TECNOLOGIAS DA INFORMAÇÃO & SOCIEDADE

Teoria · Usos · Impactos

INFORMATION TECHNOLOGY & SOCIETY

Theory · Uses · Impacts

Organizado por / Edited by : João Lopes Alves

APDC
associação portuguesa
para o desenvolvimento
das comunicações

SPF
sociedade
portuguesa de
filosofia

Sociedade
Portuguesa
de Filosofia

TECNOLOGIAS DA INFORMAÇÃO & SOCIEDADE

TEORIA · USOS · IMPACTOS

INFORMATION TECHNOLOGY & SOCIETY

THEORY · USES · IMPACTS

AGRADECIMENTOS

A publicação deste livro beneficiou dos apoios concedidos por:

ACKNOWLEDGEMENTS

The publication of this volume was supported by the sponsorship of:

DATA GENERAL PORTUGAL

EUROFONE – TELECOMUNICAÇÕES MÓVEIS, SA
(Empresa do Grupo SONAE, Cable & Wireless
e Banco Português de Investimento)

FUNDAÇÃO CALOUSTE GULBENKIAN

TECNOLOGIAS DA INFORMAÇÃO & SOCIEDADE

TEORIA · USOS · IMPACTOS

INFORMATION TECHNOLOGY & SOCIETY

THEORY · USES · IMPACTS

Organizado por / Edited by
João Lopes Alves

ACTAS DO SIMPÓSIO INTERNACIONAL SOBRE COMUNICAÇÃO, SIGNIFICAÇÃO E CONHECIMENTO FACE ÀS TECNOLOGIAS DA INFORMAÇÃO

PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON COMMUNICATION, MEANING, AND KNOWLEDGE FACING THE INFORMATION TECHNOLOGY

Lisboa, 13–15 Setembro, 1989
Lisbon, September 13–15, 1989

A P D C

associação portuguesa
para o desenvolvimento
das comunicações

S P F

sociedade portuguesa
de filosofia

FICHA TÉCNICA

Título TECNOLOGIAS DA INFORMAÇÃO E SOCIEDADE. TEORIA, USOS, IMPACTOS
INFORMATION TECHNOLOGY AND SOCIETY. THEORY, USES, IMPACTS

Actas do Simpósio Internacional sobre Comunicação, Significação e Conhecimento face às Tecnologias da Informação/ Proceedings of the International Symposium on Communication, Meaning, and Knowledge Facing the Information Technology

(Lisboa, 13-15 Setembro, 1989/ Lisbon, September 13-15, 1989)

Edição Associação Portuguesa para o Desenvolvimento das Comunicações (APDC)
Rua Tomaz Ribeiro, nº 41 - 8º Esq., 1000 Lisboa, Portugal

Sociedade Portuguesa de Filosofia (SPF)
Avenida da República, nº 37 - 4º, 1000 Lisboa Portugal

Organizador João Lopes Alves

Capa João Braamcamp Sobral

Composição e Impressão *Colibri* - Sociedade de Artes Gráficas, Lda.
Alameda da Universidade - 1699 LISBOA Codex, PORTUGAL

Depósito Legal Nº: 52.769/92 Acabado de imprimir em Fevereiro de 1992

ÍNDICE / CONTENTS

João Lopes Alves: *Apresentação* 7

I · USOS DA TEORIA / USES OF THEORY

1 · Computação & Inteligência Artificial / Computation & Artificial Intelligence

Aaron Sloman: *The relevance of computation to the study of mind* 41
Marcelo Dascal: *Why does the language matter to Artificial Intelligence?* 52
Luc Steels: *Issues and open problems in subsymbolic representation* 79
Daniel Andler: *Intelligence Artificielle, Philosophie et Sciences Cognitives* 95
Avron Polakow: *ASS (Artificial Stupidity Systems)* 112

2 · Lógica & Racionalidade / Logic & Rationality

Newton C. A. da Costa: *Lógica e Inteligência Artificial* 127
Antonio A. Martino & Carlos E. Alchourrón: *Lógica sem Verdade* 133
Ruy J. G. B. de Queiroz: *Meaning, Use, Consequences – Interconnected concepts* 151
Eugene J. Meehan: *Logical models of human judgements: limits and possibilities* 167
Ruth Manor: *Rational question raising* 176

3 · Inteligência Natural & Inteligência Artificial / Natural & Artificial Intelligence

Roland Fischer: *Neuroepistemological aspects of natural and artificial intelligence* 191
Hung-Yul So: *The Artificial and the Popperian World* 201
Manuel Liz: *Mental contents – Attributions and Superveniences* 206
Ng Tai-Kee: *Static and dynamic structures of consciousness in Psychology* 214

4 · Informação, Comunicação, Representação & Significação / Information, Communication, Representation & Meaning

Pedro Roberto Medina Martins: *Comunicação, Conversação e Cibernética* 223
Rui B. Vieira Marques: *Information Theory – Communicative Data* 243
Mukesh J. Patel: *Relationship between Meaning and Representation: an experimental approach* 253
Jean-Claude Perez: *Peut-il y avoir mémoire sans représentation?* 260
Klaus D. Dutz: *Models of models of Meaning. Representation, Knowledge and Communication in Cognitive Linguistics and in Leibniz* 271
Danilo Marcondes de Souza Filho: *A questão das linguagens artificiais no pensamento moderno* 278

5 · A Ciência perante a complexidade e a incerteza / Science facing complexity and uncertainty

Abraham A. Moles: *Y a-t-il des Sciences de l'Imprécis?* 287
Manuel José Lopes da Silva: *Hierarquia e Complexidade* 303

Carlos Fiolhais: *Física, Computação e a "inteligência" da Natureza* 310
Basilio Catania: *The Physics of the Boolean Observer* 317

II · TEORIA DOS USOS / THEORY OF USES

1 · Concepção & Organização / Design & Organization
Ronald Marlin Lee: *Limits to Technological Rationalization: Sociology of Semantics* 331
Margarita Vasquez & Manuel Liz: *Designing Systems* 345
Ionna Tsivacou: *Generation and transference of meaning in Generic Design* 349
David Keever: *Consensual design of an organization's management system* 358
F. Ross Janes: *Planning for an industrial training Board Team: an application of interactive management* 367
James T. Coulter Wright: *Análise e Estruturação de Modelos na selecção de projectos de pesquisa* 375
Ilona Kovacs & A. Brandão Moniz: *As tendências tecnocêntricas e antropocêntricas na concepção de sistemas integrados de produção* 384
J. N. Lacerda Fonseca: *Algumas relações entre tipos de distribuição dos conhecimentos e as estruturas da cooperação e da realização material* 394

2 · Educação / Education
Duarte Costa Pereira: *Modelos como representações nas estratégias de confrontação para aprendizagem da ciência* 411
João Pedro Ponte: *O computador como ferramenta. Resultados de investigação* 431
Teresa Levy: *Computers in use, communication in process* 438
Asher Idan: *Educational ideologies in the Informational Age* 446

3 · Linguagens & Hermenêuticas / Languages & Hermeneutics
André Robinet: *Les applications de l'informatique à l'étude des textes* 453
Ephraim Nissan: *Derivational Knowledge and the common sense of coping with the incompleteness of lexical knowledge (ONOMATURGE)* 462
José António Barreiros: *A arte combinatória jurídica no século das TI* 478
E. Nissan, G. Puni & T. Kuflik: *Finding excuses with ALIBI: alternative plans that are deontically more defensible* 484
António J. R. Ferreira: *Atitudes musicais e metáforas computacionais: uma visão pessoal* 511

III · IMPACTOS GLOBAIS / GLOBAL IMPACTS

Jesus Mosterín: *The role of communication in the dynamics of cultural change* 523
Valdemar W. Setzer: *The Computerized Society: new life or destruction?* 531
John N. Warfield: *The Artificial Philosophy and the decline of Technology* 542

João Lopes Alves
Lisboa (Portugal)

Apresentação

IT makes things possible which have
never been possible before.
Iain Mclean

La question maintenant est de maîtriser
la maîtrise et non plus la nature.
Michel Serres

A presente colectânea de estudos teve origem num Simpósio Internacional realizado em Lisboa.* Quiz-se com aquele Simpósio promover um debate interdisciplinar sobre as repercussões, inesgotavelmente em aberto, que a invasão intensiva de todas as práticas humanas pelas tecnologias da informação <TI> está a introduzir nos percursos cruzados da comunicação, da significação e do conhecimento; o campo de debate estender-se-ia desde os alicerces teóricos dos diversos avatares das TI até aos impactos sociais de toda a ordem provocados pela generalização de usos da computação electrónica, conjugada com a produção de efeitos de "inteligência" pela máquina. Como é natural, o sentido dos estudos ora editados reflecte as intenções do encontro científico de origem.

Centrado o temário, genericamente, em torno dos fenómenos "materiais" e "espirituais" que irradiam das novas TI, há que admitir à partida que se tem de pisar um terreno minado de ambiguidades.

Na verdade, quando se fala hoje de TI pode-se estar a designar, sob cobertura da mesma expressão, campos de análise relacionáveis sem dúvida, mas não coincidentes.

O alvo das preocupações pode ser um corpo de engenharias particulares, nas suas tecnicidade e economia industrial próprias: as tecnologias do computador, em sentido estrito (mecânicas e sistemas de computação digital e de inteligência artificial <IA>) as tecnologias ditas de controlo (robótica, projectos e/ou fabricos assistidos por computador, sistemas de gestão e planeamento de base informática, etc.) e as novas tecnologias da comunicação (combinatórias da computação digital e das técnicas e sistemas de comutação e transmissão à distância dos vários tipos de mensagens: voz, texto, dados, gráficos, imagens). É a famosa tríade de CCs das TI — computar, controlar, comunicar [1] — que tem como núcleo duro comum aquele artefacto por excelência polimórfico e polivalente, verdadeiramente digno de Proteu, para usarmos uma expressão de Simon (1981, p. 17), actuante no cerne de todos os desenvolvimentos tecnológicos do trato com a "informação": o computador dos nossos dias.

*Simpósio Internacional sobre Comunicação, Significação & Conhecimento face às Tecnologias da Informação, 13-15 de Setembro de 1989, promovido e organizado por a Sociedade Portuguesa de Filosofia (SPF) e a Associação Portuguesa para o Desenvolvimento das Comunicações (APDC). O autor desta apresentação partilhou com o Prof. Eng. M.J. Lopes da Silva funções de coordenação da comissão organizadora do Simpósio.

Podemos também, no entanto, querer aludir sobretudo a algo que, pressupondo as proezas tecnológicas, se não esgota problematicamente nessa dimensão: as múltiplas valências de uso associadas às TI. E se na primeira abordagem caberia falar principalmente de tecnologia industrial, na segunda ver-nos-íamos compelidos, se aspirássemos à exaustão, a considerar um caleidoscópio de pontos de vista: sociologia e psicologia, lógica, antropologia, neurofisiologia e medecina, linguística, educação, documentalística e biblioteconomia, a investigação científica e tecnológica em geral, todo o vasto mundo dos *media* de comunicação, evidentemente, mas também arquitectura e urbanística, belas-artes, ciências da organização, mercados financeiros, as grandes linhas da macroeconomia, políticas de emprego, de defesa, etc., etc., etc.. A generalidade das ciências e artes de paz e de guerra, em suma. Fazendo bem as contas, podemos descobrir-nos a considerar tudo que à vida humana interessa, ou pouco menos de tudo, desde as grandes estratégias do poder mundial que passam pelos Estados, pelas agências internacionais, por conglomerados empresariais de projecção planetária, até às práticas individuais mais anodinas (mas serão anodinas?) como transferir electronicamente fundos de uma conta pessoal ou fazer alguma das múltiplas consultas caseiras facilitadas pelos serviços de videotex.

Mas, mesmo porque tudo é muito, e este muito talvez seja demais como mapa de estudos, podemos ainda preferir concentrar a nossa ambição e restringir-nos às fronteiras de rigor de um domínio científico tipicamente "informacional". É, caracteristicamente o caso da ciência, ou, melhor dito, do *Konzern* científico compósito — a ciência , ou ciências da cognição — que, florescendo com raízes nos modelos computacionais, alimentou a pretensão de, *grosso modo*, tomar o próprio conhecimento "em-si" como alvo directo, materializado, da análise ("computer science makes it possible to manipulate ideas as though they were things": Minsky, 1983) emancipado de vinculações que, nas disciplinas tradicionais da psicologia e da epistemologia, prejudicam a sua possível (e altamente discutível) autonomia de objecto de uma ciência própria. Só que esta restrição de focagem não aboliria a ambiguidade, bem pelo contrário, tantas são as perplexidades acerca do preciso recorte (inteligência? "espírito" <mind>? cognição? conhecimento?) do que deva ser objecto das novas disciplinas, ou corpos disciplinares, emergentes do "cognitivismo", assim como das suas orientações e dos seus resultados (cfr. Andler, *infra*).

Como quer que seja, um dos traços característicos das modernas TI na ligação às ciências cognitivas é que, além de constituirem um bom testemunho da confusão de águas entre tecnologia e pesquisa fundamental a que se assiste hoje em dia na maior parte dos ramos da ciência, "o conhecimento passou a estar vinculado *tangivelmente* a uma tecnologia que transforma as práticas sociais sobre que repousa — sendo a inteligência artificial o exemplo mais flagrante disso mesmo" (Varela, 1988, p. 11)

Se os impactos destas novidades históricas[2] na filosofia, na psicologia, na linguística ou, de um modo geral, nas áreas da chamada ciência cognitiva são patentes, acresce que a observação de Varela acerca do conhecimento se pode reproduzir, *mutatis mutandis*, a propósito de todas as demais expressões, cooperativas ou conflituais, da interacção humana — e eis-nos, outra vez, perante o amplo e diversificado espectro de intromissões das TI nas actuais estruturas de vida (as "materiais" e as "ideais") da humanidade.

As razões de fundo desta vocação proteiforme radicam na aptidão inerente aos sistemas computacionais de operarem a síntese entre linguagem e técnica. É uma simbiose que permite aliar nas TI o poder material das máquinas à plasticidade espiritual dos discursos simbólicos abstractos. Lidamos aqui não tanto com máquinas mas com linguagens mecânicas.

De facto, a ciência e as técnicas da computação conseguiram traduzir o "paradigma simbólico" de conceitos e teorias cujo campo de operação se confinava principalmente às ciências humanas (lógica, linguística, psicologia) em estruturas tratáveis mecanicamente (vd. p.e. *infra* Dascal, Steels, Nissan et al.). Podendo colonizar, por dentro, todos os dispositivos, de certo modo preparados para as receber, do mundo industrial, as inovações "informacionais" como que conseguiram jogar o mesmo lance ganhador em dois tabuleiros simultaneamente: envasar a força simbólica do "espírito" na materialidade maquinal e trazer o poder da máquina às tarefas espirituais.

Foi este logro notável que conferiu às TI o potencial — único no concerto das tecnologias — de *a sua grande novidade consistir em que são fonte, agente, matriz e acelerador de todas as mais novidades*. Stonier (1988) observou, a propósito, que um dos aspectos mais interessantes nas TI, hoje, é o de conduzirem <drive> a evolução tecnológica *em geral* de modo análogo àquele como a tecnologia conduziu a evolução social em fases históricas anteriores; na condição de se substituir (ou de se entender) conduzir por induzir, a analogia é provavelmente muito instrutiva. Seja como for, o êxito no terreno mostra-se fulminante: pouco mais de trinta anos após a comercialização dos primeiros computadores digitais, quando, na actualidade, se invoca o vocábulo cintilante do "novo" — novas tecnologias, novas ciências, novos produtos e serviços, novas profissões, em resumo: novas realidades — pode estar-se certo de que, de maneira directa ou indirecta, no proscénio ou nos bastidores, é a presença activa de TI o que dá forma à novidade daquilo que se diz ser novo.

Donde que a temática das TI se tenha aberto a um outro horizonte de indagação, entre todos o mais ambicioso mas também o de maior risco especulativo: não será que a importância expansivamente crescente e as múltiplas aplicações das TI exibem virtualidades de poder vir a sinalizar como que o advento de uma rede de "infraestruturação" global das sociedades humanas, historicamente inédita entre os modos de conceber, produzir e repartir os bens da vida, chame-se-lhe — reflectindo matizes que são significativos na sua diferença — "sociedade tecnológica", "sociedade pós-industrial", "sociedade da informação"?

Fundamentalmente, estaria, então, em causa não (apenas) a tecnologia particular de dadas gerações de máquinas ou sistemas de mecanismo, nem (somente) os traços específicos de usos sociais e privados, mas que se houvesse afirmado, como força motriz e modelador principal das relações humanas, um certo tipo de "mecanização virtual", para o dizermos à maneira de Sloman, embora em distinto contexto. Seria algo a entender como "an abstract structure which can undergo various changes of state" e que, emancipando-se do *hardware* físico, teria muito de comum com o tipo de sistemas formais estudado por matemáticos e lógicos (Sloman, 1984, p. 36; cfr. também o seu estudo, *infra*); daqui irradiaria uma peculiar "logotecnicização" difusa das formações sociais, com alargamento dos tipos de "performatividade" da máquina — mais precisamente: de um modo novo de mecanização "abstracta" — à generalidade das esferas da vida (cfr. Rodrigues, 1990).

Podemos, então, interrogar-nos sobre se não estamos a enfrentar qualquer coisa como um mundo de super-máquinas, produzindo nas interrelações com os sujeitos humanos efeitos de gigantescos misturadores automáticos de almas, a uma escala vertiginosamente macroscópica, se comparada à do triturador ainda artesanal de corpos que era a máquina dos "Tempos Modernos" de Chaplin.

À primeira vista, a resposta talvez deva ser negativa. É que a trituração, se a há, mostra-se muito mais subtil, muito mais vaporosa. Em lugar dos mastodontes mecânicos dos primórdios da informática, realmente super-máquinas, os avanços da micro-

electrónica permitem que a teia ideal-material suportada pelas TI se apoie, hoje em dia, não num exército de *Panzers* gigantescos mas em redes de conexões mediadas por "infra-máquinas". De facto, as arquitectónicas produtivas/distributivas que caracterizam diferencialmente a "era industrial" e a "era informacional" exibem o contraste de, na primeira, os processos produtivos dominantes (grandes concentrações de fontes de energia termodinâmica) se configurarem como destruições transformativas de macro-matéria realizadas em macro-espaços enquanto o quadro de referências da segunda, em que o elemento estruturante passou a ser a energia virtual da informação como factor de organização de matéria e energia "real", aponta a condensar micro-matéria em micro--espaços. O que se vai construindo de novo, em *overlay* com as estruturas industriais existentes, baseia-se em sistemas muito finos de redes de transmissão, comutação e recepção de mensagens, cada vez mais densos, mais envolventes e de polivalência acrescida nas utilizações que facultam. São sistemas em que o factor de produção preponderante passou a residir na qualidade do *software*, ou seja, no trabalho intelectual (sobre as consequências sociais deste aspecto, em particular na estrutura de classes, cfr. p.e. Bell, 1983; vd. também Idan, *infra*); paralelamente, o *hardware* vai perdendo, pouco a pouco, a importância determinante que lhe pertenceu no modo de produção típico das primeiras "revoluções industriais", com as suas tecnologias pesadas. Ao mesmo tempo, expandiram-se, em mancha de óleo, as teias do que se designa usualmente por "economia de serviços" (recorro à expressão sem cuidados de rigor) para, nos últimos tempos, graças ao suporte operativo proporcionado pelas novas tecnologias da comunicação (i.e. das vias e meios de emissão, transporte e recepção de informação) se gerar uma dinâmica que se alimenta muito mais de fluxos financeiros negociados em contínuo, a uma escala verdadeiramente mundializada, do que dos termos de troca de mercadorias e prestações de serviços. Anote-se, de passagem, que as consequências cruciais, em todo o mapa da interacção humana, de a esfera da "aparência" improdutiva se ter tornado a nova *realidade* dominante por sobre (se não *contra*) a racionalidade da "economia real" da produção e dos intercâmbios de bens e serviços ainda não foram assumidas estruturadamente pela teoria económica. Está por aparecer o Keynes do que já se chama "Symbol Economy".

Não são estes os únicos contrastes entre as linhas de força da "revolução industrial" e as da "revolução informacional" (que talvez seja mais apropriado designar por "reforma revolucionária" ...). O diferente significado histórico do que está, e esteve, em jogo deixa-se ainda ver, também muito instrutivamente, pela comparação entre o ambiente convulsivo em que decorreram os processos de industrialização de formações sociais agrárias e o estilo *soft* da invasão paulatina das estruturas urbano-industriais pelos sistemas e redes de base informática. Recorde-se que as chamadas "revoluções industriais" foram sempre fenómenos drasticamente agressivos: arrazam-se os eco-sistemas agrários tradicionais, criam-se condições de ruína das massas de campesinato que as convertem compulsivamente em proletariado de indústria, promovem-se concentrações fabris e urbanas com impactos devastadores no ordenamento "humanizado" do território e de gestão quase impossível pela desmesura, desencadeiam-se cataclismos ambientais em cascata, ergue-se o cenário de dolorosos choques culturais, de violentas lutas de classes, de terremotos sociais e políticos demolidores, de competição feroz pela apropriação de recursos naturais, suscitando-se um estado endémico de guerra, declarada ou latente, nas relações internacionais ... Em primeira leitura, não são essas as expectativas razoavelmente admissíveis na fase de passagem à "information-oriented society": não se detectam factores de agravamento da conflitualidade social, pelo menos em larga escala, imputáveis às TI, há, apesar de tudo, indícios de alguma distensão nos confrontos internacionais, não é visível que a

disseminação de usos das TI provoque agressões aos eco-sistemas físicos da vida (a preocupação por efeitos poluentes parece circunscrita às zonas do "espírito"). Lobriga-se, na comparação destes distintos *backgrounds*, um traço diferenciador fundamental: enquanto a industrialização se faz, necessariamente, *contra* as estruturas sócio-produtivas pré-existentes e as destrói, a informatização tem-se feito *nas* e *com* as estruturas industriais estabelecidas de que nasce e que potencía, antes de as reformar por dentro.

Num cenário como o esquissado, o estabelecimento de analogias entre as redes de informação suportadas pela computação electrónica e os sistemas nervosos biológicos foi sempre uma tentação irresistível, até porque, contrastando com a corrente eléctrica "inerte", a forma de "energia" circulante nos circuitos electrónicos passou a ser corrente "informacional" inter-activa: os sistemas informacionais relacionar-se-iam com a sociedade não apenas como núcleos tecnológicos exteriores, à espera de oportunidade (mercantil, designadamente) para entrarem no jogo social, mas, muito mais e outra coisa do que isso, como estruturas lógico-materiais complexas que fariam parte, intrínseca e constitutivamente, do organismo social, tal como o sistema nervoso participa, de modo substantivo, do organismo biológico.

Trata-se de uma perspectivação das coisas (fortemente estimulada, nos últimos tempos, pelas inovações técnicas aparatosas — é o termo! — obtidas com recurso às TI) que nos convoca a fazer contas, na base, com a matéria prima "ideal" trabalhada pela computação digital: a "informação". Sem que devamos perder de vista que o próprio conceito de "informação" mais do que um campo estável de dados de referência para a investigação recobre uma zona de problemas em grande ebulição teórica, certo é que nasceu daqui uma visão do mundo que esperou, e ainda espera, poder explicar, mediante sucessivas transposições, a estrutura da matéria, os fenómenos cognitivos e a complexidade social através do estudo das redes de circulação e comutação de informação actuantes nas várias formas de vida, quer as "naturais", quer *a fortiori* as "artificiais" (vd. *infra* p.e. Mosterín, Fischer, Catania). Concomitantemente, foi colhendo seiva teórica neste húmus a convicção de que as trocas de informação entre homens e homens, entre homens e máquinas e entre máquinas e máquinas estão destinadas a um papel de crescente importância nos rumos da espécie humana. Note-se que o reconhecimento da importância tanto pode acenar a promessas de emancipação (vd. Mosterín, *infra*) como suscitar juízos condenatórios, em razão das ameaças de dispersão e definhamento digamos ético-cultural que o fenómeno traria no bojo (cfr. p.e. Setzer, *infra*). Os argumentos de condenação crítica são diversos, embora animados, todos eles, por uma intenção comum de recusa do que se pode chamar o fenómeno de logo-mecanização generalizada das estruturas da vida humana: há quem denuncie o *formalismo* e a *abstracção* do tipo de discurso processado pelos programas informáticos, alicerçados numa modelação algoritmica que empobrece irremediavelmente a informação por esvaziamento dos conteúdos semânticos (é uma das tónicas do criticismo de Setzer), assim como se põem embargos à adscrição à máquina de tarefas que, por natureza, deveriam ficar reservadas, sob pena de graves danos, às faculdades de empatia compreensiva, de globalização intuitiva ou de avaliação valorativa próprias do espírito humano (cfr. *infra* p.e. Meehan, Levy, Setzer).

Como substracto mais profundo dos temores, latejam atitudes intelectuais que, não sendo inéditas nos horizontes problemáticos do pensamento ocidental face à técnica, colhem considerável alento das virtualidades proteiformes e expansivas evidenciadas pelo tipo de dispositivos materiais/espirituais que as TI viabilizam. O grande receio, como exprimiu recentemente um filósofo espanhol, anda à volta de que o domínio da razão sobre a natureza ameaça resolver-se numa "dissolução natural do racional" (Duque, 1988, p.

299) e os danos diagnosticados dispersam-se em múltiplas direcções. Desde o forte impulso àquela desagregação de *o saber* pela criação contínua de linguagens especializadas que, mau grado parecer inerente ao progresso das ciências, foi sempre vista com olhar hostil do lado da filosofia — "el especialista en lo particular (y todos lo somos ya, seamos o no científicos) es lo analfabeto *en general*" (Duque, *ibid*, p. 303) — até ao risco orwelliano dos sistemas de regulação e de hetero e autocontrolo técnicos que a lógica do modo "informacional" de produção e reprodução dos bens da vida incubaria nas modernas sociedades de massas, totalitárias ou democráticas ("todo está permitido porque todo está controlado y prohibido", *ibid.*), é largo o espectro de pânicos que a "sociedade computorizada" espalha à volta. A sociedade "pilotada" e "programada", aparentemente implícita na cibernetização electrónica do mundo, como que consumaria, ao limite, a pretensa usurpação da "essência do homem" pela "essência da técnica" de que "já só um Deus nos pode ainda salvar", anunciada num dos tropos finais da lúgubre retórica heideggeriana(3).

Julgo que aproximações deste tipo, tanto as pessimistas como as optimistas, tendem a ser muito vulneráveis ao característico virus do espírito que debilita as teorias totalizantes: a "generalização do caso claro" (Wittgenstein) que leva a incluir-se numa dada focagem de análise "all social phenomena, or the 'essence' of society", deixando abusivamente de fora as perspectivas ou experiências que não cabem no critério "totalizador" (Poster, 1990, pp. 21 ss.). Exibem, não obstante, o mérito provocativo de relevar a *generalidade* e a *interconexão* dos fenómenos de "electronically mediated communication" na geração de um novo contexto social.

De facto, pense-se o que se pensar da explosão "informacional" de massas, trazida pelos avanços da electrónica e, sobremaneira, pelo casamento entre informática e, telecomunicações, afigura-se que é dificilmente recusável a verificação de que, pelo menos nas formações sociais ditas desenvolvidas e nas de desenvolvimento intermédio(4), os indivíduos e os grupos humanos estão já, para o bem ou para o mal (talvez mais precisamente: para o bem *e* para o mal) a conceber, a organizar, a normar, a executar, a memorizar, a habitar — resumindo: *a viver* — os seus pensamentos, as suas inter-relações, os seus direitos e deveres, as suas tarefas, os seus lazeres, os seus prazeres, os seus territórios, de maneiras substantivamente novas e dentro de coordenadas de espaço / tempo radicalmente alteradas, por efeito da disseminação das TI de base computacional. A verdade é que está a mudar a relação do homem com muita coisa — a do homem "científico", a do homem "produtivo", a do homem "comum": com a técnica e com as dimensões do "natural" e do "artificial" no humano; com o que e como se pensa, com as representações simbólicas, com a linguagem, com os textos, com os sons, com as imagens; com o que se pode conhecer e o que se desconhece; com os modos e as instâncias de registo e de transmissão (ou de ocultação) do conhecimento; com os meios e os limites da racionalidade instrumental na produção e distribuição dos bens da vida; com a arte; com a organização e a fruição dos espaços públicos do trabalho e dos espaços privados do lar; com os problemas e o desenho dos imperativos ético-jurídicos, com as estruturas e fontes do poder e dos poderes, com as maneiras de fazer a guerra e com as expectativas de paz.

Num contexto de questionamentos tão amplo a apontada ambiguidade dos conceitos operativos mostra-se irreduzível. Não tem, contudo, de significar obrigatoriamente um vício ostensivo ou redibitório da teoria. De certa maneira, guarda, pelo menos, a utilidade de designar, *a contrario*, o único modo que a indagação teórica pode explorar na especialidade, se quizer fazer justiça, sem incorrer em trivialidades, a uma tal vastidão de situações imbricadas e heterógeneas: o modo *interdisciplinar* no acercamento a e na

explanação de problemas que nos pressionam inescapavelmente enquanto actores do mundo do trabalho, enquanto protagonistas de cidadania e enquanto pessoas privadas. Também enquanto praticantes responsáveis de labor técnico-científico ou de reflexão filosófica, e não apenas como problemas que se nos apresentam *de fora*, sob forma de objecto de análise. Ao mesmo tempo que solicitam do exterior as atenções das ciências e da filosofia, as TI acometem-nas *por dentro*, como uma infra-estruturação reticular cada vez mais decisivamente influente no desenho dos instrumentos materiais da investigação e dos modelos ideais do pensamento teórico (cfr. *infra* p.e. Fiolhais e Pereira).

Ora convém atender a que, neste como noutros campos, o carácter na origem instrumental dos meios técnicos propende a desenvolver, com rapidez, lógicas de conversão de uma economia de meios numa economia de fins, e é um traço interessante, p.e., que no estudo de Martino & Alchourrón, *infra*, se releve como vantagem de uma "Lógica sem verdade" a de assegurar margens de manobra alargadas à intervenção informática, uma vez que os computadores seriam receptivos apenas a instruções sintáticas. Entre as tendências emergentes, não só nem talvez principalmente na Lógica, pode desenvolver-se, a par da utilização da máquina como instrumento de potenciação e renovação do trabalho científico, a de se predispôr as orientações do trabalho científico a tratamentos afeiçoáveis às características processuais da máquina. É tópico a pôr questões tudo menos tecnologicamente neutras; limito-me a anotá-lo, já que a sua apreciação útil exorbitaria das normais restrições desta apresentação preambular.

* * *

A colecção dos estudos agora editados não esgota a cartografia de uma tal fermentação problemática em toda a sua extensão, longe disso, até pelo que se visou com o Simpósio de origem: não um inventário de "estados-da-arte" mas a suscitação interdisciplinar de "estados-de-questionamento". Sem embargo, oferece um leque muito instrutivo do alcance, da multiplicidade diversificada e da profundidade do que está em jogo com o desenvolvimento do novo arranjo das coisas humanas a que Poster, parafraseando a teoria marxiana dos modos de produção, chama "the mode of information".

O leitor encontrará nas páginas seguintes testemunhos científicos que cobrem um naipe alargadíssimo de abordagens. De como se pensa e, sobretudo, se está a pensar o pensamento sob focagens renovadas (p.e. Sloman, Steels, Andler, Polakow, Liz, Tai-Kee); de novas pistas de estudo desvinculadas de rotinas disciplinares e institucionais nas áreas de tradição da linguística (Dascal), da lógica (da Costa, Martino & Alchourrón), da teorização das formas de racionalidade (Meehan, Manor, Queiroz); de como são outros os meios da leitura e da interpretação (Robinet), bem assim do trato com linguagens textuais e musicais (Nissan, Ferreira) e se bifurcam diferentemente as encruzilhadas da representação e do sentido (Dutz, Patel, Perez); se ensina, se aprende e se investiga em ambientes instrumentais, conceptuais e ideológicos modificados em profundidade (Ponte, Pereira, Levy, Idan); se redefine, a novas luzes, o horizonte científico dos nossos conhecimentos e desconhecimentos (Moles, Catania, Fiolhais, da Silva) e se colocam questões inéditas na prática do direito (Barreiros), na possibilidade de tratar computacionalmente as controvérsias deônticas (Nissan *et al.*), na concepção, organização, gestão, métodos e enquadramento social das tarefas humanas (Lee, Vasquez & Liz, Tsivacou, Keever, Wright, Janes, Kovács & Moniz, Fonseca). São igualmente chamadas à colação perspectivações inovadoras nas formas e nos conteúdos da acção comunicativa (Martins, Marques, Mosterín), no binómio "naturalidade" / "artificialidade" (Fischer, So, e, em avaliação histórica iluminável

retrospectivamente pelos adquiridos do presente, Souza Filho), na oposição "espiritualidade" / "mecanização" (Setzer), nas difíceis relações entre ética e tecnologia (Meehan, Warfield).

Bastaria este arrolamento para convencer que a invocada interdisciplinaridade mais do que uma conveniência é aqui uma exigência. Como ilustra também, mesmo que só de maneira aproximativa, a capacidade de infiltração inexaurível do fenómeno TI.

Se o potencial de novidade e de mudança do que se encontra em marcha é, segundo cremos, incontestável, há que reconhecer, no entanto, que as *performances* técnicas, sem dúvida impressionantes, se fizeram acompanhar, em muita da literatura recente, de um triunfalismo teórico nem sempre confirmado pelos resultados. Pior: de um triunfalismo que obnubilou alguns dos problemas reais e as limitações próprias do discurso computacional.

Ora, outro dos méritos da presente colectânea assenta justamente em que não compartilha ilusões triunfalistas. Pode mesmo dizer-se que dá testemunho de uma certa deflação naquele tal ou qual embevecimento perante as seduções do discurso computacional que tem sido uma das tónicas culturais fortes do últimos anos.

Índice expressivo deste distanciamento crítico é que a noção dos limites dos modelos computacionais compareça como problema comum de vários estudos que, pelo demais, evidenciam acentuadas disparidades nos ângulos de focagem e nas opções doutrinárias dos autores.

Com efeito, são diagnosticadas limitações teóricas fundamentais, digamos assim, no sentido de embargos oponíveis a expectativas por cumprir, ou, mais drasticamente, sem possibilidade de cumprimento: limitações da reivindicação para a IA do papel de a ciência fundamental da cognição (Andler); limitações da concepção, que fez escola, de que os sistemas computacionais esgotam em si todos os tipos de mecanismo susceptíveis de intervir na configuração de sistemas inteligentes (Sloman); limitações do paradigma simbólico para resolver alguns dos problemas comummente cometidos aos sistemas de IA, em particular (mas não só) as questões da complexidade e do dinamismo (Steels); limitações do discurso computacional, pelo menos nas versões da "ortodoxia computacionalista", para dar conta de todas as dimensões de uso da linguagem (Dascal); limitações na legitimidade conceitual da atribuição de capacidades epistemológicas de "inteligência" à máquina (Polakow); limitações do cálculo computacional como metáfora explicativa da "inteligência" à obra na natureza física (Fiolhais) ou como base de tratamento científico satisfatório dos fenómenos inilidíveis da incerteza e da ambiguidade (Moles, *infra*; vd. desenvolvimentos posteriores em Moles, 1990).

Fala-se também de limites na racionalização tecnológica das organizações e das tarefas, não obstante actuar aqui um dos factores mais influentes na expansividade de usos das TI: limites no potencial de racionalização tecnológica do computador impostos pelos dispositivos necessariamente sociais, com negociação de interesses, valores e preferências, da inovação semântica (Lee), limites da perspectiva tecnocêntrica no desenvolvimento de sistemas integrados de produção (Kovács & Moniz), limites insitos na distorção biunívoca que significa a concepção tecnomórfica dos seres humanos e a concepção antropomórfica da máquina, a não negligenciar quando se introduz o computador como instrumento e como objecto do processo educativo (Levy).

Este último tipo de limitações inclui-se já em ambientes de criticismo normativo que, aliás, têm sido uma constante no trabalho de reflexão sobre a "revolução informática", desde os seus inícios. É que, ao lado de limites que se produzem na órbita do "ser" (o que se não pode esperar dos sistemas computacionais), se alegam ainda limites no plano do "dever-ser": o que se lhes *não deve pedir*, sob pena de graves lesões humanas. Assim, reclama-se que se não deve sacrificar ao "sindroma teorético-físico" do nosso tempo,

subordinando as escolhas sociais a modelos de lógica formal, incorporados ou não em dispositivos mecânicos, quando é logicamente demonstrável que aquele tipo de opções não admite a formalização e requer, por natureza, justificações de valor (Meehan), que se não deve incorrer em pragmáticas ansiosas que omitam o "hazard gap", impossível de estimar com rigor, entre a introdução de novos sistemas tecnológicos de alto risco de catástrofe e o desenvolvimento científico dos meios de controlo do risco (Warfield), que se não deve obliterar incriticamente as ameaças de empobrecimento cultural implicadas pelo tratamento generalizado da informação mediante programas ou pela substituição dos agentes humanos por computadores, em particular no processo educativo (Setzer).

* * *

Se a topologia de "limites" extractável do presente volume se esgotasse em discursos retóricos opostos, numa relação de exterioridade, às realizações e às virtualidades inovadoras das TI, o balanço de leitura seria, provavelmente, decepcionante. Não se andaria longe de mais algum rol de manifestações da típica ideologia de contra-poder que a afirmação dos poderes sociais da ciência e da tecnologia vem provocando compensatoriamente, como seu fantasma sazonal, pelo menos desde quando as primícias da era industrial encetaram o redesenho das arquitectónicas de base das sociedades humanas, com todo o lastro de custos e de agressões nos equilíbrios tradicionais de vida que bem se conhece.

Não é o caso, porém, dos elementos de criticismo explanados ao longo deste livro. Os novos traçados de legitimidade teórica, de adequação prática e de admissibilidade ética na vasta constelação das TI aqui propostos produzem-se como expressão de questionamentos não exteriores mas intrínsecos ao "modo-de-ser" dos modelos informacionais, sem por isso incorrerem na recusa estéril dos seus adquiridos. O que prevalece é a preocupação de os fazer prosperar, em termos de maior-valia humana, pela obtenção de níveis mais consistentes de fundamentação nas reivindicações teorético-práticas e de consciência crítica nos usos.

O leitor ganhará em reflectir detidamente sobre o sentido — que não é unívoco — dos limites que ficaram enumerados em esboço, pois o seu aprofundamento ajuda à partilha de águas, cada vez mais necessária na consideração do novo "mode of information", entre expectativas razoáveis e o que são miragens, entre o rastreio de ameaças reais à humanização dos homens e o jogo de sombras de fantasmagorias inócuas. Antes de mais, convém fazer a destrinça das diferentes significações de limites que se produzem em três planos distintos: na ambição teórica, na aptidão racionalizante dos usos ou por imposições de exigência ética.

Diagnosticarem-se pretensões teóricas infundadas no "computacionalismo ortodoxo" ou no "cognitivismo" em geral, abandonar-se a hegemonia computacional no desenho do inteiro campo de possibilidade de sistemas inteligentes, prospectar-se as limitações internas do "paradigma simbólico", contestar-se genuínos atributos de inteligência na maquinaria digital por comparação com a inteligência natural, etc., não significa, necessariamente, uma descoberta de limites debilitadores das expectativas teóricas globais abertas pelo surgimento das TI. Pode significar, pelo contrário, o enriquecimento da teoria, na sua dinâmica incessante de "verdade em construção", mediante a profilaxia de pretensões mal colocadas, de equívocos no perspectivar de objectivos, de enunciações falhas de sentido.

Vale por dizer que a identificação de limites postos a descoberto no território da ciência graças ao próprio trabalho científico nunca inculca, no que isso oferece de mais fecundo,

o levantamento descoroçoante de um mapa de impossibilidades ou de frustrações mas a proposta de novas fronteiras para a investigação, a redefinição de objectivos viáveis e um clareamento problemático pela limpeza de ambiguidades. Afinal, neste tipo de indagação há sempre limites, e os limites (tal como os adquiridos) são sempre provisórios.

Abundam exemplos neste livro. O mais interessante no estudo de Sloman não será tanto a prospecção dos limites da "Turing equivalence" para dar conta da generalidade de sistemas inteligentes mas que, a partir desse trabalho crítico, se avance com a proposta de um conceito mais amplo de "mecanismo"; as contestações de Dascal ao fechamento formal do computacionalismo ortodoxo acicatariam menos o leitor se não viessem encastoadas, como vêm, no seu programa onto-pragmático de plena compreensão dos actos linguísticos; Steels apela explicitamente a uma estratégia de superação dos limites do paradigma simbólico mediante uma "subsymbolic representation", por construir; ao despistar as ambiguidades e carências teóricas do cognitivismo, Andler não encerra a investigação num lago de impossibilidades, antes sugere o seu desenvolvimento por cursos mais produtivos cientificamente; quando Moles contrasta as pretensões de plena exactidão mensurável, acalentadas pelo pensamento científico ocidental, com os fluxos irredutíveis da incerteza e da imprecisão, não é para nos dirigir um convite de renúncia ao controlo cognitivo desses fenómenos mas como primeiro passo de consciência crítica na feitura de uma genuína "ciência da imprecisão", etc., etc. etc.

Semelhantemente, também os limites, arguidos por Lee, ao desígnio genérico de racionalização tecnológica das organizações e das tarefas advindos do processo socio-político (em sentido lato) que subjaz à mudança ou criatividade semântica, tal como as críticas de Kovacs & Moniz ou de Levy a certas expressões do peculiar "tecno-centrismo" estimulado pela "logo-técnica" dos dispositivos computacionais (seja no mundo da produção e circulação dos bens materiais, seja no contexto da emergência e difusão de bens do "espírito"), não têm de balizar, necessariamente, pontos fracos nas aptidões racionalizantes daqueles meios. Contribuem, sim, para situar a sua racionalidade *possível* pelo dignóstico de distorções insinuadas nos processos de utilização ou de más avaliações do potencial racionalizante do meio por ignorância da natureza do que está globalmente em jogo. E é evidente que os ganhos de consciência crítica quanto a distorções e quanto a capacidades efectivas conflituam não com a economia racional dos meios mas com usos equivocados, se não indevidos.

Em qualquer dos exemplos focados, aponta-se a reorientar, segundo expectativas verdadeiramente razoáveis, a bússola da investigação fundamental e do desenvolvimento das aplicações, o que passa pela dissipação de miragens. Com o dissipá-las, não se está a restringir mas a desobstruir de bloqueios o campo das possibilidades. Veremos, adiante, como o tipo de limites invocáveis do lado da ética transfere o debate para outro plano: não se trata aí de fazer uma profilaxia de impossibilidades (teóricas e práticas) mas de opôr argumentos de indesejabilidade ao que se reconhece ser possível.

* * *

O desenho dos limites do discurso computacional mostra ainda a virtude de trazer para a ribalta algo que o substracto fortemente "tecnicista" de todas estas questões tende a empurrar para os bastidores: o potencial problemático mais significativo da reflexão que se vai fazendo em torno à "information age" tem a ver não tanto com os aspectos mecânicos desta nova linguagem universal e seus usos mas com aspectos psicológicos, linguísticos, sociológicos. Isto é, com a compleição *humana* complexa, a níveis quer de "subjectividade",

quer de "objectividade" que, por um lado pré-existe à estruturas maquinais, por outro determina os respectivos campos de utilização, ante de encaixar, em *feedback*, o impacto dos efeitos novos que a "computorização" das sociedades por seu turno implica.

Pode-se ilustrar este comentário com alguns exemplos.

Assim, a polémica, que já começa a envelhecer, em torno à legitimidade de se adjudicarem efectivos atributos de "inteligência" aos artefactos de IA — o que, diga-se de raspão, aufere o principal do seu sentido como *problema* de que tomemos por paradigma *único* de "inteligência" (ou "espírito", ou "cognição", etc.) a mente humana natural[5] — projecta os efeitos mais produtivos teoricamente não sobre questões de estrita "engenharia do conhecimento" (seja como for, os engenheiros vão concebendo e construindo arquitecturas de *software* e *hardware* que, chame-se-lhes "inteligentes" ou não, ou mesmo que se lhes chame "estúpidas", conseguem resolver, pior ou melhor, mas sempre em termos de resposta útil que sem elas *não seria possível*, os problemas de intervenção "inteligente" ou assim parecida que as suscitam) mas sobre os *aprofundamentos*, sem exagero revolucionários, introduzidos em retroacção na linguística, na lógica, na psicologia da consciência, na teoria da acção racional, etc.

A verdade é que foi o surgimento deste campo novo de artificialidade que nos compeliu, e compele, a aprofundar o sentido das relações de semelhança / dissemelhança entre mente natural e "mentes" mecânicas. Obrigados a saber *tudo* sobre a mecanicidade dos nossos artefactos (nesta área, as falhas de dominio cognitivo pagam-se implacavelmente ao preço forte de fracassos de operacionalidade), fomos chamados, em paralelo, a renovar o sentido das investigações e das idéias vigentes sobre a zona de considerável opacidade que continua a opor-nos a relação *mind/body*. Artificialidade e naturalidade iluminam-se reciprocamente numa dialectica inter-activa em que a sucessiva desproblematização intrínseca à técnica que os avanços desta lhe trazem se desdobra numa re-problematização adensada e agudizada da teoria do "natural". O livro recente de Penrose (1990) sobre os processos mentais testemunha muito bem o modo como este jogo de tensões contraditórias produz efeitos na teoria.

Analogamente, o debate em aberto entre defensores das virtudes educacionais do computador, à Papert, e opositores, à Setzer, além de se poder reflectir em contributos de melhoria na engenharia informática pertinente, já nos fez, e isso com certeza, progredir alguma coisa de importante na consciência crítica do que significa ensinar, aprender, educar.

Na mesma linha de detecção do tipo de problemas meta-tecnológicos desencadeados pelas TI, damo-nos conta, na análise económica, de que os novos meios técnicos de comunicação, animados de uma dinâmica de inovação sem exagero vertiginosa, consumaram a transformação dos mercados, de espaços virtuais de troca, nas redes públicas e privativas que viabilizam o intercâmbio integrado, em tempo real e à escala do planeta, de mensagens de diversos tipos, de um modo que vai abolindo, tendencialmente, as coordenadas físicas e institucionais de espaço / tempo dentro das quais se moviam as actividades da produção, da distribuição, da intermediação e do consumo. Do mesmo passo, são subvertidos os próprios limites da geografia política. Mas quando tateamos, por isso, na difícil procura de modelos de reorganização político-económica adequados a esta realidade explosiva e implosivamente nova, as questões verdadeiramente candentes que nos aparecem não são as tecnológicas, são as de economia política. Os produtos da tecnologia, que *já lá estão*, vão continuar a estar expansivamente, salvo a superveniência de catástrofes sociais apocalípticas, como comenta, de passagem, Mosterín, *infra*; entretanto, suscitam-se-nos interrogações com uma radicalidade de perturbação que transborda por todos os lados da esfera

tecnológica: o que passou a *significar* o feixe de fenómenos que apreendiamos sob a noção de mercado[6]? Como "modelar" estratégias de conhecimento e de actuação, individuais e colectivas, com pregnância nas novas realidades? O que pressupõem e o que implicam nos arranjos (e desarranjos) estruturais das sociedades humanas? etc., etc.

Mil outros exemplos acodem em catadupa — e este aluvião problemático, já o dissémos, é, justamente, uma das características maiores da simbiose tecnologia / sociedade activada pelas modernas TI.

Concorre aqui, a par de uma considerável dose de novidade, a reconfirmação de nódulos problemáticos que a consciência teórica ocidental há muito mantém em tela de juízo.

Na verdade, desde Hegel e Saint-Simon, pelo menos[7], a elucidação do papel das técnicas no inteiro contexto das realidades humanas tornou-se um dos tópicos maiores da reflexão filosófica e da teoria social. E de estranhar seria que assim não acontecesse, tão avassaladoramente começara a impôr-se a omnipresença do "meio técnico" por sobre o "meio natural" na vida da espécie humana, inscrita e escrita, de maneira irreversível, numa malha de múltiplos artefactos, sejam-no os objectos materiais, as sequências de símbolos ou os grandes "artifícios" colectivos das formas sociais da interacção (acerca da supremacia ganha pelo reino da artificialidade nos eco-sistemas da existência humana, vd. p.e. Simon, 1981).

As questões cruciais de caracterização, de periodicização histórica e de avaliação dos impactos das técnicas e da tecnologia na economia global das relações humanas permanecem polemicamente em aberto (sobretudo, o problema, posto por Marx na ordem do dia das preocupações humanas, de qual a força operativa principal no processo de interacções que liga o progresso técnico às transformações e à morfologia das sociedades). No entanto talvez não haja abuso em admitir que se foi formando um depósito de idéias-força que tendem, se não à aceitação, pelo menos ao aproveitamento teórico generalizado. Será o caso do reconhecimento, atrás aflorado, de que vigora nas técnicas humanas uma propensão irresistível a converterem-se de simples "meios" em "fins" autónomos de sociabilidade e cultura (é uma das noções chave remontáveis à peculiar teleo-lógica de Hegel — segundo a qual o *quid* da finalidade humana consiste em que o fim realizado se torna um meio na justa medida de o meio se tornar o fim realizado[8] — e que, com distinto grau de elaboração, se prolonga no estatuto basilar atribuído às "forças produtivas" pela teoria das transformações sociais devida a Marx); ou a decorrente compreensão de que a "essência da técnica" desborda as balizas de neutralidade inferível de uma leitura superficial do fenómeno como mera produção de instrumentos (já o Bergson de *L'Évolution Créatrice* e o Spengler de *Der Mensch und die Technik*, por exemplo, sublinharam que, muito mais do que instrumentalidade, a técnica significa, segundo Bergson, uma dimensão originária do homem, ou, segundo Spengler, algo como uma "táctica vital da espécie" que decide dos sentidos ganhos pela história geral da espécie que a humanidade corporiza[9]; idéias estas que, em combinatória com outras referências alimentadoras do registo agónico próprio do pensamento de Heidegger, infiltrarão *a contrario* o seu célebre filosofema de que a técnica moderna é interpretável não em termos de neutralidade de instrumento mas como o desvendamento terminal de um persistente processo de "olvido do ser" que corresponderia a qualquer coisa de semelhante a um alongado suicídio espiritual da humanidade); ou ainda as várias (amiúde discrepantes) expressões da noção de que o conteúdo "ideal" da comunicação é sobredeterminado pelas arquitecturas técnicas de suporte do contexto comunicacional (filão ideológico, tornado lugar comum pelos trabalhos de McLuhan, de que se socorrem todos os desenhos do

"mode of information" como gramática geral da interacção humana, designadamente a influente teoria da acção comunicativa de Habermas).

Embora eminentemente discutível todo ele e campo de travessia de grandes turbulências polémicas, ou talvez por isso mesmo, este espólio reflexivo continua a guardar importantes virtualidades de iluminação do sentido da tecnicidade no relacionamento dos homens com o seu mundo. Formou-se daí um património teórico-crítico acumulado que convém não alijar no trato com a "information-oriented society", tanto mais que, neste como noutros domínios (ou talvez mais neste do que noutros domínios), a genealogia das criações humanas se fabrica, não de rupturas absolutas, mas de um processo de continuidades / descontinuidades em que do "velho" se faz dialecticamente o "novo" (quando não é o "velho" a sufocar o "novo").

Os traços de continuidade levaram mesmo alguns autores a sustentar que o tema da novidade de fundo das TI recobre uma grande ilusão ideológica, se não intuitos de propagandismo puro e simples. Alega-se, neste sentido, que, no plano da inovação tecnológica, as TI não passam de meros desenvolvimentos de linhas de força já actuantes na tecnologia industrial; que a existência de correntes íntimas de interacção entre estruturas tecnológicas e estruturas sociais se inclui numa antiga constante histórica, em cuja lógica as TI nada introduzem de substancialmente novo; que os fundamentos e as leis da actividade económica não sofreram alterações por efeito da disseminação das TI, a não ser em aspectos de superfície; ou que a acção comunicativa nas sociedades contemporâneas, embora podendo exibir importantes modificações adjectivas devido às novas tecnologias, permaneceria, não obstante, idêntica na sua natureza, nos seus objectivos e nas suas consequências. Resumindo: em vez de reclamar novidades, deveríamos, sim, fazer o diagnóstico de uma potenciação de linhas de continuidade.

Esta polémica resolve-se muitas vezes, como tantas outras polémicas, numa competição de equívocos contrapostos. É verdade que as pretensões de ineditismo da "information age" aparecem, nalguns casos, revestidas de uma irritante insensibilidade histórico-sociológica que leva à reivindicação de prerrogativas de novidade absoluta, radical, obviamente improcedentes ou a entusiásticos exercícios de "computopia" (a expressão é de Masuda, 1981), imersos de pleno na velha ilusão do determinismo tecnológico que atribui à tecnologia, como tal, poderes de *deos ex machina* soberano para "modelar" as formas e as relações de sociabilidade; e daí, muita da exasperação reactiva de contestações que ora caem no exagero, simetricamente antagónico, de recusar atributos de efectiva novidade a fenómenos que, de toda a evidência, são novos nas formas tecno-sociais que os manifestam, ora propendem a secundarizar a relevância global daquilo que é patentemente relevante. Mas, se é certo que não há, nunca houve, revoluções *ex nihilo* – – muito menos a "revolução informática", que, nas origens, nos produtos, no sentido das suas consequências, em resumo: em toda a dimensão da sua *possibilidade* histórica, se imiscui indissoluvelmente no mundo industrial, tendendo a "re-in-formá-lo", em *overlay* com as estruturas existentes, como já tive ensejo de anotar — isso não acarreta que se deva negar o significado de novidade daquilo que, como factor de um ambiente social efectivamente novo, vai emergindo nos percursos da humanidade. E o complexo fenómeno TI, tão ambíguo nos seus contornos técnicos e sociais (que são sociais enquanto técnicos e técnicos enquanto sociais), embora prolongando ou agudizando muitas das linhas de força caracterizadoras daquele estádio nas "tácticas vitais" da espécie que Mumford, nos já distantes anos trinta, chamou, com sugestivo poder de caracterização, "sociedade neotécnica"[10], recorta um traçado de impactos que avaliza convincentemente, segundo creio, o ponto de vista dos que tomam a "information society" como o quadro de referências de um novo patamar do processo histórico.

Limitar-me-ei a recensear, ligeiramente, alguns aspectos que, sem quaisquer pretensões de originalidade, que será nenhuma, tenho por mais marcantes no delineamento diferenciador das novas realidades.

Registe-se, em primeiro lugar, o impacto revolucionário que tem, obrigatoriamente, de significar, no quadro geral das interacções humanas, a computorização intensiva não só de tarefas produtivas como de *toda a concepção, organização, gestão e sistemas de circulação dos bens da vida*.

Como quer que a teoria se decida acerca da questão da "inteligência" dos dispositivos computacionais, parece irrecusável que esta entrada, e em força, de factores mecânicos na reserva de caça da pretensa especificidade humana que sempre fora o trabalho intelectual constitui um grande acontecimento histórico. *Acontecimento espiritual*: algo de novo se intrometeu na auto-apreciação que os homens vão fazendo de si e das suas criações técnicas, gerando-se, mais do que uma nova consciência teórica, uma nova "consciência cultural". *Acontecimento técnico-económico*: basta lembrar que a automação "inteligente" das tarefas da concepção, da produção e da racionalização procedimental tende, em medida crescentemente alargada, a transformar o trabalhador humano *médio* das organizações administrativas e produtivas *(latu sensu)*, à obra nas sociedades ditas avançadas, de principal protagonista activo nos cenários do trabalho em supervisor ou capataz de agentes mecânicos. Por outras palavras: sem deixar de trabalhar, o homem retira-se do âmago dos processos produtivos, e estranho seria que esta alteração verdadeiramente radical não devesse transtornar toda a economia de equilíbrios e desiquilíbrios, de cooperação e de conflitos, de continuidade e de crises, em formações sociais que, reconhecidamente, têm como centro de gravidade os modos de organização do trabalho colectivo. *Acontecimento sociológico*, enfim, no sentido mais forte e compreensivo do termo: com o impacto na economia geral da comunicação decorrente da multiplicação versátil e integrada das redes de transmissão/comutação de informação que inter-correm imbricadamente desde a macro-dimensão planetária à micro-dimensão dos lares e dos actos da vida quotidiana, o que se altera não são apenas os parâmetros do modo de produção em sentido técnico-económico mas todos os lugares e todo o tecido da unificação social.

Em segundo lugar, deve ter-se em conta o significado do impressionante aumento de *escala* que os avanços das TI trouxeram às linhas de força mais influentes da "sociedade neotécnica". Mesmo que as mudanças apuráveis se reconduzissem apenas a este fenómeno de ampliação estariamos perante uma situação típica de comutação da quantidade em nova qualidade. Isto é tanto mais assim quanto a intensificação dos efeitos passa por uma compressão drástica dos dois constrangimentos genuinamente *fundamentais* dos intercâmbios humanos: o espaço e o tempo. Os cenários de "instant universal information", já tecnicamente actuantes, ao articularem na mesma teia "nervosa" todos os níveis macro e micro da interacção humana abrem perspectivas de transformação das estruturas e dos processos sociais em que a única coisa hoje certa é a amplitude planetária da mudança; quanto ao sentido dos efeitos deve reconhecer-se que se mantém ainda muito incerto, até porque tudo continua em fase de explosão. Por este ângulo, o ineditismo histórico das questões parece incontestável.

Como incontestáveis na sua relevância e na novidade dos seus problemas são os consecutivos fenómenos, a que já se aludiu ao de leve nesta apresentação, da *globalização planetária dos mercados* e da *transnacionalização das actividades económicas*, seja nos actos da produção, da circulação e do consumo, seja no levantamento e na gestão dos recursos terrenos, quando não espaciais (o que se cruza com os primeiros florescimentos de uma "ecologia transnacional"). A globalização mundial dos mercados é uma das

expressões da "global dimension" que parece caracterizar a "era da informação". Os saltos em frente num efectivo sistema mundial de telecomunicações que o progresso integrado das TI tem permitido criaram a base operativa para que se possa falar de "global market" no mundo dos negócios em sentido análogo ao que McLuhan popularizou com a noção de "global village" no mundo da comunicação.

Quanto ao fenómeno de *transnacionalização* da actividade económica, convém atender a que sinaliza uma realidade qualitativamente diferente do bem conhecido e muito estudado fenómeno de internacionalização das forças produtivas que marcou a *praxis* económica sobretudo desde o pós guerra. Ao passo que a internacionalização desenha um processo de complementaridades e dependências dos espaços nacionais, que não abole a importância destes como referencial de pertinência estratégica, o cenário da transnacionalização secundariza drasticamente as territorialidades nacionais (políticas, jurídicas, económicas, culturais) que, continuando a representar instâncias significativas, perdem, no entanto, a relevância de referencial estratégico principal. As estratégias de mundialização (assumidas hoje não apenas pelo actor-empresa: há poderes políticos transnacionais de nova espécie a emergir, assim como poderes corporativos, universitários, de investigação científica e tecnológica, de protecção do meio ambiente, etc.) passam, literalmente, por cima das fronteiras e, uma vez mais, as redes mundiais de comunicação são o pressuposto infra-estrutural do processo.

Finalmente, algo que enforma toda a situação actual dos assuntos humanos: o grau superlativo introduzido pelas TI na simbiose interactiva entre *sociedade, ciência* e *tecnologia*. Mercê da sua versatilidade polimórfica, as novas TI tomaram, como já referi, um papel central e indutor neste processo *integrado* que passou a confundir-se com a vertebração constitutiva das formações sociais contemporâneas.

E digo 'passou' porque, se é certo que as técnicas humanas sempre andaram intrinsecamente associadas aos processos sociais num entrecruzamento tão antigo como o próprio homem, a situação actual de fusão orgânica entre as formas da interacção humana, a inovação científica e a tecnologia configura um tipo de interpenetração constitutiva que se não observa em estádios históricos anteriores.

Nestes, os inter-relacionamentos das técnicas e das morfologias sociais deixam-se ainda analisar mediante quadros de referência de dupla entrada do tipo "efeitos sociais da tecnologia" / "efeitos da envolvente social no desenvolvimento científico-tecnológico". As correspondentes grelhas explicativas reportam-se a uma situação geral em que os produtos tecnológicos podiam ser pensados como criações técnicas *contrapostas* às estruturas sociais, num posicionamento relativo de exterioridade portanto, embora ambas as esferas de realidade comportassem virtualidades de intercepção recíproca. Porém, as intercepções não ocorriam no contexto de um processo integrado. Nem o desenvolvimento científico se articulava organicamente com o das técnicas (pelo contrário, detectam-se situações históricas de assimetria bloqueante entre o nível cognoscitivo da ciência e o potencial operativo das técnicas), nem o processo social apelava à ciência e às tecnologias como forças *principais* de produção e mudança. Daí, a verificação de *descontinuidades* de fundo entre progresso científico, progresso técnico e progresso social[11].

Como o advento do sistema produtivo industrial da era contemporânea, este cenário começou a acusar modificações verdadeiramente estruturais. O fenómeno de *centralização* de todo o processo da interacção sobre o *pivot* constituído pela função produtiva e seus aparelhos tende a fundir incindivelmente as esferas da sociabilidade, da inovação científica e da instrumentalidade técnica. Entre relações sociais, investigação científica e engenharia técnica foi-se tecendo uma rede de vasos comunicantes e inter-activamente propulsores;

a sintese resultante traduz-se em que as formas e os modos da produção e circulação dos bens da vida passaram a absorver os desenvolvimentos científico-técnicos como variável *endógena* e *fundamental* de um mesmo sistema integrado, no qual as orientações da ciência se reconduzem a objectivos de produção e os objectivos da produção são estruturados pelo estádio do desenvolvimento científico-tecnológico, o que tudo pressupõe e ao mesmo tempo estimula uma incessante interpenetração do trabalho científico e da criatividade técnica.

Ao potencial dinâmico desta simbiose, a um tempo tripla e una, na expressão de Cotta (1971), fica bem, como propõe o mesmo autor, a designação de "energia tecnológica", e, com efeito, o que pode estar em causa, presentemente, na tríplice relação sociedade/ciência/técnica não é apenas, como no passado, a aptidão ou inaptidão social de dados avanços no conhecimento científico ou de dados constructos técnicos mas todo um processo objectivamente auto-regulado na sua lógica de reprodução expansiva.

A explosão "informacional" dos nossos dias garante ao cenário "neotécnico" o ambiente perfeitamente adequado à plena explicitação da essência cibernética latente *ab ovo* no sistema. Obtiveram-se os meios tecnológicos para se aceder a um estádio de profundidade no enraizamento social, de versatilidade ramificada nas aplicações e de banalização nos usos que vai a caminho de abolir, por integração simbiótica, as últimas linhas de fractura entre morfologia social e morfologia tecno-lógica.

Pode dizer-se que na "era informacional" *o efeito social primário das tecnologias consiste na sua própria existência*. O que chamamos hoje tecnologias — nomeadamente as TI — são em rigor *sistemas* tecno-económicos altamente complexos ("a technological system is defined as a dynamic network of agents interacting in a specific economic/industrial area under a particular institutional infrastructure and involved in the generation, diffusion, and utilization of technology": Carlsson & Stankiewicz, 1991, p. 93). De facto, assim como as sociedades se vão "tecno-logicizando" progressivamente (e, segundo alguns, perversamente), as arquitecturas técnicas que suportam o processo apresentam-se, de ponta a ponta, como fenómenos intrinsecamente sociais em si mesmos. Se tem algum sentido falar-se de algo como "sociedade tecnológica" isso implica, no reverso da medalha, tecnologias por essência sociais; como resume Lyon (1990, p. viii) "new technology is as much a social product as the shape of society is a technological product". Ao falar de inovação científica hoje estamos a falar cada vez menos de proezas individuais de cientistas e cada vez mais de uma combinatória de conexões interactivas ligando unidades de trabalho científico colectivo, novas prioridades do investimento, tanto público (nacional e transnacional) como privado, formas de organização da produção susceptíveis de potenciar a "energia tecnológica" que decorre da articulação dos sistemas da investigação, do ensino e das empresas e modelos de interface que assegurem a economia das articulações da produção (oferta de bens "tecno-científicos") e do uso (procura desses bens)[12].

Num cenário como este os casos de descontinuidade reduzem-se a meros erros de cálculo industrial ou comercial que não afectam a lógica geral da situação. Compreende-se que o domínio teorético-prático de semelhante imbricação orgânica requeira bastante mais e sobretudo outra coisa do que uma como que contabilidade de efeitos recíprocos entre a esfera tecnológica e a esfera social, ao modo das contabilidades de custos/benefícios. A exigência passou a ser substantivamente diferente: impõe-se reconstituir cartas de marear para os itinerários de um arranjo global dos assuntos humanos em que as paisagens da sociabilidade, da sensibilidade, da cultura, se vão "tecno-logicizando" *intrinsecamente* com a exacta contrapartida de todo o processo científico-tecnológico se

haver tornado *constitutiva* e *operativamente* social. É um dos grandes desafios do nosso tempo, se não *o desafio*, e a resposta que a humanidade lhe souber dar compromete todas as dimensões do presente, assim como as do futuro das gerações vindouras.

* * *

É evidente que num mundo de fusão íntima e generalizada entre, por um lado, estruturas de vida e estruturas tecnológicas e, por outro, as TI e a evolução tecnológica em geral, o questionamento ético sobre as TI (e, logo, há que não esquecê-lo, sobre as orientações deontológicas do próprio trabalho científico que fundamenta as tecnologias) assume uma importância a bem dizer *decisiva*. Na verdade, a modalidade de questionamento introduzida pela perspectivação ética no debate dos limites das TI abre campos problemáticos particularmente melindrosos, pois enquanto os limites teóricos ou os limites de aptidão racionalizante dos usos jogam com questões de impossibilidade — ou de certas pretensões da teoria, ou de certos desideratos práticos — os limites invocáveis do lado da ética[13] traçam fronteiras de *indesejabilidade* tanto mais inquietante quanto pressupõe que o indesejável é plenamente possível.

Que se levantem embargos éticos à pretensa nocividade, efectiva ou virtual, de efeitos sociais, culturais, ambientais, etc. de determinadas formas ou utilizações de tecnologia, ou ao risco de desastres que possam desencadear, não é uma atitude de espírito inédita. Como se sabe, a noção de que as criações da técnica transportam uma irredutível carga de ambivalência positiva e negativa nos efeitos, bem ilustrada pelo lugar comum metafórico do aprendiz de feiticeiro, comparece recorrentemente nas experiências da consciência humana. Reencontramos aqui, no entanto, como problema, esse sim inédito, dos nossos tempos, o da *escala*: agora, a da nocividade e a das ameaças, e a questão do nuclear, tão viva, justificadamente, nos debates ético-políticos da contemporaneidade, aí está, como situação limite, em absoluto esmagadora, dos recortes novos adquiridos por antigas inquietações. Escusado seria acrescentar que o papel fulcral das TI em toda a infra-estruturação tecnológica das actividades humanas, tanto as de paz como as de policiamento e guerra, aberta ou silenciosa, as coloca no centro deste tipo de questionamentos.

É sintomático que as reflexões de Warfield *(infra)* sobre o perigo de que a tecnologia decline como força de progresso social partam do significado ético da escala: quando os sistemas de base tecnológica alcançam um patamar de disseminação espacial totalizante e condensam a velocidade de eclosão e propagação de efeitos no grau explosivo do instantâneo, o "hazard gap", que sempre existiu, entre a implantação de sistemas de alto risco e o controlo científico do risco ganhou um potencial dramático obviamente agudizado. Os poderes científico-tecnológicos reuniram, como todos sabemos e alguns indivíduos e grupos humanos já experimentaram, capacidades de destruição em que verdadeiramente alucinante, mais do que aquilo se sabe delas, é o que se pressente. Com a prodigiosa aceleração que "the advanced computing for science" trouxe ao processo dos conhecimentos *em geral* — principalmente (mas não exclusivamente: as vantagens irradiam em múltiplas direcções) mercê das possibilidades novas de se pôr à prova as consequências de teorias simples em situações de grande complexidade (vd. a propósito, da Silva, *infra*) — estimula-se a euforia tecno-cientificista a que Warfield chama "Artificial Philosophy" e introduz-se um formidável efeito de multiplicador não só de vantagens como, paralelamente, no paiol de perigos calamitosos disseminados pelo mundo dos homens.

Se ao efeito de escala somarmos, como risco, a *irreversibilidade* dos danos o panorama tinge-se de cores de susto ainda mais carregadas. Para além da ameaça de descalabros

civilizacionais e ambientais em rigor inestimáveis, a humanidade enfrenta hoje — e é a primeira vez que o enfrenta *a sério* — o risco *efectivo* de a sua "energia tecnológica" poder lesionar irrecuperavelmente patrimónios genéticos de formação multi-milenária, quer a destruição revista a espectacularidade sinistra de Hiroshima ou de Chernobyl, quer tome os micro-contornos silenciosos que estão latentes em desenvolvimentos das biotecnologias sem controlo ético adequado.

Uma ética "cosmopolítica", para o dizermos como Kant, *deve* compreender, nos tempos que correm, o imperativo categórico de, sem cultivar a insensatez de querer parar o progresso técnico ou despejar o bébé com a água do banho, saber meter travões nos modos de fazer face aos problemas humanos que não podemos dominar com segurança; designadamente, deveríamos dispensarmo-nos de tomar decisões que empurrem a humanidade para caminhos sem regresso. Acrescentar à soma dos conhecimentos eficazes disponíveis a sageza de tornar exequível a tarefa das gerações futuras preservando-lhes opções é, talvez, o máximo e o mínimo que qualquer geração humana pode fazer pela sua descendência, como comentou recentemente Simon (1990, p. 156; neste sentido, cfr. também os estudos fundamentais de Jonas, 1984).

O que, dir-se-ia, parece óbvio. Tão óbvio que a assunção, quaisquer que sejam as vantagens esperáveis, de certos riscos de desastre para comunidades, para regiões, para povos, para as próprias condições de sobrevivência da espécie, antes de ofender a exigência ética ofende o senso comum. *Porém, são assumidos*, e o que mais nos deve preocupar é isso mesmo. Ante a notável teimosia humana em apostas absurdas segundo as luzes mais comuns do entendimento, o impressionante — i.e. o que merece ser interrogado — consiste não tanto na demonstração do absurdo, *mas em que possa haver reiteradamente opções estratégicas pelo absurdo*.

Warfield aponta o Terceiro Reich como exemplo, entre outros, de um "invasive system" (sistemas de forte sustentação tecnológica susceptíveis de ferir desastrosamente, em larga escala, as vidas, os bens e os valores de massas consideráveis de pessoas que não deram assentimento ao impacto calamitoso sofrido). O exemplo é oportuno mas passa-se talvez por alto o seu traço mais significativo para uma reflexão necessária não já sobre os limites éticos da tecnologia, pois o que fica em causa são os limites da própria ética; a saber: que milhões de pessoas — a massiva maioria delas portadora, como todos nós, das faculdades do senso comum que são, dos bens da terra, o melhor repartido, muitas inteligentíssimas e culturalmente preparadíssimas, algumas exibindo cotas de génio em artes, ciências e filosofia — se tenham empenhado ou deixado empenhar, com convicção de vida ou de morte, naquele projecto de selvajaria "tecnológica" monstruosa.

Embatemos aqui numa dimensão problemática das coisas humanas que não é somente meta-tecnológica; é também *meta-moral*, se tomarmos a moral como uma comunhão universalmente partilhada de normas racionais de vida.

Veja-se, a propósito, a crítica acerada de Meehan *(infra)* à prevalência de modelos decisionistas tecno-lógicos na condução dos assuntos humanos: Meehan fundamenta na sua reflexão crítica a proposta construtiva de que a *desejabilidade* das condições de vida das populações afectadas por uma qualquer decisão deva ser, ao mesmo tempo, a preocupação primária e o critério justificador da escolha; ora, é possível que se não consiga arranjar melhor critério, assim como é provável que a proposta reúna requisitos de anuência intelectual generalizada, e, todavia, parece-me manifesto que serve para *localizar* o problema, não para o resolver. Aliás, Meehan não escamoteia a dificuldade, longe disso; basta atender à inilidível variedade, por ele salientada, dos quadros de referência (sociais, culturais, políticos, de conjuntura, etc.) que definem o quadro circunstancial de cada

decisão para termos de assumir o reconhecimento de que o critério de justificação proposto deixa a questão a pairar sobre um abismo de incertezas (Quem decide da desejabilidade? Com que metodologia política recomendável de tomada de decisões? A que mínimos de objectividade racional se pode apelar? etc.).

No fim de contas, penso que o questionamento ético da tecnologia nos devolve todas as grandes interrogações defrontadas por qualquer modalidade de moral discursiva (fora a questão liminar de se saber se pode haver outro tipo de moral). Com o reenvio, traz-nos ainda o acrescento de um valor de nitidez muito considerável acerca da nervura problemática do que efectivamente está em causa. Evidencia-se aqui, à luz extremamente crua do poder de vida ou de morte para grandes grupos humanos, se não para a inteira humanidade, que é hoje um poder por excelência de alta tecnologia, como as antiquíssimas contradições entre a comunidade ideal do discurso e a comunidade real das interacções humanas, entre a universalidade racional dos consensos do espírito (a ''comunidade instituída das consciências'' de que fala Hegel no prefácio da *Fenomenologia do Espírito*) e a racionalidade estratégica da acção, transcendem o esquema em que correntemente são colocadas que põe a um lado, a razão, e a outro, os desastres do mundo, vista a razão na forma de instância de julgamento *erga omnes* alheia aos desastres que condena.

É que a razão humana não pode pretender postar-se majestaticamente perante a ciência moderna e a tecnologia, que são suas concretizações, como tribunal exterior aos pleitos. Assim como, p.e., a lei de que a má moeda expulsa a boa moeda obriga, para ter sentido, a pressupôr-se o conceito englobante de moeda, o problema dos ''bons'' e ''maus'' usos das faculdades racionais da humanidade passa não por fora mas pelo interior da própria razão, evidenciando, porém, a inquietante propensão de as "más" razões poderem neutralizar as "boas". Ou, dito de outra maneira que significa o mesmo, *a razão faz parte do problema*, no ''universal concreto'' do complexo de manifestações que dão forma e conteúdo, dentro do mundo dos homens, à luta porfiada para sobrepôr a paixão da razão às obras irracionais que da razão se geram.

Digamos que tal como se não pode esperar da tecnologia que resolva as questões meta-tecnológicas da própria tecnologia — o que é uma velha ilusão que a tipologia "inteligente" das TI tem feito prosperar — também se não pode confiar na transparência ideal do discurso da pura razão para fazer vingar a razão das razões humanas contra as razões, quase sempre, se não sempre, muito fortes, da irrazão. A opção pelas possibilidades racionais da razão coenvolve um interesse da vontade, um pré-conceito a favor da razão que deve contemplar, como adversário a combater, a contra-possibilidade de a razão se mostrar irracional.

Hoje como sempre a opção decide-se na cidade. Aí, e só aí, auferem sentido e não-sentido os módulos racionais e irracionais de realidade da razão, aí, e só aí, se joga a tensão, permanentemente em aberto, de um princípio de esperança no futuro do homem e das suas criações, pois (soube dizê-lo admiravelmente esse desesperado filósofo da esperança que foi Ernst Bloch) a razão não pode florescer sem esperança mas a esperança não pode ser dita se faltar a razão, nenhuma outra ciência senão esta pode ter futuro, nenhum futuro outra ciência. Trata-se de contradições sobre as quais qualquer acção "civicamente" útil deve atender a uma dupla perspectiva de abordagem: nem as tecnologias, de informação ou outras (mas sobrarão outras que não estejam em vias de se tornar "informacionais"?) podem constituir, por si, quadros de solução para os problemas, nem há soluções idóneas em que se possa prescindir de saber fazer contas com as tecnologias. É que a cidade mudou de qualidade e de escala. As ambivalências da razão passaram a apoiar-se numa nova ''táctica vital'', de extremo a extremo "tecnológica", cujo travejamento operativo envolve

todo o planeta e vai a caminho de alargar as redes de "inteligência colectiva" da espécie até dimensões transplanetárias. Isto que já não é "science fiction" está a requerer formas renovadas de razão económica, de razão jurídica, de razão política na difícil gestão global da cooperação e dos conflitos, e importaria sobremaneira que o necessário controlo ético do processo se não esgotasse em exercícios de "moral fiction". Ora receio bem que haja muito de ficção em alguns moralismos correntes do tipo condenatório alegados contra a "essência da técnica", ainda quando se mostrem justificáveis: se confiar nas competências próprias da tecnologia e dos tecnologos para, como tais, resolverem os problemas de meta-tecnologia surgidos dos sistemas tecnológicos se perfila sempre como um caso limite de *ignoratio elenchi* — a típica ilusão tecnocrática — até porque, além de outras razões, um problema é problema justamente quando lidamos com as situações anti-tecnológicas por excelência de se não dispôr de toda a informação para agir e, no entanto, se ter de agir[14], não é menos ilusório esperar que os homens renunciem às suas técnicas, mesmo que em situações altamente problemáticas, por efeito apenas de exortações de natureza moral, pois, observou-o Marx na célebre carta a Paulo Annenkov, parece ser sestro da humanidade nunca se desfazer dos seus adquiridos históricos de *know-how*. Para que os poderes discursivos da moral ao nível da comunidade ideal de discurso se tornem persuasivos ao nível da comunidade real da acção impõe-se que a moral se faça ética. I.e.: *que impregne a interacção humana em vez de a querer morigerar*. O que, no contexto tecnológico do mundo nosso contemporâneo, implica *eticizar* a ciência e a tecnologia *por dentro*, nos seus planos de criação/aplicação, e não propriamente querer moralizá-las de fora.

Simplesmente, esta exigência, para ter êxito, limitado que seja, demanda o trato com algo mais do que a dimensão comunicacional do discurso: demanda a mobilização, no campo de operações da *polis*, dos argumentos, dos objectivos, das armas de luta pela plena cidadania que é um dos sinónimos de luta pela razão.

Para esse esforço — arriscado, vigilante, obrigatoriamente paciente; mas como esperar que pudesse ser de outra maneira? — as novas TI podem contribuir com mais do que ameaças. *Podem* abrir, também, um arsenal de oportunidades favoráveis, e a escala é a mesma que exibem as ameaças: gigantesca.

Sublinha-se 'podem' porque o *quid* problemático está aí. Não há fatalidades neste domínio, sejam tingidas de rosa, sejam tingidas de negro. As "tácticas vitais" da técnica nunca são neutras nem quanto às formas nem quanto aos conteúdos da acção, mas não são tão potentes que absorvam todo o sentido das estratégias normativas.

Por exemplo, os meios de uma "direct on line democracy", baseada na tecnologia das novas redes de inter-acção comunicativa, são, como o mostra alguma experiência já invocável, vulneráveis a uma gama diversíssima de astúcias, de distorções, de efeitos perversos, e nunca será demais a denúncia, politicamente vigorosa, dos embustes que viabilizam e das fraudes que podem travestir; porém, *abrem*, do mesmo passo, um horizonte de afirmação de democracia que, sem eles, seria absolutamente impensável (cfr. sobre este tópico Mclean, 1989).

A questão, de resto, é mais uma das que não são inteiramente novas, e já que de técnicas de informação falamos pode-se lembrar que a democracia de discurso directo dos atenienses, assente sobre a fascinante proeminência ganha pela palavra como tecido conjuntivo do poder e dos poderes no "universo espiritual da polis" (cfr. as sínteses magistrais de Vernant, 1975), serviu, por sinal em um processo de impecável regularidade democrática, para se condenar à morte o paradigma da razão chamado Sócrates; mas terá servido *somente* para isso? Terá servido *principalmente* para isso? Sem dúvida que não; em todo o caso, serviu *também* para isso.

Reencontramos nestes envolvimentos do tema a possibilidade, de que se tem fabricado a história atribulada dos percursos da razão humana, de a razão se virar contra si própria e a malha das contradições revela-se bem mais densa, bem mais problemática do que uma oposição trivial entre a "razão dos valores" e a "razão dos procedimentos". A razão tecnológica não é excepção, e nisso se enganam os inevitáveis vendedores da ilusão tecnocrática; mas a luta não se pode ganhar contra a tecnologia, repito-me, ganhar-se-á, na medida em que for ganhável, no interior da própria tecnologia, como processo social que também é, desde a criação científica até aos desenvolvimentos de aplicação, e nisso se enganam os moralismos de recusa dos adquiridos tecnológicos. O espaço da nova *polis* "tecnológica" não é, não deve ser, não pode ser um espaço de neutralização da ética política pela tecnologia ou de neutralização da tecnologia por cominações éticas. Pelo contrário: a própria vertebração tecnológica do inteiro corpo social, com os "hazard gaps" de toda a ordem que a nova escala dos sistemas implica mas também com os "venture gaps" que paralelamente ficam em aberto, torna instante que se *politize* em profundidade e com conhecimento de causa tecnológico a supervisão ética dos processos globais activados hoje pela tecnologia. ('Politize' no sentido de se assegurar efectividade política aos controlos éticos; e este tem de ser, obrigatoriamente, um dos campos de renovado aprofundamento de uma política democrática, sob pena de declínio não da tecnologia mas da democracia).

Uma ética dos usos da tecnologia nas sociedades contemporâneas não é só problemas para a moral o que em definitivo recorta mas sobretudo problemas da e para a política, no sentido forte e valorizado do termo, como empreendimento de cidadania. A necessidade, realmente imperiosa, de que os indivíduos e os grupos humanos aprendam a viver com as suas máquinas revela-se, na "information age", mais do que nunca isomorfa da aprendizagem de saberem viver uns com os outros, e isso, como já alguém disse, não significa orar nas catacumbas mas preservar atributos de humanidade no meio dos arranha-céus.

Entretanto, podemos esquecer, sem perda, alguns fantasmas típicos da era informática, o principal dos quais, provavelmente, vai continuar a ser o receio sado-masoquista de que os indivíduos mecânicos "inteligentes" substituam, em geral, os agentes humanos no trabalho e na tomada de decisões[15].

No que a trabalhar concerne lamentamos não poder confirmar a boa notícia que seria as máquinas inteligentes mostrarem aptidão de nos dispensar, de vez, das penas do trabalho necessário. Não que seja despiciendo o seu impacto no mundo do trabalho humano. Como se mencionou atrás, de passagem, o impacto é já estremecedor e não parece desrazoável antever efeitos de abalo sísmico agravado num futuro próximo, isto qualquer que seja o ângulo de consideração: o do volume do emprego, intimamente correlacionado, como sempre, com o dos níveis de produtividade, o das qualificações profissionais relativas, em que se imbricam exigências de reforma radical nas estruturas e nos critérios, socialmente tão sensíveis, da educação e da formação, o dos parâmetros no arranjo óptimo das organizações complexas da produção, da comercialização e da administração, por onde passa o essencial das questões de economia do poder nas sociedades modernas, o da retroversão sociológica destas mudanças e seu embate nas formas tradicionais de representação político-sindical dos interesses das populações laborais, etc., etc. (dentre uma literatura especializada em vias de se tornar inesgotável, pode consultar-se, com proveito, os capítulos sinópticos sobre esta matéria em Lyon, 1990, e Laver, 1989). Porém, não me parece que haja no cenário indícios a apontar para a abolição geral do próprio trabalho. Do que podemos ter uma certeza razoável é de que a generalização do uso de dispositivos

mecânicos "inteligentes" nos vai compelir a trabalhar de maneiras radicalmente diferentes e que o nível da diferença penetrará, a fundo, em todos os estamentos (operativos, sociais, jurídicos, económicos) da relação de trabalho.

Quanto ao risco da hegemonia de "decisores" mecânicos na gestão dos negócios humanos, o que já se disse aqui e o que se vai poder ler adiante neste livro, justifica o ponto de vista de que aquilo que nos deve inquietar, e muito, não são propriamente as "decisões" de máquinas mas as provindas do tipo de homens que decidem deixar que haja sistemas maquinais a "decidir" (cfr. Meehan, *infra*). Se se conseguir contrariar o poder de mando destes perigosos espécimes de humanidade — i.e. se se conseguir controlar os mecanismos sociais que os forjam e em que se forja o seu perigo — as pobres máquinas "lógicas" podem ficar em paz, a prestar-nos serviços de maior ou menor utilidade.

Que elas, afinal, as máquinas, estão sempre em paz, na sua rasa inocência de metal, *chips* e plástico. Se não têm o dom de nos presentear com a *password* de acesso à idade de oiro da humanidade, igualmente não dispõem de tanto poder maléfico como para desculpabilizar os actores humanos, individuais e colectivos, da insensatez, do crime, da injustiça, da violência.

Postos perante as esperanças e as ameaças do potencial tecnológico genuinamente novo que significam as modernas TI, no seu cruzamento orgânico com todas as metamorfoses da sociabilidade, os homens deste nosso tempo em que somos chamados a estar aprendem, uma vez mais, que, na história da sua caminhada no planeta, se a mudança é uma certeza, o progresso na mudança tem de ser uma empresa.

Não há deuses que nos salvem nem demónios que nos percam. Nas apostas de esperança ou desespero a humanidade é de si para consigo que tem de se haver. Ganhar ou perder as apostas caberá à responsabilidade individual e colectiva dos homens com as *suas* máquinas, e suas por serem criações da *sua* razão, como o são as idéias, as crenças, os valores pelos quais vivem, pelos quais morrem, pelos quais matam. Mesmo que desejássemos o contrário, as máquinas sem nós por isto não respondem.

Lisboa, Maio de 1991.

O meu agradecimento a Marcelo Dascal, cujas críticas a uma primeira versão deste texto contribuiram para se esbaterem alguns dos defeitos de fundo e de forma. Os que subsistem são-me imputáveis em exclusivo.

Notas

(1) Cfr. Coelho (1986, pp. 41-2). Numa definição escolarmente simples pode dizer-se que as TI são as tecnologias de recolha, armazenamento, recuperação, processamento e comunicação de informação sob forma de dados, texto, gráficos, imagem e/ou voz.
Anote-se que a identificação genérica dos novos tratamentos tecnológicos da informação pela expressão "Information Technology" <IT> que, com fonte sobretudo na literatura britânica, foi tendo correspondência directa noutras línguas, não se encontra ainda de todo estabilizada, subsistindo alguma flutuação terminológica de país para país (p.e. nos EUA são utilizadas também, com frequência, as designações mais sectorizadas de "computing technology" e "communication technology", assim como "informatique" e "telematique" na literatura em língua francesa; cfr. a propósito a "note on terms" em Lyon, 1990).

(2) Novidade apenas quanto aos contornos específicos das actuais TI, pois a importância histórica que sempre tiveram as inovações técnicas de tratamento e registo da informação como marcos de "saltos de civilização" (invenção dos caracteres de escrita, do papel, da imprensa, etc.) é,

de há muito, reconhecida e aceite.

Aliás, o grau superior de complexidade, sob todos os aspectos do termo, designadamente os sociais, que o actual contexto do processamento da informação alcançou, tem contribuído, com luzes de reforçada compreensibilidade retrospectiva, para o entendimento da relevância crucial, em toda a estruturação das sociedades, dos estádios precedentes das técnicas de recolha, preservação e transmissão de informações (cfr. Poster, 1988, pp. 6 ss.). Comprova-se aqui o princípio explicativo de que, nos percursos do conhecimento, é o patamar de maior complexidade num dado processo que permite, em retroacção, esclarecer os sentidos do mais simples.

(3) Na célebre entrevista a *Der Spiegel*, concedida em 1966, com publicação póstuma só 10 anos depois, por exigência expressa do filósofo (vd. Heidegger, 1989). Aqui fez Heidegger a afirmação surpreendente de que o antigo posto da Filosofia passou a ser ocupado pela Cibernética.

(4) O que não deixa de restringir consideravelmente a amplitude ecuménica actual do fenómeno. Convém, no entanto, não pôr demasiada ênfase na restritiva. Se for verdade que a pretensa "information age" designa um novo modo dominante de organização dos relacionamentos humanos, o atributo de domínio nem envolve um carácter de exclusividade (a co-existência de sedimentações histórico-sociais várias é uma constante dos mapas da interacção humana) nem, à partida, significa uma presença maioritária, mas tão-só a detecção de linhas de força em processo de desenvolvimento. Como comenta Poster, "industrial capitalism was when Marx theorized it both a minority phenomenon and the outcome of a long historical trajectory; the same emergent character may be attributed today to the mode of information" (*ibid.*, p. 20).

(5) Realmente, a polémica à volta da reivindicação de atributos de efectiva inteligência em sistemas computacionais propendeu sempre, sem prejuízo de acentuadas especialidades no argumento, a incluir-se em um horizonte comum de referências ideológicas a bem dizer temático da cultura ocidental: o de que o traço de fronteira da "essência" da humanidade passa por aquelas faculdades racionais da "alma" que Aristóteles identificava com o *telos* específico dos entes humanos e que Descartes teorizou como a *res cogitans* cuja vinculação a dadas extensões corpóreas se revestiria de alcance meramente contingente (cfr. *infra* as indicações de Dascal e de Polakow, sobre este tópico).

Donde que, não obstante os caminhos de progressivo refinamento percorridos pela teoria pós Descartes, se continui a ligar as competências linguísticas e lógico-matemáticas próprias dos indivíduos humanos ao recorte da racionalidade em geral e à fundação de valores ético-morais, de que tudo, por definição, estariam excluídas as esferas de existência animal e mecânica. Donde, também, o sopro perturbador introduzido neste cenário pelo advento de indivíduos mecânicos com toda a aparência de "inteligentes" e muitas das resistências intelectuais opostas ao reconhecimento da genuinidade de tal atributo. Duvido que, ainda hoje, o debate se processe com descontaminação das sobrecargas ideológicas obviamente inerentes a semelhante quadro, embora um dos aspectos mais interessantes da conceptualização do problema em Polakow (*infra*) consista, justamente, no intento de contrariar a legitimidade teórica da adjudicação de verdadeira "inteligência" aos artefactos computacionais, mas "without any repeal to humaneness".

Este fenómeno de rejeição que, como comenta p.e. Pereira (1988), teria mais do que uma analogia com a exasperação dos confrontos ideológicos associados a outros derrubes científicos de bastiões da "essencialidade" humana no mundo (lembre-se os casos limite de Copernico/Galileu, de Darwin, de Freud), foi espicaçado pelas reivindicações maximalistas de alguns teóricos da IA que vêm na sua disciplina, mais do que a teoria de uma certa forma de inteligência (o que já seria muito), o laboratório da pura inteligência "em-si" e "para-si" (o que é desmesuradamente muitíssimo). Até porque daqui a prospectivas cósmicas, ao modo de Stonier, para quem a "machine intelligence", que não conheceria os limites de evolução

da mente humana, jungida ao seu *hardware* cerebral, "will outpace human intelligence and very likely will do so during the lifetime of our children", o que "will create an evolutionary discontinuity as profound as the origin of life" e "will presage the end of human species as we know it" (1988, p. 133), o passo é mínimo.

O debate, relançado recentemente por Penrose, em defesa de pontos de vista "humanistas", foi ganhando densidade e dimensões teóricas que excluem qualquer tratamento útil ao nível de uma brevíssima nota como a presente. Ouso, no entanto, sugerir algumas cautelas de senso comum na aproximação ao tema que, não o resolvendo, evidentemente, nem chegando sequer a demarcar os seus zonamentos problemáticos, talvez ajudem a não se perder o pé em águas de excessiva emotividade.

Em primeiro lugar convirá atender a que uma coisa é sinalizar-se e aprofundar-se teoricamente a diferença entre os mecanismos psico-biológicos discerníveis no modo de funcionamento da mente humana e a lógica dos dispositivos de produção de efeitos de inteligência por máquinas computacionais, outra é inferir-se, da diferença, que estes são obtusos ou estúpidos apenas porque se produzem de maneira substancialmente distinta daqueles. A menos, claro está, que se pressuponha, de entrada, que só o tipo de inteligência dos humanos é inteligente e todos os mais fenómenos de resposta inteligente a estímulos da realidade configuram, no fim de contas, desde que provindos de máquinas (de animais também?), meros simulacros que escondem estupidez ontologicamente insanável; mas então o sofisma do argumento "essencialista" fica demasiadamente a descoberto.

Conta-se que o legendário campeão de xadrês Robert Fischer, após jogar uma partida contra um programa de computador e, naturalmente, a ganhar, teria comentado: "a única diferença verdadeiramente de fundo que vejo entre mim e o computador é que eu sei que estou a jogar xadrês". Fidedigno ou não o episódio, a observação mostra-se fulgurantemente certeira: identifica com precisão o que parece ser um dos atributos específicos do tipo de inteligência natural humana (a auto-consciência reflexiva do sentido da acção) mas não nega por isso a especificidade da inteligência da máquina, no exercício restrito para que foi concebida (jogar xadrês, no caso de espécie). Julgo que um dos aspectos mais profícuos na investigação de Sloman irradia, precisamente, da sua idéia de um espaço *pluri-dimensional* de "possible minds", subtraindo, portanto, a noção do "mental" às armadilhas essencialistas (embora seja discutível, no caso de Sloman, se o seu conceito de "mechanism" não reintroduzirá, pela porta das trazeiras, uma nova "essência" no debate). É que fica entreaberta, assim, uma perspectiva de pluralidade de entes mentais heterógeneos, sem pretensões de hegemonia isomórfica nem dos contextos humanos por contextos mecânicos, nem de contextos mecânicos por contextos humanos. Torrance (1984, p. 25) conclui daqui, e penso que proveitosamente para um entendimento sereno da questão, que "so there's surely a lot of point in seeing computers as the source, not so much of mock-mental abilities, mechanical analogous for, and poor substitutes for the real things, *as of new forms of the real thing*" (sublinhado meu).

Aliás, parece que as arquitecturas de IA só são em verdade úteis não desde que se pretenda que "pensem" como os agentes humanos pensam, mas, pelo contrário, se e na medida em que *não pensem* como e naquilo o que e em que o pensamento humano pode actuar directamente. Se "pensassem" (fizessem) apenas aquilo que agentes humanos pensam (fazem) da mesma maneira, não se vê bem para que serviriam tais máquinas (excepto talvez como artefactos de jogo, o que, não deixando de ser um escopo lúdico perfeitamente respeitável, não alcançaria, com certeza, a provocar tanta controvérsia). Fabrica-se (e vende-se) este tipo de máquinas sobretudo quando podem intervir "inteligentemente" (se o advérbio ofende, podemos dizer p.e. "eficazmente") em domínios de acção inacessíveis ao pensamento humano, por razões que nada têm de misterioso, ou em que não há interesse económico que o agente humano continui a trabalhar, porque o faria pior, ou faria mais caro (o que, segundo a lógica social dominante, vem a dar no mesmo).

De resto, o desiderato característico de qualquer máquina, ou, mais em geral, de qualquer

instrumento utilizado pelos homens, nunca foi propriamente o de duplicar capacidades de acção ao alcance de agentes humanos, mas, sim, o de as ampliar, criando margens de manobra novas em termos de operacionalidade ou de menor dispêndio de recursos. Enquanto esteve em causa poupar bolhas nas mãos ou explorar *performances* mecânicas baseadas no potencial físico, não houve, por este lado, escândalos intelectuais de maior; qualquer que seja o ácume do criticismo que rodeou as "revoluções industriais", não consta que os operários invejassem a superioridade da força do vapor (os seus motivos de queixa eram outros) ou que os protagonistas da era romântica da aviação se lastimassem por não conseguir voar como os seus aviões, assim como não é plausível que os guerreiros de Anibal se sentissem humanamente desqualificados perante as patentes vantagens de operação dos elefantes no cumprimento de dadas obrigações militares. Porém, quando, mercê das máquinas computacionais, se passa do trabalho manual para o trabalho intelectual, interferindo-se nos domínios "nobres" da subjectividade e do conhecimento, sendo justamente isto o que as caracteriza enquanto tipologia de mecanismo diferente de todos os outros e podendo mesmo, sem abuso, carregar-se o computador com a energia poética do inolvidável HAL, de Stanley Kubrick, está-se a mexer em algumas crenças fundadoras da filosofia ocidental, e daí que as contestações mais veementes ao "mecanismo" provenham dos meios da *inteligentsia*. Não deixa de ser curioso, e talvez novo no historial das atitudes humanas perante a máquina, que a recepção da maquinaria "inteligente" encontre as principais dificuldades ideológicas junto de intelectuais, já que os homens e as mulheres que um tique de linguagem eivado de paternalismos de *clerc* denomina de 'comuns' a usam sem grandes inquietações anímicas.

Justamente porque é indiscutível a seriedade dos problemas teóricos e práticos que a IA polariza, penso que a investigação só terá a ganhar se as *performances* de "inteligência" naturalmente fora da alçada do cérebro humano que as máquinas criadas pelo engenho inventivo dos homens realizam passarem a ser vistas com a mesma isenção serena como se assiste às proezas de força de uma grúa que nenhum exército de trabalhadores humanos poderá igualar. Afinal, é precisamente por isso que concebemos e construímos tanto grúas como computadores.

(6) A economia, em todas as dimensões da "Macroeconomics" e da "Microeconomics", está a ser um dos domínios em que se comprova, com grande poder de visibilidade, a observação de Varela acerca de novas experiências "tangíveis" na interacção, materializada pelos avanços das TI, entre transformações no conhecimento e transformações nas práticas sociais. Com os extraordinários progressos dos sistemas de telecomunicações baseados nos desenvolvimentos da tecnologia digital de comutação e transmissão, das comunicações por satélite e da capacidade, praticamente ilimitada, dos novos materiais condutores do trânsito de informação (proporcionando, no seu conjunto, redes digitais com integração e multiplicidade de serviços), é toda uma nova "formatação" das actividades económicas que está em curso. Mas mesmo antes da explosão mais recente das TI e dos seus impactos nas actividades económicas, o núcleo cibernético por excelência do mercado polarizara cruzamentos de investigação altamente frutuosos entre teorias da computação, da decisão racional e da economia da informação (cfr. p.e. Simon, 1981, e Arrow, 1983).

Presentemente, o reconhecimento no terreno de que "markets are no longer places, but electronic networks" (Stoet) e que, como já asseverou um especialista da envergadura de Walter Wriston, ex-presidente da Citicorp, a actividade bancária passou a ter por objecto, não dinheiro, mas "o valor temporal da informação" (*apud* cit. em Liscia, 1989), traz reforços muito "tangíveis" à idéia de que o mercado-instituição social (no mesmo sentido em que dizemos da moeda que é uma instituição) funciona, no essencial da sua "lógica", como um mecanismo de comunicação inter-activa de informações entre pontos dispersos, cuja propriedade mais notável "é o modo como reduz e localiza as necessidades informacionais e computacionais" (Simon, 1981, p. 76). Produz-se também nesta área um grande potencial de iluminação retrospectiva do sentido de experiências do passado por efeito de cristalizações

de maior complexidade no presente.

(7) Reporto a fronteira de periodicização a Hegel e a Saint-Simon (poderia ter acrescentado o contemporâneo Augusto Comte) por se tratar de pensadores que fazem já contas de enfrentamento directo com os fenómenos tecnico-económico-sociais da grande indústria, mas isso não desvaloriza o interesse das intuições genialmente premonitórias de um Francis Bacon ou de um Descartes, nem os traços de actualidade problemática que guarda muita da reflexão dos gregos sobre a *techné*.

(8) Pode-se ilustrar esta dialéctica com o exemplo da emergência da agricultura: os homens terão começado por destinar excedentes de energia e recursos à tarefa de lançar sementes à terra tendo em vista primariamente a melhor satisfação da finalidade de comer; mas, ao mesmo tempo que alimentos, germinará deste trabalho (das suas rotinas, da relação com os seus instrumentos, do prazer dos seus achados) uma finalidade "espiritual" nova, e, em fim de percurso, o homem descobre-se *agricultor*. A finalidade primária fora transcendida pelo meio de satisfação, o qual se tornou o fim determinante na economia do desejo.

(9) Note-se que em aproximações como a de Spengler ressalta que a técnica não é um fenómeno que seja apanágio da espécie humana mas a expressão genérica da vida animal na luta pela sobrevivência de cada espécie, face à natureza "inanimada" e à natureza "animada" constituída pelas outras espécies. Julgo que se podem encontrar traços desta noção da tecnicidade como "facto zoológico" em todas as teorias da evolução das espécies animais de matriz darwiniana e reencontramo-la, p.e., nas notáveis análises de Leroi-Gourhan (1963) sobre a fraca especialização anatómica do homem, se comparado com os outros animais, e o significado da primeira produção de instrumentos pelo homem paleolítico, interpretados como extensões da polivalência "aberta" do corpo humano, na luta pela vida.

(10) O esquema de civilizações tecnológicas proposto por Mumford (1964) distingue historicamente o estádio "eotécnico" (baseado no recurso ao vento, à água e à lenha como principais fontes de energia), o estádio "paleotécnico" (a fase do carvão, do vapor e do aço) e o estádio "neotécnico" (marcado pela cientificização sistemática das actividades ao nível da vida corrente, pela predominância da electricidade e da petroquímica como fontes de energia e pela criação de novas ligas metálicas e dos plásticos como estruturas materiais). Luke (1990) observa que é ainda possível perspectivar, a partir dos escritos de Mumford, as grandes linhas de um estádio "prototécnico" pré-medieval: como fontes de energia o esforço muscular humano e animal, prolongado pela água e pelo fogo, e uso da pedra e da madeira como matérias primas essenciais ("essentially, prototechnics is a slave labor, stone, and animal and water energy complex", Luke: p. 31).

(11) Exemplos de escola muito instrutivos são os das "blocagens" na evolução do pensamento técnico e científico, com efeitos de decadência regressiva, que se detectam na fase terminal do mundo greco-romano e nos decursos da civilização chinesa e da civilização islâmica. Não obstante se ter alcançado, em qualquer dos casos, níveis avançadíssimos no domínio dos conhecimentos teóricos e práticos necessários a um *take-off* tecnológico sustentado como o que se verificaria na era moderna europeia, acabaram por prevalecer bloqueios de natureza social que laquearam as virtualidades de desenvolvimento técnico-científico que estavam reunidas. Sobre este aspecto na história da antiguidade greco-romana, cfr. Magalhães-Vilhena (1984).

(12) Embora se aceite generalizadamente hoje que a inovação tecnológica é um dos factores determinantes do crescimento económico, os modelos neoclássicos que predominam nas análises do crescimento não conseguem deixar de a tratar na óptica de factor exógeno, o que, como comentam Carlsson & Stankiewicz (1991), empobrece a compreensão das conexões causais que interferem no processo; daí que aqueles autores, colhendo expressamente inspiração nas idéias de Schumpeter, proponham uma aproximação integrada que privilegia, à partida, o papel das "micro units" de articulação entre inovação tecnológica e iniciativa empresarial, passando depois à análise das junções que ligam as redes de micro-

relacionamentos daquela esfera ao plano macro dos grandes sectores da economia.

No mesmo número do *Journal of Evolutionary Economics* pode ler-se, da autoria de Andersen (1991), um outro estimulante contributo para a teoria económica da inovação e da evolução. Segundo Andersen, os "paradigmas técnico-económicos" (actualização mais compreensiva do conceito de "paradigmas tecnológicos" proposto recentemente por Dosi) devem ser entendidos não somente como meios de coordenação entre produtores de conhecimento tecnológico mas também como meios de coordenação dos grupos de produtores e dos grupos de utentes de dados tipos de artefactos, mediante especificações partilhadas de *interfaces* típicas entre ambas as partes.

O mais significativo nestes desenvolvimentos recentes da investigação económica é, do nosso ponto de vista, o reconhecimento e a prospecção teórica dos ligames doravante *endógenos* entre dinâmica tecnocientífica e dinâmica económica nas suas várias interfaces (o mesmo é dizer, contemporaneamente, dinâmica social).

Aliás, no caso das TI e mesmo, mais em geral, da electrónica, a existência de ligações imediatas ou quase-imediatas à sociedade, consubstanciais ao próprio processo da criação científica, foi sempre uma constante. Lembre-se o drástico efeito de aceleração constitutiva que um fenómeno dramaticamente social na sua "insociabilidade" como é a guerra trouxe à emergência e aos desenvolvimentos da moderna computação (vd. a este respeito as observações sempre actuais de Wiener, 1962, e, mais recentemente, de Edwards, 1988) e é interessante registar que praticamente todas as principais inovações da electrónica e da informática têm sido produto do trabalho em laboratórios de empresa (pode consultar-se uma listagem à data exaustiva em Mizrahi, 1986), articulando-se, desde a origem, com projectos de utilização. Também não será casual que grandes inventores como Bell, Hollerith, Marconi, exibissem igualmente notáveis capacidades empresariais, um traço que não era comum, nem ainda o é, fora deste domínio.

(13) Aludo a Ética tendo em vista uma acepção próxima daquela que, consubstancial à crítica de Hegel à moral "abstracta" de tipo kantiano, conheceu recentemente uma notável revitalização teorético-problemática devida a Williams (1987): não um programa universal da razão legisladora mas forma de vida em comum, concretamente experienciada, "that made particular sense to the people living in it" (*ibid.*, p. 104).

(14) É a questão a que procuram dar resposta as teorias de "racionalidade limitada", às maneiras de Hayek ou de Simon, e que Moles (*infra*) enfrenta, de outro ângulo.

(15) É interessante que a imagética de robôs antropomórficos, nimbada ainda de um halo de suspicácia e de vagos receios nos países ocidentais, funcione no Japão como metáfora emblemática amigável, popular, das novas tecnologias na *interface* homem/máquina (cfr. Marques, *infra*). Haverá provavelmente nestas diferentes atitudes a influência de contextos culturais que não coincidem: um, o oriental, predisposto a acolher com simpatia formas não humanas de inteligência, o outro, o ocidental, hostil de raiz ideológica a essa possibilidade (vd. nota 5, *supra*).

Referências

E.S. ANDERSEN: 1991, *Tecno-economic paradigm as typical interfaces between producers and users*, em: *Journal of Evolutionary Economics*, V.1, n. 2, Berlin/Heidelberg, Springer.

Kenneth J. ARROW: 1983 ([1]1979), *The Economics of Information*, em: Dertouzos, M.L. & Moses, J. (eds. lit.), *The Computer Age: A Twenty Year View*, Cambridge, Mass., MIT Pr.

Daniel BELL: 1983 ([1]1979), *The Social Framework of the Information Society*, em: Dertouzos, M.L. & Moses, J. (eds. lit.), *The Computer Age: A Twenty Year View*, Cambridge, Mass., MIT Pr.

B. CARLSSON & R. STANKIEWICZS: 1991, *On the nature, function and composition of technological systems*, em: *Journal of Evolutionary Economics*, V.1, n. 2, Berlin/Heidelberg,

Springer.
Helder COELHO: 1986, *Tecnologias da Informação*, Lisboa, Dom Quixote.
Sergio COTTA: 1971, *O Desafio Tecnológico*, trad. port., Coimbra, A. Amado (v.o.: *La Sfida Tecnologica*, Bolonha, Mulino, 1968).
Felix DUQUE: 1986, *Filosofia de la Tecnica de la Naturaleza*, Madrid, Tecnos.
Paul E. EDWARDS: 1988, *The Closed World: Systems Discourse, Military Strategy and Post WW II American Historical Consciousness*, em: *AI & Society*, V.2, London, Springer.
Martin HEIDEGGER: 1989, *"Já só um Deus nos pode ainda salvar"* (entrevista a *Der Spiegel*, em 23.9.1966, publ. no nº 23/1976), trad. port. em: *Filosofia*, v. III, nº 1/2, Lisboa, Soc. Port. Filosofia.
Hans JONAS: 1984, *The Imperative of Responsability. In Search of an Ethics for the Technological Age* (trad. ingl. pelo próprio autor) Chicago/Londres, The Univ. of Chicago Pr. (v.o.: *Das Prinzip Verantwortung: Versuch einer Ethik für die technologische Zivilisation*, Frankfurt a.M., Insel., 1979).
Murray LAVER: 1989, *Information Technology: agent of change*, Cambridge, Cambridge Univ. Pr.
André LEROI-GOURHAN: 1963, *Le geste et la parole*, Paris, A. Michel.
Roberto LISCIA: 1989, *Il Tempo e lo Spazio, risorse della Nuova Economia*, em: Doc. reprod. do Convénio de RESEAU "Le Telecomunicazioni saranno la differenze", Veneza, 5-7 Abril 1989.
Timothy W. LUKE: 1990, *Social Theory and Modernity*, Newbury Park (California)/London/N. Delhi, Sage Pub.
David LYON: 1990 ([1]1988), *The Information Society. Issues and Illusions*, Cambridge/Oxford, Polity Pr./Blackwell.
Vasco de MAGALHÃES-VILHENA: 1984, *Surto técnico e científico e "blocagem" social na Cidade Antiga*, em: *Antigos e Modernos. Estudos de História Social das Idéias*, Lisboa, Horizonte.
Y. MASUDA: 1981, *The Information Society as Postindustrial Society*, Bethesda MD, World Futures Soc.
Ian MCLEAN: 1989, *Democracy and New Technology*, Cambridge/Oxford, Polity Pr./Blackwell.
Marvin L. MINSKY; 1983 ([1]1979), *Computer Science and the Representation of Knowledge*, em: Dertouzos, M.L. & Moses, J. (eds. lit.). *The Computer Age: A Twenty Year View*; Cambridge, Mass., MIT Pr.
Jean MIZRHAI: 1986, *L'Échiquier de l'Électronique*, Paris, Hachette.
Abraham A. MOLES: 1990, *Les Sciences de l'Imprécis*, Paris, Seuil.
Lewis MUMFORD: 1964 ([1]1934), *Technics and Civilization*, NY, Harcourt.
Roger PENROSE: 1990, *The Emperor's New Mind. Concerning Computers, Minds, and the Laws of Physics*, Oxford, Oxford Univ. Pr.
Luiz Moniz PEREIRA: 1988, *Máquinas Humanas*, em: VV.AA., *Pensar a Ciência*, Lisboa, Gradiva.
Mark POSTER: 1990, *The Mode of Information. Poststructuralism and Social Context*, Cambridge/Oxford, Polity Pr./Blackwell.
Adriano Duarte RODRIGUES: 1990, *As Estratégias da Comunicação*, Lisboa, Presença.
Herbert A. SIMON: 1981, *As Ciências do Artificial*, trad. port., Coimbra, A. Amado (v.o.: *The Sciences of Artificial*, Cambridge, Mass., MIT Pr., 1981).
——————— 1990, *A racionalidade humana e os seus limites*, trad. port. em: VV.AA., *Balanço do Século*, Lisboa, IN-CM.
Aaron SLOMAN: 1984, *The structure of the space of possible minds*, em: Torrance, S. (ed.lit.), *The Mind and the Machine. Philosophical Aspects of Artificial Intelligence*, Chichester, E. Horwood.
Tom STONIER: 1988, *Machine Intelligence and the Long-Term Future of the Human Species*,

em: *AI & Society*, v.2, London, Springer.

Steve TORRANCE: 1984, *Philosophy and AI: some issues (Editor's Introduction)*, em: Torrance, S. (ed.lit.), *The Mind and the Machine. Philosophical Aspects of Artificial Intelligence*, Chichester, E. Horwood.

Francisco J. VARELA: 1988, *Connaître les Sciences Cognitives. Tendances et perspectives*, trad. fr., Paris, Seuil (v.o.: *Cognitive Science. A cartography of current ideas*).

Jean-Pierre VERNANT: 1975 ([1]1962), *Les origines de la pensée grecque*, Paris, PUF.

Norbert WIENER: 1962, *Cybernetique et Société*, trad. fr., Paris, UGE, col. 10–18 (v.o.: *The human use of human beings. Cybernetics and Society*, Boston, Houghton Mifflin, 1954).

Bernard WILLIAMS: 1987 ([1]1985), *Ethics and the limits of Philosophy*, London, Fontana/Collins.

Acrescem a esta relação todos os estudos publicados no presente volume.

I

USOS DA TEORIA

USES OF THEORY

1

COMPUTAÇÃO & INTELIGÊNCIA ARTIFICIAL

COMPUTATION & ARTIFICIAL INTELLIGENCE

Aaron Sloman

School of Cognitive and Computing Sciences,
Univ. of Sussex, Brighton (UK)

The Relevance of Computation to the Study of Mind

Abstract

Although the development of Artificial Intelligence research based on the use of computers has transformed philosophy, psychology, linguistics, and other areas of cognitive science, it is argued that the key notion for understanding how minds work remains "mechanism". Computational systems are merely one type of mechanism, and there is no reason to restrict ourselves to these in attempting to explain intelligent capabilities or design new machines. This leaves us with the problem of studying different types of mechanisms and different architectures within which mechanisms can be combined: the space of possible designs for intelligent systems.

Resumo

Embora o desenvolvimento da pesquisa em Inteligência Artificial baseada no uso de computadores tenha transformado a filosofia, a psicologia, a linguística e outras áreas da ciência cognitiva, é sustentado aqui que a noção chave para a compreensão de como a mente funciona continua a ser a de "mecanismo". Os sistemas computacionais são apenas um dos tipos de mecanismo, e não há razões para nos restringirmos a este tipo nas tentativas de explanação das aptidões de inteligência ou concepção de novas máquinas. Este ponto de vista coloca-nos perante o problema de se estudar diferentes tipos de mecanismo e diferentes arquitecturas de combinação de mecanismos: o inteiro espaço de possibilidades na concepção de sistemas inteligentes.

The Problem

I once argued (Sloman 1978) that computational concepts and theories suffice to explain the nature of perception, learning, creativity, motivation, emotions, feelings, consciousness etc. because all these are essentially computational states, events or processes. But this claim, implicit in much work in Artificial Intelligence and Cognitive Science, presupposes that we know what computations are. Do we? New models of computation, including "neural computations" (e.g. McClelland and Rumelhart 1986) suggest that the concept of computation needs to be broadened. But how far?

We could define "computation" in the normal logico-mathematical sense (Turing-equivalence). We then risk too narrow a concept.

We could define it to include all the kinds of processes that appear to play a role in intelligence, including the effects of drugs and hormones, for instance. This leaves the distinction between computational and non-computational processes quite blurred.

Instead, I suggest we abandon the idea that computation is the *key* notion underlying intelligence, and study different kinds of mechanisms to see what kinds of roles they can play in the design or explanation of intelligent systems.

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposyum, 1989, 41/51
1992, APDC & SPF, Lisboa

The standard concept of computation.

If our study is to include asynchronous distributed processes, with sub-processes only making contact intermittently, e.g. via messages or remote procedure calls, then we can't require computational processes to have well-defined states and transformations, as a Turing machine does. Similarly, if we wish to include continuous changes in neuronal states, then we are looking at processes that Turing-equivalent computations can only approximate.

Many brain processes relevant to mental states involve chemical processes, as shown by drugs and hormonal changes that alter the character of mental processes — are these also computational? Why stop there? Why not include digestion, the growth of a tree, or even the processes in a thundercloud?

Continuous geometrical processes are simulated in computer graphics systems, but they are rarely used internally in AI programs. (Even Funt 1977 merely simulates a discrete approximation to continuous change).

Even if all physical processes are discrete at some level, as required by quantum physics, it is often useful to conceive of them as continuous (e.g. in some feedback systems studied in control theory), and this could be true of some mechanisms required for intelligence.

If all computational approximations to such mechanisms inevitably run more slowly, that may mean that the computational mechanisms are not part of the explanation of how actual systems work - unless we broaden the concept of computation.

It may also be the case that some of the probabilistic processes postulated by quantum physics cannot be accurately modelled on a classical computer.

Even if all physical processes really were computational, then it would be trivially true, and therefore uninteresting, that intelligent systems require computational mechanisms.

I am not saying that non-Turing processes are essential to intelligence, merely that there is no apriori reason to rule them out. Equally I am not saying that Turing-equivalent processes are unnecessary for the design of intelligent systems. We may need both.

Defining "computation" in terms of discreteness or Turing equivalence (Pylyshyn 1986, Haugeland 1985) arbitrarily excludes legitimate scientific and engineering attempts to explore computations using continuous processes, like slide rules and analog computers, or soap bubbles on stretched frames used as computers to work out minimum stress designs. Pinning intelligence down to computation so defined might wrongly exclude explanatory theories that go beyond currently known forms of computation.

Explanatory mechanisms cannot be entirely formal.

There is a sense of "computation" in which patterns made by leaves blown about on the forest floor could form a computation, for instance a derivation of the solution to an equation, as long as there is a description of the patterns according to which they conform to some set of rules for computation (or proof).

Physical embodiment is not required for this formal concept of computation: a sequence of Gödel numbers, itself represented by a single large Gödel number, would be a computation. Mathematical functions can be defined to determine whether a particular (very large) number is or is not a computation of a particular sort in virtue of its decomposition into powers of prime numbers, where those powers are themselves

decomposable into representations of intermediate states in a computation. A structure so defined can satisfy the formal conditions for being a computation. But it could not, in itself explain behaviour. It cannot make anything happen. A number suitably decomposed may satisfy a formal definition of "computation", but that shows that satisfying the formal definition cannot in itself explain how intelligent *behaving* systems are possible.

Avoiding circular definitions.

Clark (1988) suggested that "computation" could be defined in terms of the origin and use of a mechanism. To be computational it would need to have been designed or used for a particular sort of purpose or to have evolved biologically to serve certain needs of some organism. At this conference Steels suggested an elaboration of this idea: a mechanism is computational if it is used to model something else for the purposes of prediction or explanation. On this account, "computation" is not a structural concept, but a functional concept, like "screwdriver". Just as a wide range of physical objects have the potential to be used as a screwdriver, so might any physical process have the potential to be a computation, if it is used in the appropriate way.

If this characteristic function is defined in terms of notions like "model", "representation", "prediction", or "explanation", then "computation" cannot be used to explain how intelligence is possible: for our purposes, we risk circularity if we use a concept of "computation" that presupposes the existence of intelligent agents, and then try to use the same concept to explain how intelligence is possible.

For similar reasons we cannot define computation in terms of manipulation of *meaningful* structures or as *symbol* manipulation (where "symbol" implies having a meaning), if we wish to use computation to explain semantic abilities. This invalidates any definition of "computation" in terms of manipulation of structures with semantics or content.

Defining computations as "rule-governed" processes also risks circularity, if being *governed by* rules involves *understanding* the rules. Understanding is what we wish to explain: we must not just assume it as a primitive.

Causation, control and computation

We could try to avoid notions like "symbol", "rule", "meaning", "interpretation", and try to define computational processes as those in which behaviour is controlled by structures. This would just make use of the notion of structure and control, or causation. But this is so general that it includes all physical processes: the structure of a stone controls its behaviour when thrown into the air. Is that a computation?

The notion of control is just a special case of the notion of "cause": one thing controls another if there is an ongoing causal relation between the controller and the controlled. (If it is a two-way relation, then the control includes feedback.) Analysing the concept of "cause" is one of the hardest problems in philosophy, but I shall, for the purposes of this paper, assume we all understand it, and will therefore uninhibitedly use "cause" and related concepts like "control".

Apart from its excessive generality, the attempt to define computation in terms of processes controlled by structures has a problem: computer programs don't control behaviour in any simple sense. In particular, programs need not *totally* control the

behaviour of a computer, since their behaviour is often limited by operating systems. For instance on a time-shared computer each process is frequently interrupted by the operating system and suspended while another program runs. Access rights can also be controlled. So computer programs normally have only limited, or "bounded", control over the processes they generate.

Most programs written by human programmers are not even in direct control of the processes they generate: instead, they are often compiled to a different, machine-code, program, and after that the compiled version controls the process. The original program no longer plays any role: if it were altered it would make no difference to the behaviour. Its control is "ballistic", not "online".

An interpreted program has more direct "online" control than a compiled one. But even there, what is in control depends on a point of view. You can think of the program as actively controlling the process, or you can think of it as a passive structure, while the interpreter controls processing by examining this structure. Similarly, a compiled program interpreted by the central processing unit can be viewed as a passive structure used by the CPU to generate behaviour.

If neural processes are to be called computations, then we cannot require computation to involve a separable stored program using an explicit programming language. The controlling structure is implicit in the topology of neural interconnections and connection weights, the mechanisms that adjust the weights, and the mechanisms that use inputs and weights to determine activation levels.

If we are prepared to call those programs, why not every physical or chemical structure that controls the behaviour of something?

Virtual machines.

If we are tempted to say that when a program is compiled it's the machine code that is in control, version then there's usually a mismatch between the structure of the events that occur at machine code level (manipulation of bit-patterns) and the events represented in the original high level language (manipulation of numbers, words, lists, trees, networks, arrays, etc.).

Programs in a high level language refer to 'virtual' structures and operations on them, whereas the machine code instructions will generally refer only to operations on bit patterns and locations in memory.

Even bit patterns are virtual structures: what the physical machine contains is not bit patterns but wires, switches, voltages, currents, etc.

So a computer can have different levels of virtual machine, each implemented in a lower level machine, until ultimately a physical machine is used.

Are these virtual machines mechanisms with causal powers to generate and control behaviour? Yes, because a computer may do one thing instead of another because a new subtree is added to a tree inside it. The virtual operation can cause other virtual operations. The whole process can cause physical events to occur, like something printed on a terminal.

Causal relations between states and events in a virtual machine depend on physical causation for their efficacy, but that doesn't mean there are no causes in virtual machines. Different causal relations can coexist at different levels: when a compiled, machine-code program is controlling behaviour at the bit level there is also usually another virtual program controlling behaviour of a higher level virtual machine: at a level of abstraction

closer to the original source code.

The same notion of layers of virtual machines may be applicable to neural nets. However, we are still at a very early stage in our understanding of such processes.

We need a better set of concepts

Often in the past a collection of ill-defined concepts, like "force", "energy", "momentum", has been transformed for the purposes of science, with new precise concepts co-existing and developing in parallel with the original nontechnical concepts. The same thing is likely to happen to concepts like "instruction", "program", "control", "compute", "machine". When we try to apply them in new situations to make fine, previously unnoticed, distinctions we discover that the concepts are unsuited for the job, and then have to develop and extend them.

I believe the same thing will happen to concepts like "mind", "belief", "desire", "consciousness", "emotion" and the like.

Of course, new stipulative definitions can be introduced at will, but if you want a good theory that encompasses the basic explanatory concepts of computer science and provides a non-circular foundation for explanations in cognitive science, then you cannot define "computation" as you like. But there does not seem to be a good definition that suffices for the task: not even the existing formal definitions mathematically equivalent to Turing computability.

So instead of looking for a major distinction between computational and non-computational processes we should examine many small distinctions between different kinds of mechanisms, with overlapping sets of similarities. The task of a theory of mechanisms for intelligence is then to explore the space of possible mechanisms and designs capable of explaining or replicating intelligent behaviour. There is no need restrict the space to processes that can run on conventional computers.

What does the rest of the space of processes relevant to intelligence look like? What are the interesting and important differences between sub-spaces?

Independently variable, causally interacting, sub-states

At the risk of vacuity, lets move to a high level of abstraction. A key notion appears to be that seems to emerge from all this is that of a system having sub-states that can change independently and interact causally with one another, where the sub-states may be physical or virtual and causation may be direct or indirect, bounded or total, online control or ballistic control, continuous or momentary, mediated by an "interpreter" or physically direct. Different sub-states have different causal roles in the total system, and we need to understand this internal "functional" differentiation to understand how intelligent systems work.

An important characteristic common to computers and brains is that their lowest level physical structure allows enormous variability of sub-states: there are astronomically large numbers of significantly different possible states between which a computer, or a brain, can change, with very rapid switching between them, without the need to traverse intermediate states. This follows from the existence of large numbers of units (transistors, neurones etc.) that can change state. N switches each with K possible states permit $K \wedge N$ possible total states. If N is large, even small values of K (e.g. 2) produce unimaginably large sets of possible states.

Complex systems of different kinds can be distinguished according to the kinds of independently variable sub-states that can co-exist, the kinds of variation of the sub-states, the kinds of interactions between sub-states, whether the number of sub-states is fixed or can change, and so on. While a program is running in a computer the number of causally significant sub-structures (e.g. parse trees, or items in a database, or even incrementally created subprograms) can change drastically.

Chairs, tables, and even steam-engines, have nothing like the variety of causal interactions between different substates that can be found in a brain or the simplest modern computers. They lack the functional differentiation, the variety of causally distinct sub-states, and the ability to remove old or add new sub-states.

Instead of worrying about which such categories of mechanisms are and which are not computation, we should instead ask which of them are useful, or even necessary, for different kinds of intelligence. We should not expect that human intelligence requires exactly the same kinds of mechanisms, or combinations of mechanisms, as ants or even mice, though there may be some overlap.

So we should now start trying to classify behaving systems according to how many different kinds of sub-states they can simultaneously support, what kind of variability they have (e.g. is it continuous or discrete variation, is it variation in an N-dimensional vector space or is it *structural* variation), what their causal interactions are, how much internal functional differentiation they have, whether the number of independently changeable sub-states is fixed or can develop, and so on. Even Turing-equivalent systems may differ significantly on these criteria.

Two kinds of sub-state that are of particular interest could be defined in terms of the causal roles characteristic of belief-like and desire-like states. Further elaboration requires an exercise in the global design of intelligent systems: too big a task for this paper.

"Structural" variability

Engineers and physicists often represent change in terms of a moving point in an N dimensional vector space. Richer kinds of variation, allowing changing complexity, are required for biological and intelligent systems. For example, growth is not merely a process of changing measurements: it can produce an entirely new structure, such as an eye or wing, or a new relationship such as connection via a nerve fibre.

Similarly, states of understanding different sentences or perceiving physical structures cannot easily be represented by fixed dimensional vectors of measurements, because of the unbounded hierarchical complexity of sentences in natural languages, or of physical structures.

Cognitive processes can therefore involve state changes in which new substructures are created or old structures rearranged. This *structural* variation is different from either continuous or discrete variation along a onedimensional scale, or simultaneous variation in N different dimensions in a uniform multi-dimensional space. Structural variation, unlike the other kinds, can involve increasing or decreasing complexity.

The kind of structural variability a system is capable of limits the kinds of interactions it can have with the rest of the world. For instance a system that can have only N distinct states cannot distinguish N+1 kinds of environmental situations. (However, sometimes the environment has to be counted as part of the state - a point elaborated in Sloman 1978.)

All sorts of different virtual machines can be embedded in the same physical computer. Thus structural variability may be a property of a virtual machine implemented in a fixed

dimensional lower lever machine, e.g. in a bit–array.

Many mechanisms with variable internal sub–states can be connected to an environment via transducers in such a way that changes in the internal sub–states are very sensitive to the nature of the environment, including structural features of the environment. (AI work in perception studies such mechanisms.) The internal sub–states produced by the environment can then control behaviour. (Compare hearing "Your house is on fire" and "your hat is on fire".)

Discrete variability, including creation of new sub–structures and new connections between old structures, seems to be required for some aspects of human intelligence. But that does not rule out other kinds of mechanisms, including some with continuous variability.

Getting a machine to understand.

Under what conditions can a machine treat some of its internal substructures as referring to anything, i.e. as having a semantics *for the machine*, and not just for us? Does it have to be a computational machine? Must all its computations be Turing equivalent?

Some of the requirements for human–like semantic capabilities are structural (i.e. mechanisms are needed with certain capabilities) some functional (i.e. the mechanisms need to be used in certain ways by the whole system).

I have previously argued (Sloman 1985, 1986) that there is no single sharp distinction between systems that do and systems that don't understand the structures they manipulate. There is a cluster of prototypical semantic capabilities and different subsets may be instantiated in different mechanisms. There is no necessary and sufficient subset of capabilities for "real" understanding: our concepts are not that precise.

A significant subset of semantic capabilities can be found even in the relationship between a digital computer and its machine language: the computer clearly uses some sub–structures of its internal states to refer to locations in its memory, and others to refer to possible internal actions. Of course this is a simplified kind of semantics, without any links to motivational states. Nevertheless, it's the machine, not a person, that uses the bit–pattern in an address register to determine which location is to be interrogated or changed, and so on.

More sophisticated semantic capabilities, and therefore semantic competence closer to human abilities, would require both richer formalisms (i.e. sub–states with richer kinds of variability) and a richer functional division of the system into sub–mechanisms (some of them virtual ones) with different causal roles. An architecture that included sub–mechanisms concerned with motivation is necessary if the machine is to interpret symbols as having meanings that matter to it e.g. "Your batteries require re–charging".

Some structural conditions for semantic mechanisms

Sentences can contain disjunctions, negations and quantifiers, and pictures can have both a smaller dimensionality than the scenes they depict and often also a quite different topology (e.g. a typical 2–D picture of a 3–D wire–frame cube has more junctions than the cube has vertices). So there is no simple mapping between representing structures and what they represent.

Tarski (1956) showed how a set of axioms in a logical formalism can be modelled by sets of objects. We now need a more general theory that is applicable to a wider class of

formalisms, including perhaps continuously variable structures, and allows a wider class of semantic relations, including many kinds of indeterminateness of meaning.

If a mechanism accommodates a logical formalism and supports transformations corresponding to valid rules of inference, then the Tarskian model theory will allocate an indefinitely large set of possible models to the formulas in the machine at any one time. This set of possible models can be reduced by adding new (independent) 'axioms', but this will never suffice to pin down the model to a particular bit of the physical world.

Causal links between the mechanism containing the formalism and bits of the world can further constrain the semantic mappings, as addresses in a machine are causally related to locations in its memory by the electronic mechanisms. Perceptual mechanisms and motors controlled by the internal sub-structures can set up causal links in both direction between internal sub-structures and aspects of the external environment. These causal links can be used to reduce the possible Tarskian interpretations of internal "axioms", "predicates", "individual constants", etc.

However residual ambiguity always remains. I see no reason to require any guarantee of uniqueness of reference, as long as the mechanisms function adequately for the organism or agent concerned. Some philosophers will be unhappy about this, but that's only because they adopt unrealistic and unsatisfiable standards for rationality.

These ideas need to be extended beyond formalisms to accommodate analogical representations and semantic roles for continuously variable sub-structures. A more general notion of a portion of the world being a model for some representing structure in a machine might be defined by relating the sub-states to the *roles* that they can play in a behaving system, and the ways in which these roles interact with the environment.

A thermostat connected to a room heater and cooler has a variable (primitive) "belief-like" sub-state that represents the current measure of temperature of the room. It also has a variable (primitive) "desire-like" sub-state corresponding to the required temperature setting. Causal interactions between these states and a switch determine whether a heater or cooler is turned on or off.

Any such sensor/controller pair corresponds to a variable aspect of the environment. The correspondence need not be exact: for instance, temperature control settings may be a discrete set while the range of possible environmental temperatures is continuous. Also the correspondence may not be one-to-one because of noise, lack of resolution, time-delays or other aspects of the measuring device or controller.

In some AI vision systems the sensor produces not just a one-dimensional measurement but a structural description in a logical language. Here the causal link between the external and internal states is very complex and indirect, and usually discards some incoming information (e.g. fine detail) and adds other information (e.g. inferences about hidden surfaces).

Suppose a controller can separately measure and control a range of properties P1, P2,...Pn, of a machine or plant, but can only work on one of them at a time. For each such property Pi it has some kind of sensor Si that will produce or change the corresponding belief-like sub-state, a settable control knob (or set of keys, etc) Ci that modifies the corresponding desire-like sub-state Di and an output-channel Oi that controls the relevant property Pi. Suppose only one such output channel can be turned on at a time, and a selector can switch between the different properties to determine which is controlled.

In this case, internal desire-like and belief-like sub-states still correspond to possible environments but it is no longer true all the time that each desire-like sub-state tends to change the environment to correspond to it. However, at any time when the *ith* controller

is disabled we can talk about the effect the *ith* desire-like state WOULD HAVE in that context (including the *ith* belief-like state estimating Pi) IF its output channel were selected. If the different sensors can be temporarily disconnected, it is then no longer true that the environment tends all the time to produce a selection from the possible belief-like sub-states. Nevertheless there is a well defined notion of what the causal correspondence would be if the particular sensor or controller were connected, and working normally.

Far more complex designs are possible where instead of a fixed number of internal belief-like sub-states connected to the same number of sensors there is a more complex and more richly variable collection of sub-states in a virtual machine, based quite indirectly on input from sensors. There might be N different sensors, and K different belief-like states all derived from different combinations of the sensor readings, where K can vary over time, like a visual system building 3-D interpretations of a 2-D array of retinal information. (Ballard and Brown 1982.)

Some of the belief-like sub-states thus produced may be stored for use as records of past situations, instead of being "overwritten" as new information comes via the sensors.

In such cases the causal links between the environment and the current set of belief-like states will be far more complex and indirect than in a thermostat. Moreover, the more complex the process that creates internal structures from sensory input, the more scope there is for malfunction in the internal mechanisms, so that the set of counterfactual conditionals linking the internal states to the environment becomes even more complex, and the correspondence depends less and less on direct causal links and more and more on structural properties of the internal states that constrain possible interpretations.

Similarly, we can design a high level virtual machine containing a variable database of desire-like states, created by "motive-Generator" rules, with context sensitive "motive-comparator" rules and decision-making rules for determining relative priorities of desire-like states, selecting a subset for action, retrieving or creating plans, and executing plans, possibly over an extended period with different plans interleaved if necessary. This extra complexity of processing (elaborated in Sloman 1987) reduces the directness of the causal links between desire-like states and states in the environment. Instead of using simple discrepancy measures to determine whether a particular switch should be turned on or off (as in the thermostat) it may require quite complex internal processing of the relationship between belief-like and desire-like states to determine whether a desire-like state is "satisfied" or not, and if not what action to take.

So besides mechanisms whose internal belief-like and desire-like sub-states are directly causally linked to each other and to the environment there are designs where the causal links are very complex and indirect.

The more complex and indirect the causal links, and the more different sub-states have to time-share the same input and output causal channels, the more any semantic correspondence is going to depend on structural properties of the representing states (as in Tarskian model theory or other kinds of more analogical modelling relationships), and the more scope there is for indeterminacy in the semantics.

The more indirect and abstract semantics together with generative capabilities can also enable sub-states to refer to things remote in space and time or even refer to possibilities that are never realised, for instance options considered and rejected in planning. This ability to give an internal sub-state a "supposition-like" instead of a "belief-like" role depends on the causal links with the environment being less direct than in the thermostat.

I suspect that biological evolution includes developments along the directions indicated here, with decreasing causal coupling of internal and external states going hand in hand

with increasing structural complexity and functional differentiation of internal virtual sub-states.

Although it is not clear exactly how, I believe that these comments provide a basis for extending Tarskian semantic ideas to cover a wider class of formalisms and semantic relationships (Sloman 1990).

In practice there will not be a determinate class of possible environments that are models of an agent's beliefs or desires – i.e. the semantics for the internal states will be indeterminate in various ways. This in itself should not disturb us if we are interested in explaining human intelligence, since there is plenty of evidence that human languages (and probably internal representations too) are indeterminate in various ways. (E.g. how big is a big man? How much water must fall on a rainy day? Is a circle an ellipse? Is it 3 o'clock on the moon? Are liquids mixtures or compounds? Where are the boundaries between species of birds?)

Some of the requirements for such a generalised semantics are as follows:

– Unlike the languages discussed by Tarski and logicians there is no need for a precisely delimited syntax to be used: intelligent systems can creatively extend the variety of representing structures they use, and humans frequently do this.

– The semantics assigned to particular notations by an individual need not be fixed: even one-off interpretations are possible (including "lets pretend" games by children).

– The relation between notation and portion of the world may depend not only on an individual's mechanisms and functional architecture, but also on the agent's being part of a social system made of other intelligent individuals.

Conclusion

I have argued that instead of assuming that computation is the key to understanding intelligence we should fall back on the more general concept of "mechanism", having learnt how to extend the concept through our studies of computational mechanisms, including virtual machines of many kinds.

Unlike Searle (1980) I am not arguing that a digital computer cannot understand symbols that it manipulates. Rather I am simply drawing attention to the need to consider a wider range of mechanisms than Turing equivalent machines, for the purpose of explaining or designing intelligent systems that can function effectively in our world.

Attempts to achieve the same generalisation by extending the notion of computation founder on the problems of finding a suitably general notion that does not trivially incorporate all physical processes, and does not presuppose the prior existence of intelligent agents.

The hoped-for single boundary between computations and non-computations should be replaced by sets of features of computation-like mechanisms defining a *variety* of design boundaries with different implications, which we still need to explore, in order to characterise a taxonomy of types of mechanisms relevant to mind or intelligence. Most of the work remains to be done, and will impact not only on AI but also on philosophy, cognitive science, studies of biological evolution and comparative ethology.

Note:

This is a shortened version of a paper originally written for the Turing conference at Sussex University, April 1990.

References

Ballard, D.H. and C.M. Brown, *Computer Vision*, Prentice Hall 1982.

Clark A *'Computation, connectionism and content'* in (ed) Yves Kodratoff, *8th European Conference on AI*, Munich, 1988.

Fodor, J.A. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, MIT Press, Cambridge Mass, 1987.

Funt, B.V. *'WHISPER: a problem solving system utilizing diagrams and a parallel processing retina'*, in *Proceedings 5th IJCAI*, MIT, 1977.

Haugeland, John, *Artificial Intelligence: The Very Idea*, Bradford Books, MIT Press, 1985.

McClelland, James L. D.E. Rumelhart et al., *Parallel Distributed Processing, Vols 1 and 2*, MIT Press 1986.

Pylyshyn, Zenon W., *Computation and Cognition: Toward a Foundation for Cognitive Science*, Bradford Books, MIT Press, 1986.

Searle, John, *'Minds Brains and Machines'* in *The Behavioural and Brain Sciences*, 1980.

Sloman, A, *The Computer Revolution in Philosophy: Philosophy Science and Models of Mind*, Harvester Press and Humanities Press, 1978.

Sloman, A, *'What enables a machine to understand?''* in *Proceedings 9th International Joint Conference on AI*, Los Angeles, 1985.

Sloman, A, *'Reference without causal links'* in L. Steels, B. du Boulay, D. Hogg, editors, *Advances in Artificial Intelligence-II* Proc 7th European Conference on AI, Brighton, 1986, North-Holland.

Sloman, A, *'Motives Mechanisms Emotions'* in *Emotion and Cognition* 1987

Sloman, A, *'Beyond Turing Equivalence'*, in *Proceedings Turing 1990 conference* Sussex University, April 1990.

Tarski, A, *'The concept of Truth in formalised languages'*, in *Logic Semantics Metamathematics*, trans J.H. Woodger, Clarendon Press, Oxford, 1956.

Marcelo Dascal

Tel Aviv Univ., Tel Aviv (Israel)

Why does Language matter to Artificial Intelligence?

Abstract

Artificial intelligence, conceived either as an attempt to provide models of human cognition or as the development of programs able to perform 'intelligent' tasks so far performed only by humans, is primarily interested in the uses *of language. It should be concerned, therefore, with* pragmatics, *the theory of language use. But pragmatics should be conceived broadly, as covering, in addition to the theory of communication (or the social use of language), also the 'mental' uses of language (in reasoning, for example) and the 'existential' dimensions of language as a determiner of the world we live in. In this paper the relevance of these three kinds of pragmatics — sociopragmatics, psychopragmatics, and ontopragmatics — for AI are explored.*

Resumo

Quer como tentativa de construir modelos da cognição humana, quer como desenvolvimento de programas 'inteligentes', a disciplina conhecida por 'Inteligência Artificial' deve interessar-se primordialmente pelos usos *da linguagem. Ela deve portanto preocupar-se mais em conhecer de perto a* pragmática, *a teoria do uso da linguagem. Mas a pragmática deve ser concebida num sentido amplo, comportando, além da teoria da comunicação (ou do uso social da linguagem), também o estudo dos usos 'mentais' da linguagem (no raciocínio, por exemplo), assim como da dimensão 'existencial' da linguagem enquanto determinante do mundo em que vivemos. Explora-se aqui a pertinência desses três tipos de pragmática — a sociopragmatica, a psicopragmática e a ontopragmática — para a Inteligência Artificial.*

Questions whose answers seem obvious have a special power. They insidiously suggest that there may be something wrong with the obvious answers; they insinuate that their apparently solid grounds are perhaps shaky. They thus call attention not so much to the answers, but to their grounds — which is what they in fact call into question.

In this paper, I purport to explore this heuristic power of my title question.[1] I will first review some typical explanations of the undisputable importance of language for AI. Next, I will propose a typology of the grounds offered for such explanations. Finally, I will indicate what kinds of tasks and difficulties face any attempt of implementing each of the approaches discerned. It is my hope that, by becoming aware of the differences in their underlying assumptions, practitioners as well as users of the new art of AI will be in a better position to judge where they may have gone wrong, to redirect their efforts, and to discern reasonable expectations from unfounded promises.

1. Three Answers

1.1 Descartes and Turing

After having summarized his theory of blood circulation, of the role of the nerves, and

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposyum, 1989, 52/78
1992, APDC & SPF, Lisboa

of other aspects of bodily physiology, in Part V of the *Discours de la Methode*, Descartes argues that, whereas well-designed machines could eventually be undistinguishable from animals, they would be easily distinguishable from human beings, for

> ...jamais elles ne pourraient user de paroles ni d'autres signes en les composant, comme nous faisons pour declarer aux autres nos pensées. Car on peut bien concevoir qu'une machine soit tellement faite qu'elle profere des paroles, et même qu'elle en profere quelques-unes a propos des actions corporelles qui causeront quelque changement en ses organes; comme si on la touche en quelque endroit, qu'elle demande ce qu'on lui veut dire, si en un autre, qu'elle crie qu'on lui a fait mal, et choses semblables; mais non pas qu'elle les arrange diversement pour repondre au sens de tout ce qui se dira en sa presence, ainsi que les hommes les plus hébétes peuvent faire (AT VI, pp. 56–57).[2]

The inability of animals and machines to use a "true language", i.e. "d'exprimer soit par la voix, soit par les gestes quelque chose qui puisse se rapporter a la seule pensée et non a l'impulsion naturelle" (Letter to Henry More, 5 February 1649; AT V, p. 278)[3] contrasts sharply with the ability of every man, regardless of his or her level of intelligence, to use such a language. Consequently, language is *the only sure sign of a latent thought in the body* and it can be taken as the *true difference* between men and animals/machines (ibid.).

Descartes's argument was based on the then prevalent conception of machines. On this conception, a machine's performance is limited — indeed determined — by the number and variety of its parts:

> Car, au lieu que la raison est un instrument universel qui peut servir en toutes sortes de rencontres, ces organes [des machines] ont besoin de quelquer particulière disposition pour chaque action particulière; d'où vient qu'il est moralement impossible qu'il y en ait assez de divers en une machine, pour la faire agir en toutes les occurrences de la vie de même façon que notre raison nous fait agir (AT VI, p. 57).[4]

Presumably, if confronted with present-day powerful programmable computers, Descartes wouldn't substantially change his position. For he might echo a claim made by many adversaries of AI: a programmed machine does what the programmer tells it to do, in the situations s/he anticipates. Descartes would probably regard any programmed abilities, even the elaborate 'learning' capabilities and heuristics that some programs display, as essentially limited, if not in number at least in variation, for they are the result of some *finite* and *given* application of the programmer's ingenuity. Hence, the machine's 'mechanicity' — as opposed to true creativity in the presence of the really unexpected — is bound to be discovered at some point. Language behavior plays also the key role as an indicator of intelligence in the well-known Turing's Test. The test is based on the 'imitation game', where an interrogator in one room puts questions to a man and a woman in another room, in order to find out who is the man and who is the woman. Replacing one of the humans by a machine that 'tries' to impersonate a human and communicates with the interrogator through a teletype, the game becomes a test of machine intelligence (cf. Turing, 1950).

Unlike Descartes, Turing is not committed to the idea that a machine that passes the test possesses 'consciousness' or any other mental attribute. He proposed a strictly *operational* test: a machine passes the test if its linguistic behavior *counts socially* (at least in the eyes of the interrogator) as a proper use of language. The test does not purport to reveal *what* 'intelligence' (or 'thought') is, but only to provide a *sign* of its presence.[5]

Descartes was convinced that no machine (nor animal) would pass such a test. He would be surprised to learn that several language-using computer programs have been quite successful in fooling even expert interrogators. For example, Colby's artificial

paranoiac, PARRY, was diagnosed as such by a group of psychiatrists interviewing mental patients via a teletype. None of the psychiatrists, nor their colleagues that reviewed the transcripts of the interviews, realized that they were interviewing a computer! (cf. Boden, 1977, pp. 99–100).

These facts, however, do not detract from the power of the idea that the appropriateness of linguistic behavior to an unrestricted set of situations is perhaps the best indicator of intelligence. For an analysis of these 'successful' programs shows that the devices that do the trick are quite simple ones, which seem to involve no 'true understanding'. Their success in the test is due, to a large extent, to the pre-understandings embedded in the interrogation setting: no doubt the interviewers' assumption that they were dealing with 'mental patients' influenced their judgment, just as human patients 'analyzed' by the 'psychoanalyst' program ELIZA tended to accept the program's vacuous questions as meaningful. Presumably, more stringent applications of the test, where such pre-judgments would be barred and where a wider variety of unexpected inputs were used, would be able to sort out intelligent from 'mechanical' uses of language.

1.2 Hobbes and Fodor

A completely different explanation for the importance of language vis-a-vis cognition is given by Thomas Hobbes, Descartes's contemporary. Hobbes claimed that language has, in addition to its communicative function, not only some minor cognitive function, but also the highest possible such function. Language, says Hobbes in an often (partially) quoted passage, is the stuff out of which judgment, reasoning and science are made. It is so important that it serves to define the highest faculty of the mind, namely *reason*:

> ...reason, in this sense, is nothing but *reckoning*, that is adding and subtracting, of the consequences of general names agreed upon for the marking and signifying or our thoughts: I say *marking* them when we reckon by ourselves, and *signifying* when we demonstrate or approve our reckonings to other men. ..... By this it appears that reason is not, as sense and memory, born with us; nor gotten by experience only, as prudence is: but attained by industry; first in apt imposing of names; and secondly by getting a good and orderly method in proceeding from the elements, which are names, to assertions made by connection of one of them to another; and so to syllogisms, which are the connections of one assertion to another, till we come to a knowledge of all the consequences of names appertaining to the subject in hand; and that is it, men call *science (Leviathan* I,5).

Consistently with this overall cognitive importance assigned to language, Hobbes does not refrain from making truth itself dependent upon words: "truth consisteth in the right ordering of names in our affirmations" (*Leviathan* I,4). This, of course, was considered scandalous by Hobbes's contemporaries, for it implied the relativization of truth. His computational/linguistic conception of thought, though taken up and developped by such thinkers as Leibniz and Condillac, was in general relegated to oblivion.

Such a conception was discovered anew with the upsurge of computer technology. Stripped of its contention that *natural language* is directly involved in cognitive processes, it was soon to become orthodoxy in the new cognitive science and in AI. According to J.A. Fodor, one of the chief defenders and articulators of this conception, "mental processes [are] computational", i.e. they are "formal operations defined on representations", and the mind is, "inter alia, a kind of computer ... carrying out whatever symbol manipulations are constitutive of the hypothesized computational processes" (Fodor, 1980, p. 230). Unlike Hobbes, Fodor does not assume that the symbols or representations in which cognitive

processes are carried on are those of any public language. For public languages are learned through cognitive processes that must already be carried on in some pre-linguistic symbol system. Such a system, which Fodor — following in this respect a Cartesian rather than a Hobbesian path — does not hesitate to consider as innate, deserves the name of 'language of thought'. Since it is necessary for the acquisition of any public language, it must be endowed with a semantic and syntactic power at least equivalent in complexity to that of any learnable public language (cf. Fodor, 1975, *passim*).

On this view, then, the importance of language to cognition is obvious: cognitive processes take place in a language-like medium; they are *formal* manipulations of language-like symbols, governed by *syntactic* constraints; acquiring a public language consists in using the 'language of thought' in order to formulate hypotheses about the public language's rules and verifying their empirical adequacy; and understanding a public language utterance is nothing but 'translating' it into its language of thought counterpart.

1.3 Heidegger and Winograd

Unlike Descartes and Hobbes, Heidegger — who also stresses the importance of language — derives such an importance not from its relation to cognition, but from its role as one of the fundamental and irreducible modes of human existence: "*Die Rede ist mit Befindlichkeit und Verstehen existential gleichursprünglich*".(6)

In spite of acknowledging its fundamental role, the early Heidegger theorizes very little about language. In his later writings, he develops his views on language by offering metaphors which are supposed to let the reader see how language works. One of the best known of these metaphors (Heidegger 1947, p. 53), describes language as the *Haus des Seins* ("the house of being"). But Heidegger's emphasis is not on the closedness but rather on the 'openness' of this linguistic house: through its metaphorical potential, language enables — rather than constrains — the creative opening of new dimensions of understanding, of existence, and of language itself. From this point of view, the paradigmatic use of language is poetry, which true philosophy is supposed to emulate. In contrast to this, in daily as well as in scientific and technological discourse, the metaphorical potential of language is not enhanced, but used up. For Heidegger, 'cognition', 'theorization' and 'knowledge' — and their characteristic kind of discourse — correspond to partial or abstract ways of relating to the world. If one were to view language as essentially related to these impoverished modes of being, language would thereby be similarly impoverished: it would become a prison rather than an open house, a world-representing mirror, rather than a world-creating power.

Terry Winograd (a pioneer in developing language-using programs) and his associates have recently launched a sweeping criticism of the accepted paradigm in AI, resorting to Heideggerian ideas both as a critical tool and as an inspiration for an alternative (cf. Winograd and Flores, 1986). The key to this critique, as well as to the proposed alternative is a better understanding of the role of language:

> Our central claim in this book is that the current theoretical discourse about computers is based on a misinterpretation of the nature of human cognition and language. Computers designed on the basis of this misconception provide only impoverished possibilities for modelling and enlarging the scope of human understanding. They are restricted to representing knowledge as the acquisition and manipulation of facts, and communication as the transferring of information (Winograd and Flores, 1986, p. 78).

Accordingly, computers should be designed "on the basis of the new discourse about

language and thought that we have been elaborating'', namely:

> Computers are not only designed in language but are themselves equipment for language. They will not just reflect our understanding of language, but will at the same time create new possibilities for the speaking and listening that we do — for creating ourselves in language (Winograd and Flores, 1986, p. 79).

On this view, then, language matters to AI in a very deep sense. For, once a true understanding of the role of language is achieved and implemented in computer design and applications, the latter will no longer be alienated from man. Rather than a threatening or competing tool, the computer will then become man's valuable companion in his quest for creativity and understanding.

2. The Functions of Language

2.1 Language, mind and existence

Whereas all the three conceptions presented above agree in viewing language as crucial for AI, it should be clear that they do so on quite different grounds. Let us now review their similarities and differences.

The first two approaches share an undisguised cognitivist outlook. For them, the most significant human capacity is cognition. At any rate, this is the capacity worth emulating or simulating by means of AI's sophisticated artifacts. Language has whatever significance it has because of its relation to cognition: ''Les langues — said Gottfried Wilhelm Leibniz — sont le meilleur miroir de l'esprit humain''.(7) The two approaches part company, however, in their conception of this ''mirroring'' relation.

For the Cartesian approach, language is nothing but the external manifestation of cognition. Thought itself is pure mental activity. Were it not for the fact that we are corporeal as well as mental beings, we would not need language at all, for we could communicate directly our thoughts to each other. Given the fact that we do not have such a direct access to other minds, we need language for communication, and language — properly used — serves as the main (perhaps sole) indicator of the presence of cognition outside of ourselves. But it plays no role in the cognitive processes themselves.

The Hobbesian approach grants language a much more intimate relationship with cognition. Without language none of the higher cognitive functions (memory, judgment, reasoning) would be possible. According to this approach, it is a mistake to consider language as having only a communicative function. For even a solitary individual, quite independently of his or her communicative needs, needs language. Such a language may be public or private, and it may be conceived either as constitutive of cognitive processes or as an indispensable tool for them. Be it as it may, language has an internal rather than external connection with thought. The following diagram summarizes the similarities and differences between the first two approaches:

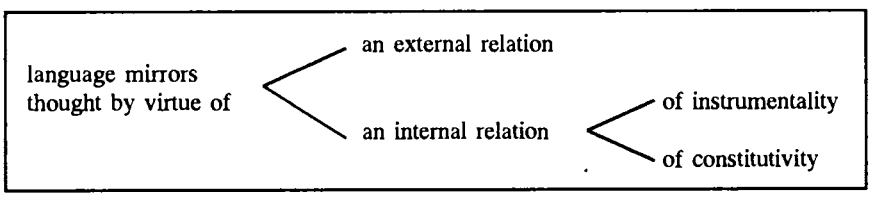


Table 1. *Language and Mind*

Unlike its predecessors, the Heideggerian approach does not locate the importance of language in its relation to cognition. Activity or praxis is the human basic existential condition, and cognition has only a secondary — even distorting — role in it. Language is important because it is directly constitutive of human praxis, not because it serves to transmit, represent, or even be instrumental in the production of knowledge. The very conception of a mind that 'cognizes' the world, represents it in a (linguistic or non-linguistic) 'model', and produces discourse about it, is misleading for this approach. For it presupposes the validity of those very concepts — e.g. 'representation', 'content', 'subject', 'object' — that, according to Heidegger, are useless and pernicious for our understanding of the human being-in-the-world (Dasein). Only a conception of language that rejects such concepts, that severs language's traditional umbilical link with cognition, can capture its real significance. And this is true not only for human beings and their praxis — would add the AI-Heideggerians — but also for their artifacts and *their* praxis.[8]

2.2 The 'place' of meaning

According to the different conceptions about the general significance of language, one should expect to find corresponding differences in the 'location' of language's most significant aspect, namely 'meaning'. For those that relate language primarily to cognition, meaning — whatever it is — should arise from this connection. Thus, for Descartes, it is the thought underlying the words we use that grants them meaning. Locke expressed this view most emphatically by claiming that the meaning of an expression is the 'idea' present in the mind of s/he who utters it.[9] Accordingly, successful communication consists in a satisfactory 'decoding' by the addresse of the ideas 'encoded' in the linguistic sign by the speaker. But the definition of language as having exclusively a communicative function, external to cognition, might as well lead to a different conception of meaning. For, communication being a *social* affair, meaning might be seen as belonging not to the private realm of Lockean ideas, but to the public domain of interaction.[10] On this view, whatever goes on in the minds of the participants in a communicative exchange is irrelevant in so far as their respective behaviors 'make sense' for each other, in the context. Understanding, thus, lies in the adequacy (for whatever reasons) of the participants' reactions to each other's utterances and to the context of the exchange, rather than in appropriate 'decoding'.[11]

For those who see language as inherent to cognition, meaning cannot lie in a correspondence of signs with a non-linguistic cognitive content. It must be somehow explained as belonging in the signs themselves — e.g. in their interrelations, interactions, operational possibilities, etc. It seems thus natural, for this conception, to reduce semantics to syntax.[12]

Given his views on language as one of the non-derivatively significant dimensions of existence, Heidegger too cannot view meaning as lying in a correspondence with 'something else'. Accordingly, he rejects the referential, the Platonic, as well as the ideational and the syntactic views. Meaning, being non-derivative, is for him an irreducible mode of human existence:

> Sinn ist ein Existential des Daseins, nicht eine Eigenschaft die am Seienden haftet, "hinter" ihm liegt oder als "Zwischenreich" irgendwo schwebt. Sinn "hat" nur das Dasein...[13]

2.3. The relevance of use

Language is a very complex phenomenon, that can be investigated from different points of view, each preferring one partition of the phenomenon over the others. For our

purposes here, it is convenient to recall the partition proposed by Charles Morris (1938). Slightly adapted, Morris's distinction between the syntactic, semantic and pragmatic aspects of language (and other semiotic systems) can be summarized as follows:[14]

SYNTAX –Segmenting and recognizing symbols. –Constructing and parsing well–formed sequences of symbols. SEMANTICS –Assigning meaning and reference to individual symbols. –Assigning meaning and reference to sequences of symbols. PRAGMATICS –Using sequences of symbols in a significant manner, in a context of use. –Interpreting sequences of symbols thus used.

Table 2. *Syntax, Semantics, Pragmatics*

Progress in the investigation of language has roughly followed the descending line in this table. Yet, although pragmatics — the theory of language use — is the most recent and less developed of the linguistic disciplines, its relevance for AI, both in its broad and in its narrow conceptions, should not be underestimated.[15] For, in narrow AI, the user/buyer is king. What counts for him or her is how the system *performs*, regardless of the structures that allow it to do so. And in broad AI, although these structures are of interest in so far as they yield insights into how the system performs intelligently, the ultimate test of the system's 'intelligence' lies not in its *possession* of such structures, but in its ability to *use* them appropriately.[16] This, of course, does not detract from the importance of the system's mastery of the relevant structures (e.g. grammatical and semantic rules), for on most theories this is a precondition for 'intelligent' behavior. Yet, exclusive focus on endowing a system with these structures, disregarding varying conditions of use, perforce leads to an automatism in the system's performance that soon reveals its 'stupidity'.

The centrality of use emerges clearly in the three different ways of grounding the importance of language we have examined. Both Descartes and Turing do not mention language structure at all in their tests; they refer only to language use. As for the computationalists, their interest is addressed to the cognitive function or use of language, and only derivatively to its structure. And Heidegger's insistence on language as embedded in and constitutive of the existential context and on the degenerative character of 'propositional' language, detached from that context, is a clear indication of the primacy of use over structure, for him.

2.4 Three kinds of pragmatics

The emphasis on use should refer us to the discipline that is supposed to account for it, namely pragmatics. Although pragmatics has been so far restricted to the study of the communicative uses of language, our examination of the alternative conceptions of the importance of language suggests that there is room for other 'kinds' or 'branches' of pragmatics as well. More specifically, in addition to a pragmatics that deals with the communicative or social uses of language, there is also room for: (a) a pragmatics that deals

with those cognitive or, more generally, mental uses of language that are not related to communication (e.g. in a person's silent reasoning, dreaming, etc.);[17] and (b) a pragmatics that deals with language use as a component of the world or of human existence. In the light of these findings, in principle we may discern, corresponding to the three approaches to language, three types of pragmatics:

SOCIOPRAGMATICS	PSYCHOPRAGMATICS
Investigates the social, other–oriented, uses of semiotic systems. *Example*: the communicative uses of language.	Investigates the mental, "private", uses of semiotic systems. *Example*: the use of language in (silent) reasoning.
Such uses must rely on shared (public) rules and contexts, to ensure comprehensibility.	Such uses may rely on idio–syncratic (private) "rules" and contexts, accessible only to the "thinker".

ONTOPRAGMATICS
Investigates the grounding of language in existence and of existence in language. *Example*: the creative power of original metaphors.

This approach transcends both the social and the mental/subjective dimensions of language use.

Table 3. *Three kinds of pragmatics*

2.5 Orthodoxy and beyond

The predominant tradition in cognitive science and in AI tends to adopt the Fodorian thesis that natural language is not directly involved in thought. Here is an authoritative expression of this position:

> The position implicit in the analysis of this book can be summed up as follows: (1) the generation or processing of symbol structures that are isomorphic with the strings of natural language or with their surface structures (in the linguist's meaning of that phrase) is inessential to human problem solving of the kinds we examine. (2) The internal symbol structures that represent problems and information about problems are synonymous with the linguist's deep structure. If 'language' means deep structure, then language is essential to thinking and problem solving. In sum, paraphrasing Dewey: (1) the surface structure of language and language strings are the garb or clothing of thought, necessary not for the thought but only for conveying it. (2) While linguistic deep structure is not thought, it is necessary for thinking (Newell & Simon, 1972, p. 66).

From this point of view, pragmatics — in so far as it is relevant at all — can only be sociopragmatics. For it can only be involved in the process of 'translating' a natural language input into the deep structure symbols that represent its meaning and that are the

only ones relevant to cognitive processes. As for the 'language of thought' it has no pragmatics because it is not 'used' by thought: it is constitutive of thought.

I surmise that such a quick and wholesale dismissal of pragmatics is damaging for the prospects of AI to develop reasonable language-using programs and for the prospects of cognitive science to produce a satisfactory theory of the mind. The problem of interpretation should be acknowledged in its full complexity, rather than compared to a more or less mechanical or algorithmic decoding. This means contributing to the development of sociopragmatic theory and of AI-applicable versions thereof. But beyond that, it may well be worthwhile for both AI and cognitive science to reconsider the validity of their dogma that precludes the possibility of a direct influence of language on cognition (and hence, of psychopragmatics). It may also be worthwhile for these disciplines to admit that ontopragmatic suggestions might not be irrelevant to their concerns as they seem to assume; for the fundamental existential role of language may well be a factor that should not be overlooked even if you are interested only in simulating or understanding cognition.

3. Sociopragmatics

3.1 The sign

What matters in communication is, of course, 'content'. But content is never directly accessible. It is always conveyed through something else, a 'sign'. The sign may be deemed to be the visible tip of an iceberg, while content lies below the surface and can only be reached through an exploratory process that begins with or at least relies on the 'stimulus' provided by the presence of the sign.

The linguistic sign (a letter or a phoneme, a word, a sentence, a text) has a structure (graphemic, phonetic, syntactic) that must be properly identified if one is to be able to recover its content. A 'grammar' in the linguistic sense is a set of rules that allow for the identification of such structures, at their various levels. One usually think of a grammar as dealing with units called 'sentences'. But the task of grammar extends both below and above sentence level, thus encompassing, respectively, morphology and phonology as well as 'text-grammar'.

To describe the sign as 'visible' or 'given'should not lead one to belittle the difficulties involved in the task of identifying and appropriately structuring it. In fact, though linguistics' first claims to scientificity were based in its achievements in dealing with phonological and then syntactic structure, the problems at these levels are far from solved, especially in a way useful for AI to handle *speech*. In spoken language, segmentation, form identification and concatenation cannot be assumed to have been already performed by 'someone', as it happens in *writing*. Not surprisingly, most AI programs concerned with 'understanding' bypass these thorny problems by dealing directly with written texts.[18] But, as we shall see, the price paid for this strategy is a significant loss in contextual information, which ultimately impairs the program's ability to 'understand'.[19]

It may well be that, as Hofstadter has starkly put it, "The central problem of AI is the question: *What is the letter* 'a'?", because "for any program to handle letterforms with the flexibility that human beings do, it would have to possess full-scale general intelligence" (Hofstadter, 1986, p. 633). No doubt from the way we solve *this* problem of pattern recognition we might eventually learn a lot about the solution of the problem of understanding in general, along the lines suggested by Hofstadter. But rather than

exploring this possibility, I will assume that the sign has been somehow properly identified and structured, and proceed to inquire about how do we get from it to its purported content.

3.2 The problem of interpretation

To interpret is to extricate the 'right' content from a given sign. The question is, what is the 'right' content, and how do we reach it?; the prior question being: "is there such a thing as *the* right content?".

Naturally, different models of interpretation arise from different answers to these questions.[20] The *hermeneutic* model contends that all interpretation is wholly dependent on the interpreter's interests, purposes, and context, so that no such thing as *the* correct interpretation exists. Instead, interpretation is to be viewed as an unending *process* where any 'solution' is to be conceived as a revisable, temporary resting point, itself purpose– and interest–dependent. For this model, content is a construct generated by the interpretation process, being thus neither 'in' the sign nor 'behind' it. The *cryptological* model stresses, on the contrary, the objectivity of content as something that 'lies there', behind the sign, something that can be 'discovered' by correctly applying the semantic rules of the 'code' to which the sign belongs. The *pragmatic* model contends that communicative intention are the primary contents the communicative process is supposed to be all about. From its point of view, semantic 'decoding' is only a step in the process of interpretation. Finally, a number of *causal* models seek to disclose the 'real' as opposed to the 'apparent' content of a communicatively used sign in a deeper layer of motivations or forces, that lies beyond the subjects' conscious intentions and control. The Marxian notion of 'ideology' and the Freudian notion of 'rationalization' illustrate this approach to interpretation.

These models do not necessarily exclude each other, for they can be each construed as addressing primarily a different 'layer' or aspect of content. The cryptographic model focusses on the relatively shallow layer of codified meanings which are the traditional concern of semantics; the causal models focus on the very deep layers of motivation, class–interests, biological drives, survival functions, etc. that are believed to ultimately underly all our meaning assignments; the pragmatic model is concerned with the speakers' communicative intentions as manifested in their utterances; and the hermeneutic model shifts attention from what underlies the iceberg (the sign) to the iceberg–watcher's contribution to the meaning assignment.

Although there are some AI programs that purport to simulate the kind of interpretation I have called 'causal' (e.g. Abelson's IDEOLOGY MACHINE, Weizenbaum's ELIZA, Colby's 'neurotic' program),[21] AI orthodoxy, when dealing with 'normal' interpretation, seems to favor the cryptological conception. On this view, understanding consists in transforming or mapping one representation (the sign) into another (cf. Rich, 1988, p. 296). In Newell and Simon's terms (see quotation in section 2.5 above), the latter representations, the ones relevant for cognition, consist in semantically interpreted (sentential) deep structures.

But semantic interpretation is never sufficient *per se*. Although interpretation may begin with semantic decoding, it cannot end there. For, as pointed out by Rich (1988, p. 296), the interpreting representation must be chosen so as to "correspond to a set of available actions ... so that for each event an *appropriate* action will be performed". And, in order to perform the appropriate action, a system must understand not just the words addressed to it, but the intentions that these words are supposed to convey in their context of use.

In communicative exchanges that employ a strictly regimented language (as any of the extant programming languages), the assumption can be made that each 'sentence', whenever 'uttered', expresses unambiguously the same intention (usually, in these cases, either directives or information-bearing statements). But, when considering usual communication and normal language, such an assumption of transparency and univocity can no longer be made. What characterizes human use of language is its pragmatic flexibility. We not only can, but very often do, say one thing and thereby mean (i.e. intend) another. Metaphor, irony, indirect speech acts, Gricean implicatures and other forms of indirectness (cf. Dascal, 1983, chap. 3) are far more widespread in our linguistic praxis than we are aware of. A system whose capabilities are restricted to semantic decoding would be unable to understand such utterances; hence, it would not be a comprehensive language understander. In order to become one, it would need to be endowed with the means necessary for pragmatic interpretation as well.

An even stronger point can be made on behalf of the need of pragmatics. Even if the use of pragmatically 'tilted' utterances were the exception rather than the rule, the sheer *existence* of such a possibility makes it mandatory for every language understanding system to be able to discern when an utterance is used in a tilted way and when not. Indeed, even the most innocent looking transparent signs can have their meaning turned upside down in an appropriate context of use. Consider a perfectly clear and understandable figurativist picture of a pipe. Once you add to it, as did Magritte, a thought-provoking caption saying "Ceci n'est pas une pipe" (This is not a pipe), the prior 'semantic' decoding of the picture is called into question and must be revised in the light of its new context. Since a 'tilting' context *might* always be the relevant one, the decision that it is not, i.e. that it is appropriate to interpret the utterance in its semantic sense, is itself a pragmatic, not a semantic decision. Transparency, thus, is nothing but a special case of pragmatic, i.e. contextually based interpretation, rather than an example of purely semantic interpretation.

3.3 The pragmatic model

For these and other reasons, language understanding in communication cannot be explained (and, *a fortiori*, implemented in a system) without taking into account the middle layer of intention-related content, i.e. without a theory of pragmatic interpretation. Let me outline some elements of such a theory.

To make a long story short, the secret of our success in understanding reasonably well both transparent and indirect utterances lies in our ability to judiciously combine semantic and contextual information. The steps involved in this process may be systematized as follows:[22]

(a) Assume, tentatively, that speaker's meaning (SM) = utterance meaning (UM).
(b) Check for fit: could the speaker, given what you know about him/her and about the context of utterance (i.e. given the relevant 'second channel' information available to you), have meant UM, only UM, and all of UM?
(c) If yes, endorse assumption (a) and proceed to utilization procedure.
(d) If not, provide diagnosis of misfit. Format: 'Speaker could not have meant UM/only UM/all of UM [choose one], because **** [specify reason]'.
(e) Check for non-intentional causal explanation[23]:
    (1) Are there indications of a general breakdown in the communicative

framework?

(2) Are there indications that the speaker is not aware of specific requirements of the situation and/or of specific rules that apply to it?

(3) Are there indications that the speaker cannot comply with such requirements and/or rules?

(4) Are there obvious, readily correctible mistakes in the utterance?

(5) Is there a misunderstanding, either by the speaker or by you?

(6) Is there any other plausible non–intentional causal explanation for the misfit?

(f) If a non–intentional causal explanation is available, proceed according to the specific procedure for the type of explanation found.

(g) If no non–intentional causal explanation is found, activate procedure for generating, on the basis of the diagnosis in step (d), one 'near' alternative meaning. If such a meaning has not yet been tested, check it for fit, as in (b).

(h) If the selected alternative meaning fits, endorse it, assign to it an appropriate degree of indirectness, and proceed to utilization procedure.

(i) If the alternative selected meaning does not fit, return to step (g)

Table 4. *Pragmatic interpretation*

Notice that the passage from semantic to pragmatic interpretation is performed through a *negative* question: "Is there any reason *not* to accept the semantic reading as appropriate, in the context?". An affirmative answer concludes the process, while a negative one serves as an indication (a *cue*) that further interpretive effort is needed.

I wish I could report that I have followed Sloman's suggestion to the effect that the best way to put to test philosophical speculations (or, for that matter, any theory) is by "expressing them in some sort of computer program or at least in a sketch for a design of a working program" (Sloman, 1978, p. 97). But I have tried to formulate the above steps in such a way that they can be easily represented in a flowchart (cf. Dascal 1983, pp. 150–151), which in turn might inspire some AI researcher to write a program along such lines.

3.4 The roles of context

The key role in the pragmatic drama is played by that ubiquitous and slippery character, the 'context'. Our iceberg, in fact, is never isolated (if it were, this significant fact and should be marked as 'null context'). Any sign is always surrounded by circumstances that must be taken into account in its interpretation. It is surrounded by other 'visible' signs, objects, and events with which it stands in syntagmatic relations. Usually invisible, but no less important, is the set of paradigmatic relations (analogies, similarities, oppositions, etc.) which the interpreting system's memory must bring to bear on the process of intepretation. Furthermore, the whole picture is drawn against a blurred, inarticulate 'background', whose role in interpretation should not be neglected either. All this, and much more, is comprised in the 'context' normally used for understanding even the most prosaic utterance.

One of the imperatives of sociopragmatics is to try to analyze and organize the different kinds and roles of the context. One such analysis is provided in the following diagram (cf. Dascal and Weizman 1987):

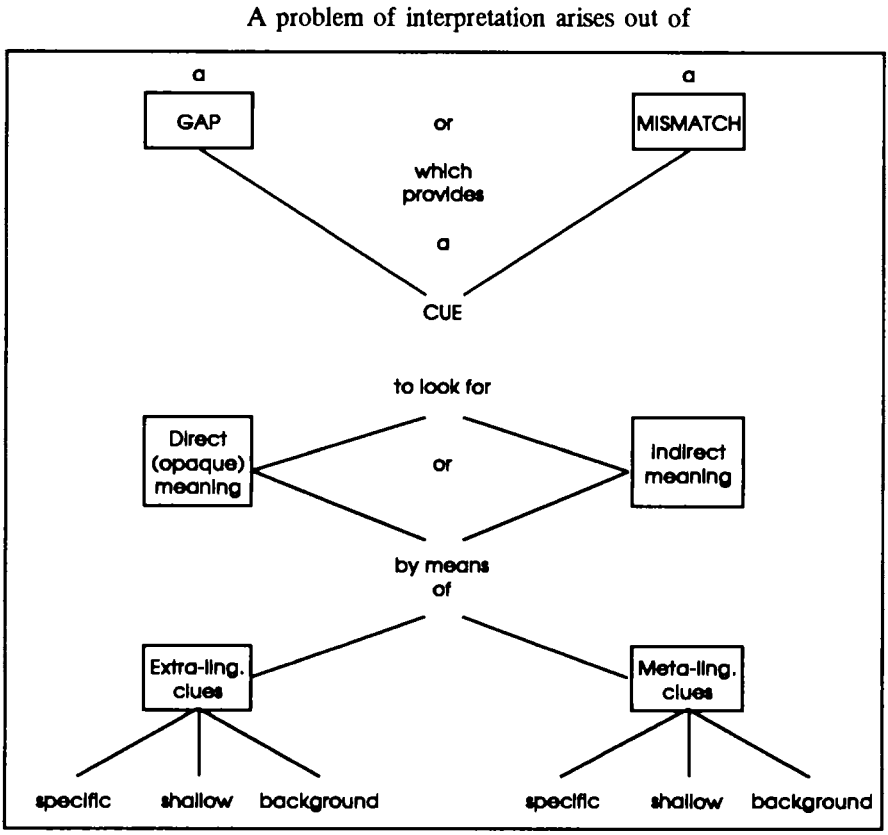


Table 5. Cues and Clues

Let me single out here for comment only the janus-like role of the context as providing both *cues* and *clues* for interpretation. In the former capacity, it is the *existence*, say of a mismatch between the semantic reading and some component of the context that signals the need for further interpretive effort. In the latter capacity, the *nature* of the mismatch leads to the formulation of the new interpretive hypothesis. For instance, if replying to my question "Why did John spank Mary?" you reply: "How much are two and two?", you do not, ostensively, match my question with a straightforward answer. This is a cue to the effect that I should not interpret your utterance as intending to convey a real question. How then should I interpret it? One clue lies in the fact that your 'question' is not an ordinary question, but rather one that has an obvious answer. I am entitled then to hypothesize that your utterance means "You know very well the answer" or the like. Notice that my interpretation of your utterance makes use, among other things, of the specific meta-linguistic context (my preceding utterance), of the background meta-linguistic context (my knowledge about what kind of speech act should follow a question), and of our presumably shared background extra-linguistic knowledge of arithmetic.

3.5 Multi-channel information and impoverished contexts

In interpretation, we make constant use of all possible kinds of contextual information available. Curtailing any of the channels of such information is likely to impair understanding. When such a curtailment is a permanent or anticipated condition, compensations are developed by the communicators, in the form of an enhanced use of, and sensitivity to, the remaining channels. But there are limits to compensation, for there are kinds of information that cannot be transmitted by alternative channels: sophisticated as they may be, verbal descriptions of acoustic, visual or tactile patterns are not the equals of their direct perception. Furthermore, the attempt to squeeze all the information required for interpretation in one single channel would soon overload that channel, rendering it ineffective.

Written communication takes place in such an impoverished context, where most of the non-verbal clues for interpretation are not available. In so far as such clues are relevant for understanding — and they certainly are — they must somehow be packed into the text itself. That is to say, the text must be made independent of them. A fully explicit or self-contained text woud be one for which all other forms of context-dependence were eliminated too. But this would involve packing into the text such a huge amount of information that would make it not more, but much less intelligible.(24)

Barring, then, the possibility of self-containment even in written texts, the confinement of communication to writing amounts to the suppression of contextual information necessary for interpretation.(25) Since the carefully contrived situation of the Turing test is impoverished precisely in this way, no wonder that 'passing' it cannot serve as a criterion for a system's *pragmatic* understanding of what is 'said' to it. For, without the possibility of contextual checking, there is no way in which the communicative intention of a speaker can be attached to his or her words.

3.6 The spirit is willing but the flesh is weak

Let me illustrate some of the complexities involved in sociopragmatics by an example, drawn from an anecdote that was very popular in the early days of machine translation. Given the English sentence "The spirit is willing but the flesh is weak", the program translated it into Russian. Since none of those attending the demonstration understood Russian, the only way to check the program's performance was to have the Russian sentence translated back into English. The result was: "The vodka is good but the meat is rotten".

Setting aside the two-language aspect of the story, let us take the second sentence to represent the program's interpretation of the first. The question I want to address is this: How did the program reach this presumably wrong reading? Suppose it begins from left to right. The word *spirit* has, in the program's dictionary, two different senses, say "soul" and "alcohol". One of the latter's hyponims is "alcoholic beverage". This concept, in turn, is likely to be construed, in different cultures, in terms of different *prototypes*. In Britain, the prototype would be whiskey, in Mexico, tequila, in Russia, vodka. Some such path of semantic relations, suitably modified by the extralinguistic cultural context, must have been followed by the program in order to distil *vodka* out of *spirit*.

But why didn't it follow the other possible semantic path, beginning with "soul"? The next word, *willing*, which is subsumed under the concept "mental attitude" fits this path, but not the alcohol one. In the absence of context, the program's choice seems to be

entirely unmotivated. So, perhaps it construes the 'conversation' as taking place in a restaurant. This fact activates in its memory the *script* RESTAURANT, where beverages, food, etc. are prominent. "Soul" would clearly be much less of a match for this script than "alcohol" (provided the latter immediately leads to "alcoholic beverage" rather than to "disinfectant", which would be more in line with the script HOSPITAL). Having thus chosen "alcohol", on the strength of the presumed extralinguistic specif contextual clue, the program must get rid of the unfitting "mental attitude" reading of *willing*. This word has a "positive" connotation, and the program knows that sometimes words are used not to convey their denotation, but only their connotation (meta-linguistic background context). It thus retains *good* as the interpretation of *willing*. Now, the ambiguity of *flesh* ("body" vs. "meat") can be solved according to the preceding choices. What about *weak*? Why to choose the path "of poor quality" rather than, say, "soft" — which would be appropriate to *meat*? Here the semantics and pragmatics of the connective *but* intervene (cf. Dascal & Katriel, 1977). There must be a *contrast* of some sort (not necessarily semantic: cf. "He is a Republican, but honest") between the two clauses linked by *but*. "Soft" is rejected because it would not provide such a contrast, whereas "of poor quality" contrasts nicely with the positive evaluation previously read into *willing*.

Hardly anybody, besides our poor program, would provide such an interpretation for the input sentence. And yet, we recognize it as a *possible* interpretation, albeit a very unlikely one. Otherwise, we wouldn't laugh; we would only be puzzled. In understanding the whole story as a joke, we *reframe* it accordingly.[26] The computer's clever interpretation now is perceived as the result of making all the wrong interpretive choices. In a Freudian vein, we laugh because we are reassured of our superiority over this apparently clever, but in fact quite stupid machine...

It is not easy to pinpoint the reasons for the program's going wrong. Perhaps simply they lie in the impoverished context provided to it. Or in the fact that it lacks a list of popular sayings, which could have led it to try to find an equivalent saying in the Russian list, rather than "translating". Or else in the program's inability to shift scripts, paths, or readings once they have been 'activated'. Were this be done by providing an overall 'meta-script', the program might then be stuck with it, unable to 'jootse' (jump out of the system — Hofstadter, 1986, p. 541), e.g. by moving from the "serious" to the "joking" frame of mind. Whatever the reasons, and the possible ways of overcoming them, it is clear that programs that purport to understand language have yet to learn a lot of sociopragmatic tricks.

4. Psychopragmatics

4.1 Dreams

Since I have concluded the preceding section with jokes, I will begin the present one with a brief discussion of language use in dreams.[27] For, as Freud persuasively argued, there is a close similarity between these two phenomena. We employ similar techniques to manufacture jokes and dreams. In both cases, 'condensation' and 'displacement', for instance, produce a camouflaged expression of some underlying (unconscious) content. Freud also noted that words are particularly suited for such a camouflaging process, because of their multiple ambiguity and richness of associative links. Consequently, language provides a powerful bag of tricks, which is constantly exploited in both jokes and dreams, along similar lines.

Yet, Freud was aware of a crucial difference between the two phenomena. Whereas a dream is "a completely asocial mental product", a joke is "the most social of all the mental functions that aim at a yield of pleasure". The former is not made "with the intention of being understood", and it has "nothing to communicate to anyone else". The latter "has the unmistakable aim of evoking pleasure in its hearers". Whereas the dream need not "set any store by intelligibility" and "must actually avoid being understood, for otherwise it would be destroyed", the joke is strictly bound by a "condition of intelligibility". The former is free to "make use of the mechanism that dominates unconscious mental processes, to the point of a distortion which can no longer be set straight". The latter "may only make use of possible distortion in the unconscious through condensation and displacement up to the point at which it can be set straight by the third person's understanding" (Freud, 1900, p. 341; 1905, p. 179).

This contrast between an eminently social activity and an entirely 'private' one epitomizes, in so far as both activities employ language, the difference between the sociopragmatic and the psychopragmatic points of view. Communication, no matter how indirect and devious, is a social game of cooperation, where lack of mutual intelligibility means a net loss for all parties involved. Intelligibility is ensured by adherence to public, shared sociopragmatic norms and procedures that regulate the communicative use of language and the interpretive practice.

In a purely mental activity, by contrast, language is not used to convey 'messages' from one person to another. Accordingly, the user is free to employ language as s/he pleases, and the norms and constraints — if any — that regulate the mental uses of language cannot be assumed to be identical to their sociopragmatic counterparts. In some mental activities (e.g. problem-solving), recoverability of content may be a desideratum. But it can be eventually achieved not by respecting the public linguistic conventions, but rather by the creation of an idyosincratic, private 'code' (as do some mnemonists, for exemple). In other mental activities, such as dreams, the desideratum (from the point of view of the dreaming subject) is rather — according to Freud — non-recoverability: only the skilled psychoanalyst knows how to recover the carefully hidden content. The dreamer makes use of linguistic materials as building blocks for her dreams, without respect for syntactic, morphological, semantic, or sociopragmatic rules. The mechanisms of condensation and displacement can put together or dissociate words, phrases, or whole utterances based on associations of any kind — e.g. phonetic, semantic, contextual. And yet, to be 'acceptable' to consciousness, the result — the dream — must display a remarkable inner coherence, and the words therein must preserve an appearance of intelligibility.

Were we to program a computer to simulate dreaming, we should somehow gather a better understanding of the psychopragmatics of language-use in dreams. As far as I know, although dreaming has been recognized as an important cognitive process,[28] AI has not yet tried to simulate it. But one of its major aims has always been to simulate other cognitive processes, such as reasoning. What is surprising is its lack of awareness of the need to take into account the psychopragmatic role of language in such processes. To this we turn now our attention.

4.2 Language and 'bad' logic

According to the cognitivist orthodoxy, once a linguistic formulation of a problem is 'understood', i.e. mapped "from its original form to a more useful one" (Rich, 1988, p. 298), the original form to a more useful one" (Rich, 1988, p. 298), the original, less useful

form is discarded and should no longer interfere with the problem-solving process. On such a view, pragmatics plays a role — if at all — only in the mapping stage, i.e. *qua* sociopragmatics. Since the reasoning involved in the problem-solving process is to be entirely performed in the "more useful", language-free representational space, there is no room at all for psychopragmatic considerations.

The trouble is that somehow language keeps 'interfering' in reasoning. This interference shows up in people's stubborn deviances from the 'right' canons of logic. Consider for example the following reasoning task, investigated by P.C. Wason. You are going to be shown four cards. You know that each card has a letter on one side and a number on the other and that either side may be visible to you. The cards will lie on a table, so that you will be able to see only one side of each card. Now you are shown these four cards:

A **F** **3** **8**

together with this statement about them:

IF A CARD HAS A VOWEL ON ONE SIDE
THEN IT HAS AN EVEN NUMBER ON THE OTHER SIDE

Your task is to pick up those cards, and only those cards, which you would need to turn over in order to find out whether this statement is true or false (Wason, 1981, p. 412). Take up a moment to solve the problem and write down your answer.

Variants of this relatively simple problem have been intensively tested, and the vast majority of subjects provide the wrong answer. They select the cards A and 8 to be turned over. But the logically correct answer is A and 3, since only these can falsify the conditional statement. Why do people persistently make this logical mistake? One answer might be that their logic is at fault, i.e. that they are not familiar with the logical properties of conditionals. Another hypothesis, favored by Wason in a later article (Wason and Evans 1975), is that the logic is O.K. and what causes the mistake is the interference of language: people simply try to 'match' the symbols in the cards with those mentioned in the conditional sentence, regardless of whether a selection based on this principle is relevant or not to the solution of the problem. That is to say, the linguistic representation of the problem is not simply discarded once it is understood, giving way to a pure logical representation, upon which reasoning operates. Rather, the linguistic representation continues somehow to play a role in the cognitive process, a role that influences its outcome, producing a 'matching bias'.

Results such as these (found also in other experiments)[29] may provoke different reactions, especially on AI designers. One such reaction is to say that, since the interference of language leads to such grave mistakes in problem-solving, stringent precautions should be taken in order to safely insulate the system's reasoning from any linguistic influence. Another, provenient from those interested not so much in 'correct' reasoning but rather in reproducing the results of actual human reasoning, might consist in proposing to replace standard by some kind of non-standard logic (e.g. non-monotonic, paraconsistent, or relevance logics), as a move towards a better fit with human 'natural' logic. As an alternative to these two drastic strategies, the psychopragmatic outlook suggests that designers should first question the orthodox view that language is or should be external to thought, and proceed to investigate more closely its role in cognitive processes, which may

well turn out to be, in general, not so negative and error-inducing as in the example discussed.

4.3 Language as a context of thought

In some mental processes, language's influence on the process may be due to its lying at its very heart, as claimed by constitutivists like Hobbes. On the Fodorian, 'syntactic' version of computationalism, where the medium of thought is not directly natural language, the latter can influence it only through its *formal* properties. On this view, an explanation for the 'matching bias' phenomenon might run as follows. The limited vocabulary of Wason's problem includes the following symbols: (i) the names of individual letters and numbers, (ii) the names of four general categories (vowel, consonant, even, odd), and (iii) the logical connectives. The symbols in class (iii) are governed by the rules of logical 'syntax', which determine the logical structure of the problem. The relevant relations between the linguistic symbols in classes (i) and (ii) is also of a 'syntactic' nature, expressed in the rule: Every symbol in (i) is necessarily an instantiation of one of the four categories in (ii). An appropriate solution to the problem requires putting this linguistic syntax to the service of the logical syntax. Yet, for most subjects, the linguistic syntax 'refuses' to accept this subordinate role. For these subjects the instructions mean not only: "Determine those instances that verify or falsify the conditional", but also: "Find instantiations of the categories named in the conditional statement". Each subject's solution for the task will depend on the relative strength, for him, of these different and conflicting readings. The phenomenon of interference between the two readings will thus be explained, on this view, as a matter of the interaction between two competing formal structures.

While this pattern of explanation may be suitable for accounting for the influence of language on thought in some well-defined domains (as is Wason's problem), it seems to unduly restrict such an influence. For only the rigidified relational structures of language are allowed to play a role in or interfere with mental computations. This conception corresponds to the Bergsonian thesis that verbal categories are rigid abstract patterns, unfit to capture an essentially fluid reality — a view echoed in Hofstadter's (1986, p. 561) claim that "a frozen verbal phrase is like a snapshot that gives a perfect likeness at one moment but fails to show how things can slip and move".

But language's potentialities go well beyond this frozen snapshot of its nature. Recall Heidegger's talk of the power of original metaphors. Even so-called frozen metaphors are far more complex and rich than usually thought: they form part of a huge underlying network of "metaphorical concepts" in terms of which we organize most of our experience and talk (cf. Lakoff & Johnson 1979, and Lakoff 1985). Language is in fact the most impressive source we have for associations, comparisons, analogies and other 'slippery' connections necessary for creative thought. Language's characteristic feature is that it is partly rigid (formally codified) and partly loose. To restrict its role in cognition only to the former part is the psychopragmatic equivalent of the sociopragmatic impoverishment of context (cf. section 3.5). This attitude amounts to assigning to language only a psycho*semantic* role (cf. Fodor, 1987). As opposed to this, a psycho*pragmatic* attitude towards language permits to conceive of it not necessarily as a formal vehicle of thought (or interfering in thought), but as a rich, open-ended, flexible, suggestive, fluid and vast *context* of mental processes. From this perspective, mental processes may or may not be linguistically constituted. What matters is that, whatever their constitutive structure, a sea of *linguistic* icebergs — from tip to bottom — surround and support them.

5. Ontopragmatics

5.1 The programme

The approach to AI I have dubbed 'ontopragmatic', as instantiated in Winograd & Flores's (1986) book, criticizes the cognitivist computationalist orthodoxy along similar lines. It stresses the 'blindness' of rigid conceptual schemes, the indispensable role of an ever-present and inarticulate 'background', the open-ended, creative aspects of language. Some of the Heideggerian theses it puts forth are (pp. 32ff.): "Our implicit beliefs and assumptions cannot all be made explicit"; "Practical understanding is more fundamental than detached theoretical understanding"; "We do not relate to things primarily through having representations of them"; "you cannot step back and reflect on your actions"; "Every representation is an interpretation"; "Language is action".[30]

This approach goes one (big) step beyond both the sociopragmatic and the psychopragmatic viewpoints when it becomes truly ontological, contending that "*nothing exists except through language*" (Winograd & Flores, 1986, p. 68; italicized in the original). It further purports to ground AI's "new conception of design" precisely in this ontological claim" "...this apparently paradoxical view (that nothing exists except through language) gives us practical orientation for understanding and designing computer systems" (p. 69).

No doubt the emphasis on the ontological dimension of design draws attention to the important fact that "the domain created by a design is a domain in which people live" (p. 179). It also rightly insists that "[design] constitutes an intervention in the background of our heritage, growing out of our already-existent ways of being in the world, and deeply affecting the kinds of beings that we are" (p. 163), and, consequently, that "our tools are part of the background in which we can ask what it is to be human" (ibid.). The dominant tradition in AI is criticized precisely on account of the type of human self-understanding it forces upon us:

> As we work with devices whose domains of action are based on an interpretation of 'data', 'goals', 'operators', and so forth, we develop patterns of language and action that reflect these assumptions. These carry over into our understanding of ourselves and the way we conduct our lives. Our criticism of descriptions of human thought as 'decision making'and language understanding as the manipulation of representations ... reflects a deeper concern with the discourse and actions than that generated by a rationalistic interpretation of human action. Computer systems can easily reinforce this interpretation, and working with them can reinforce patterns of acting that are consistent with it (Winograd & Flores, 1986, p. 178).

Yet, in spite of these dangers, the authors do not recommend that we abandon computers altogether. They believe that it is possible to "create computer systems whose use leads to better domains of interpretation" (p. 179), i.e. systems that reflect a pre-understanding which is fully aware of the ontological implications of design, of the roles of background, breakdowns, blindness, and readiness-to-hand — in short, truly Heideggerian computer systems.

The only illustration of such a system sketched in the book I am referring to is that of a managerial system. The fundamental contention is that management should not be viewed as having to do primarily with setting goals, choosing between alternative courses of action, making and implementing decisions, and solving problems. Accordingly, an organization should not be viewed as structured in such terms, and computer 'aids' to management

should not be conceived as supporting such tasks. An organization is rather a "network of commitments" (p. 150). And such commitments arise in "conversation":

> ...the key elements are the conversation among the affected parties and the commitment to action that results from reaching resolution (p. 151); the conversational dimension permeates every realm of coordinated activity (p. 158); the principal characteristic of deliberation is that it is a kind of conversation (in which one or many actors may participate) (p. 149).

The linguiatic dimension is thus raised to the position of the fundamental mode of existence of organizations and to the basic concern of management. It is linguistic acts that bring into existence problems, relevant alternatives, and commitments to courses of action.(31) Accordingly, "the most fruitful domain for understanding and facilitating management" is the "domain of speech acts and conversation" (p. 144), and the new tools the authors propose deal essentially with this domain:

> Computer-based tools can be used in requesting, creating, and monitoring commitments. They can provide answers to the question 'What do I need to do?', or as we prefer to put it, 'What is the status of my active commitments?' (p. 158).

Unfortunately, this sample illustration, even if described in more detail, cannot serve to test the viability of applying Heidegger's views to AI. For in its essentials, it obviously resorts to the less arcane theory of speechacts, rather than to the Heideggerian conceptual apparatus, even though it intersperses here and there elements of the latter. We shall return to this point in section 5.3. At present, let us consider not a particular illustration, but the more general issue of the possibility of implementing the full range of Heideggerian ideas in AI.

5.2 Heidegger against Heideggerian AI

Eventually, computer systems could be built so as to incorporate some of Heidegger's theses. But could they be designed so as to be attuned to the full import of the Heideggerian ethos?

Heidegger's 'ontopragmatization' of language is based on a thorough critique of the tendency to transform the 'theoretical' or 'propositional' mode of language use into the dominant or primary one. According to him, the assertive mood — on which the 'theoretical' use of language is centered — is derivative from a more basic use of language: "Zuhandenheit" or praxis. In this basic form of use, an utterance is naturally embedded in the wholeness of a concrete experience, and it is unproblematically and immediately understood as such. The propositional or assertive uses of language arise only as a result of a breakdown or problematization, whereby instead of relating to the impact of the concrete thing itself, we react rather to the "cognitive content" of the words. The result is an "inversion" of perspective, whereby the originally clear understanding of the belonging of the thing to its concrete context disappears giving way to hidenness:

> Das in der Vorhabe gehaltene Seiende, der Hammer zum Beispiel, ist zunächst zuhanden als Zeug. Wird dieses Seiende "Gegenstand" einer Aussage, dann vollzieht sich mit dem Aussageansatz im Vorhinein ein Umschlag in der Vorhabe. Dass *zuhandene Womit* des Zutunhabens, der Verrichtung, wird zum "*Worüber*" der aufzeigenden Aussage...*Durch* die Hinsicht und *für* sie wird das Zuhandene als Zuhandenes verhüllt.(32)

This implies not only a subversion of language use, but also a loss in existential status: the "objectified" thing now becomes "cut" from its concrete existential mode, it becomes a dead abstraction: "Es ist bezüglich seiner Moeglichkeiten der Artikulation von Verweisungsbezügen von der Bedeutsamkeit ... abgeschnitten".(33)

Heidegger views the development of Western metaphysics, coupled with the development of science and technology, as a continuous expansion of the theoretical mode, which gradually becomes the dominant one, and aspires to a monopoly on truth. And he is highly critical of this trend, because, for him, within a realm of 'truth' thus defined, Man looses the possibility of confronting his Self: "*Indessen begegnet der Mensch heute in Wahrheit gerade nirgends mehr sich selber, d.h. seinem Wesen*" (Heidegger, 1962, p. 27).

This basic orientation of the Heideggerian critique is embarrassing (to say the least) for any proposal of an alternative programme for AI based on Heidegger. For AI, whatever its pretentions to become 'humanized', remains a typical instantiation of the theoretical/technological mode of cognition and discourse. What is 'designing' a system, if not using a set of abstract concepts and procedures in order to specify the conditions that a concrete product of human activity must satisfy? Under these circumstances, how can one expect that the product of such an activity, the implemented system, will be devoid of the dehumanizing characteristics of its mode of production? From a Heideggerian perspective, the likelihood is that our computer 'companion', even if designed according to allegedly Heideggerian principles, instead of offering new possibilities for "creating ourselves in language" — as claimed by Winograd & Flores — will perpetuate the monopoly of that mode of discourse that prevents such possibilities to arise.

5.3 What can be rescued?

It would be premature to assess the prospects of the ontopragmatic programme in AI, especially on the basis of the sole examination of one of its early products. The following remarks should be taken only as preliminary evaluative comments.

To my mind, one of the main advantages of bringing pragmatic considerations to bear on ever more fundamental aspects of existence is its effect on the notions of cognition and knowledge in general, and on their use in AI in particular. The traditional conception of an objective world "out there" which we cognize by virtue of "representing" it "in our heads" (or in a system's "head") is effectively undermined by the ontopragmatic critique. Accordingly, it must be replaced by a 'pragmatic' conception of knowledge and cognition, where the relevant criterion is not accuracy of representation, but the ability of the 'knower' to play successfully the social game of justification. I have argued for the relevance of this conception to AI elsewhere (Dascal, 1989).

In such a game, the notion of commitment certainly plays a central role. But to analyze it in terms of speech-act theory does not do justice to its full range, especially to its existential aspects, for which the related — but different — notion of "involvement" is more pertinent (cf. Katriel and Dascal 1989). Furthermore, translating the ontopragmatic insights into the idiom of speech-act theory amounts to reducing them to what many consider to be the centerpiece of sociopragmatics.[34] This, in itself, can hardly be blamed. But it might imply the loss of potentially irreducible, specific and original *onto*pragmatic contributions, which in turn would undermine the hypothesized status of ontopragmatics as a separate branch or kind of pragmatics.

Furthermore, the emphasis on its sociopragmatic side leads to overlook the psychopragmatic implications of the ontopragmatic outlook. The phenomenon of the 'blindness' induced by linguistic and conceptual structures directly affects cognition, regardless of communication. No less important, from a psychopragmatic point of view, is the creative potentialities latent in linguistic and conceptual structures, which the Heideggerian poetic view of language has highlighted.

It remains to be seen how well the ontopragmatic point of view — which need not be restricted to a Heideggerian outlook — will fare as a new alternative in AI. Whatever the outcome, its contribution so far to the other two points of view are not negligible.[35]

* * *

There is no conclusion to this paper, for it has been a *tour d'horizon* of the main approaches to language relevant to AI, as I see them. Though defenders of each approach tend to exclude the others, they are not intrinsically incompatible with each other. Some judicious combination of the full power of all of them might, for all we know, pave the way for designing machines that use language in a more understanding way, in a more thought-sustaining way, and in a more ready-to-hand way.

Notes

(1) My title paraphrases that of Hacking's book (1975). Although the answers to *his* question are less obvious, their exploration reveals important basic differences in approaches to meaning and cognition. For an attempt to set up systematically these differences as they appear in the 17th and 18th centuries, see Dascal (Forthcoming c). Perhaps not surprisingly, the different assumptions regarding the importance of language for AI correlate with different philosophical conception of the role of language, as will become apparent in what follows.

(2) "...they could never use speech or other signs as we do when placing our thoughts on record for the benefit of others. For we can easily understand a machine's being constituted so that it can utter words, and even emit some responses to action on it of a corporeal kind, which brings about a change in its organs; for instance, if it is touched in a particular part it may ask what we wish to say to it; if in another part it may exclaim that it is being hurt, and so on. But it never happens that it arranges its speech in various ways, in order to reply appropriately to everything that may be said in its presence, as even the lowest type of man can do" (HR, p. 116).

(3) "...it has never yet been observed that any brute animal reached the stage of using real speech, that is to say, of indicting by word or sign something pertaining to pure thought and not to natural impulse" (Kenny, p. 245).

(4) "For while reason is a universal instrument which can serve for all contigencies, these organs have need of some special adaptation for every particular action. From this it follows that it is morally impossible that there should be sufficient diversity in any machine to allow it to act in all the events of life in the same way as our reason causes us to act" (HR, p. 116). A few paragraphs thereafter, referring to animals, Descartes says that "they have no reason at all, and that it is nature which acts in them according to the disposition of their organs, just as a clock, which is only composed of wheels and weights is able to tell the hours and measure the time more correctly than we can do with all our wisdom" (HR, p. 117).

(5) In fact, the test (if passed) would provide an indication of a *high* level of intelligence. For it requires that the machine be able to *deceive* the interrogator. The inability to lie or fake behavior is taken to be an indication of the limitations of whatever intelligence some animals have. In this respect, since it is primarily a test of the ability to deceive, the Turing Test should be compared not with Descartes's claims about language use, but rather with his Great Deceptor, the Evil Genius.

(6) Heidegger (1927, p. 161), italicized in the original. "Discourse is existentially equiprimordial with state-of-mind and understanding" (Engl. transl., p. 203).

(7) "Languages are the best mirror of the human mind" (NE, 3.7.6; p. 313).

(8) A recent ad in TIME Magazine tries to cater for the favor of all those who share the cognitivist outlook, regardless of their precise position in the diagram above, by combining the terms 'picture' and 'counterpart' in its description of the relationship between language and thought. Here is part of the text: "Language is the picture and counterpart of thought. How will man and machine work together in the future? This intriguing philosophical and functional question has a promising answer: with the spoken word." The rest of the ad mostly insists on communication ("Man will talk to Machine and Machine will respond – – for the benefit of Man"), though it hints at a use of language in machine "cognition" itself ("...developing sophisticated new word-based systems"). The "work together" suggestion, if set apart from the rest, could even be favored by Heideggerians.

(9) "*Words, in their primary or immediate signification, stand for nothing but the* ideas *in the mind of him that uses them...*" (*Essay*, 3.2.2; emphasis in the original).

(10) Alternatively, for those unhappy both with the subjective and with the social relativization of meaning, there is always the possibility of 'locating' it in a Platonic, separate realm, or else in the 'objective' reference of language to the world.

(11) On one of its versions, which seeks to minimize if not to eliminate the role of the 'mental states' of the interacting subjects, this conception has been put forth by the later Wittgenstein. On another — the 'dialogical' doctrine of Martin Buber — meaning belongs primarily to the inter-subjective domain (the "Zwischenmenschlich"), to be conceived not only as the meeting place of pre-formed 'subjects', but as the sphere that gives rise to and shapes them.

(12) Such a move is made by Fodor regarding the 'language of thought' (cf. sections 1.2 and 4.2 of this essay). It has been, to some extent, prefigured by Leibniz in the 17th century (see Dascal, 1987d).

(13) Heidegger (1927, p. 151). "Meaning is an *existential* of Dasein, not a property attaching to entities, lying 'behind' them, or floating somewhere as an 'intermediate domain'. Dasein only 'has' meaning..." (Engl. transl., p. 193).

(14) For a criticism of Morris's trichotomy, and for proposed revisions in the conceptions of semantics and pragmatics, see Dascal 1983, chap. 1.

(15) Broadly conceived, AI is a discipline that endeavors to use computer programs "as tools in the study of intelligent processess, tools that help in the discovery of the thinking-procedures and epistemologial structures employed by intelligent creatures" (Boden, 1977, p. 17). Narrowly conceived, it is a technical activity whose aim is to create programs that perform intelligent tasks for specific purposes.

(16) From this perspective, 'intelligence' should be construed as an *adverb* or an *adjective* that qualifies certain types of activity or behavior, rather than as a *noun* referring to some alleged 'faculty' (cf. Ryle 1949).

(17) Such 'inner' uses or effects of language are not easy to investigate, to be sure. I have argued for the 'right of existence' of this branch of pragmatics in Dascal (1979, 1983) and have produced some examples of its importance in Dascal (1985, 1987a).

(18) "Written language is a fairly recent invention and still plays a less central role than speech in most activities. But understanding written language (assuming it is written in unambiguous characters) is easier than understanding speech. To build a program that understands spoken language, we need all the facilities of a written language understander as well as enough additional knowledge to handle all the noise and ambiguities of the audio signal" (Rich, 1988, pp. 295–296).

(19) The title of Marguerite Duras's novel *L'amante anglaise*, where one would look in vain for an English lover, is a deliberate misnomer. It mis-segments and mis-spells what should be, instead, written (presumably) as "La menthe en glaise". On the significance of this clue for understanding the novel, see Dascal & Weizman (1990). In the light of this striking example, Hofstadter (1986, p. 589) is certainly right in contending that "translating a novel's *title* may sometimes be the most challenging aspect of translating the whole novel".

(20) For a survey and comparative analysis of different models of interpretation, see Dascal (Forthcoming b). For a discussion of the hermeneutic and pragmatic models, see Dascal (1987b).

(21) For a criticism of the IDEOLOGY MACHINE, see Boden (1977, pp. 66–75) and Dascal (1981, pp. 164–165). For an analysis and criticism of Colby's artificial neurosis, see Boden (1977, pp. 21–63). For a criticism of ELIZA, see Boden (1977, pp. 106–111). In general, programs that adopt as the sole basis for interpretation a given (causal) framework fail because they exaggerate the conceptual rigidity they are simulating. Even a stubborn communist and an obsessive neurotic are able, on occasion, to shift to other modes of interpretation. The programs, unable to do so, quickly stumble into stupidity (see Abelson, 1975, pp. 274ff.). It should be mentioned that there are some programs that try to go beyond the cryptological model. For instance, Wilensky's (1981) story–understanding program PAM, which seeks (contextual) explanations for the events described in a story in terms of plans, scripts, and goals of the characters involved. Still, the structures embedded in such programs seem to me to be far too rigid to allow for real pragmatic interpretation.

(22) Cf. Dascal (1983, pp. 149–152) for details. These steps are presented in a traditional sequential form. With the development of parallel distributed processing or connexionist techniques, it may be possible to implement the processes here described in a non–sequential way. This may eventually require a reorganization of the model.

(23) The causal explanations referred to here are to be distinguished from the 'deep structure' causal aspects of content mentioned above. They are intended to be such pedestrian things as current lack of attention, unwillingness to engage in conversation, lack of mastery of the language, etc. that may easily explain why a speaker's utterance cannot be taken, in a given context, at face value. When available, such explanations preempt the need for laborious intentional explanations of such misfits.

(24) Try to figure out all the background information that we bring to bear on our understanding of an utterance of a simple sentence such as "The cat is on the mat" (cf. Searle, 1979). For further discussion of the problem of explicitness, see Dascal (1983, pp. 85–91).

(25) Even in writing there is much 'para–linguistic' information that can provide clues for interpretation (e.g. the fluency, clarity, etc. of the handwriting). These clues are not necessarily absent in written communication with computers either. In this respect, Hofstadter (1986, pp. 522–523) points out the significant differences between: (a) the 'mail' mode, where texts of arbitrary length are put in full in a mailbox; (b) the 'write' mode, where full lines are instantly transmitted when the return key is pressed; and (c) the 'talk' mode, where every keystroke is instantly transmitted. This last mode reveals hesitations, corrections, second thoughts, etc. — which may be crucial in interpretation.

(26) For a pragmatic analysis of jokes, see Dascal (1985).

(27) For further discussion, see Dascal (1985). For some examples of the reliance of dreams on literal meaning, see Dascal (1987c, p. 275).

(28) According to Freud, when Tartini composed his *trillo del Diavolo* sonata in a dream, he accomplished an intellectual achievement "due to the same mental forces which produce every similar result during daytime" (Freud, 1900, p. 612).

(29) For discussion of the 'matching bias' and other examples of the influence of language in reasoning, see Dascal (1987a).

(30) Winograd and Flores's exposition is very systematic and clear, and there is no need to gloss it over here. It should be pointed out that they combine the Heideggerian viewpoint with the 'autopoietic' approach to living systems and to cognition, due to Maturana and his associates. This aspect of their work, which in my opinion is related to the 'deep structure'*causal* models of interpretation I mentioned in section 3.2, is not discussed here. Nor do I discuss their criticism of specific AI trends, such as the "fifth generation" project, connexionism, expert systems, etc (pp. 125–139). For my views on expert systems, see Dascal (1989).

(31) For instance: "A problem is created by the linguistic acts in which it is identified and categorized. Of course, some situation is previous to the formulation, but its existence as a particular problem ... is generated by the commitment in language of those who talk about it" (p. 147).

(32) Heidegger (1927, pp. 157–158). "The entity which is held in our fore–having — for instance, the hammer — is proximally ready–to–hand as equipment. If this entity becomes the 'object' of an assertion, then as soon as we begin this assertion, there is already a change–over in the fore–having. Something *ready–to–hand with which* we have to do or perform something turns into something *'about which'*the assertion that points it out is made... Both *by* and *for* this way of looking at it, the ready–to–hand becomes veiled as ready–to–hand" (Engl. transl., p. 200).

(33) Heidegger (1927, p. 158). "As regards its possibilities for Articulating reference–relations, it has been cut off from that significance" (Engl. transl., p. 200).

(34) In my opinion, traditional speech–act theory should be rather conceived as a part of semantics. This is no mere classificatory quibble, and it is relevant to the present discussion. The most important point to recall is that for accounting for the 'fluid structure' of converations. See Dascal (Forthcoming a) for details.

(35) I am grateful to Adam Tanenbaum for sharing with me his expertise in Heidegger, and to Itiel Dror for comments on a previous version of this paper.

References

Abelson, P. (1975) *Concepts for representing mundane reality in plans*, in: D.G. Bobrow & A. Collins (Eds.), *Representation and Understanding: Studies in Cognitive Science*, pp. 273–309 (New York, Academic Press).

Boden, M. (1977) *Artificial Intelligence and Natural Man* (New York, Basic Books).

Dascal, M. (1979) *Towards psycho–pragmatics*, in: T. Borbe (Ed.), *Semiotics Unfolding* [Proceedings of the Second Congress of the International Association for Semiotic Studies, Vienna, July 1979], pp. 1119–1126 (Berlin/New York/Amsterdam, Mouton, 1984).

Dascal, M. (1981) *Contextualism*, in: H. Parret, M. Sbisa & J. Verschueren (Eds.), *Possibilities and Limitations of Pragmatics*, pp. 153–177 (Amsterdam/Philadelphia, John Benjamins).

Dascal, M. (1983) *Pragmatics and the Philosophy of Mind*, vol. 1 (Amsterdam/Philadelphia, John Benjamins).

Dascal, M. (1985) *Language use in jokes and dreams: sociopragmatics vs. psychopragmatics*, *Language & Communication*, 5, pp. 95–106.

Dascal, M. (1987a) *Language and reasoning: sorting out sociopragmatic and psychopragmatic factors*, in: J.C. Boudreaux, B.W. Hamil, & R. Jernigan (Eds.), *The Role of Language in Problem Solving 2*, pp. 183–197 (Dordrecht, Elsevier).

Dascal, M. (1987b) *Interpretazione ermeneutica e interpretazione pragmatica*, *Rivista di Filosofia neo–Scolastica*, 79, pp. 564–579. [English version in *Philosophy and Rhetoric*, 22 (1989), pp. 239–259].

Dascal, M. (1987c) *Defending literal meaning*, *Cognitive Science*, 11, pp. 259–281.

Dascal, M. (1987d) *Leibniz: Language, Signs, and Thought* (Amsterdam/Philadelphia, John Benjamins).

Dascal, M. (1989) *Artificial intelligence and philosophy: the knowledge of representation*, *Systems Research*, 6, pp. 39–52.

Dascal, M. (Forthcoming a) *The pragmatic structure of conversation*, in: J.R. Searle et al., *On Searle on Conversation* (Amsterdam/Philadelphia, John Benjamins).

Dascal, M. (Forthcoming b) *Models of interpretation*, in M. Stamenov (Ed.), *Models of Meaning* (Amsterdam/Philadelphia, John Benjamins).

Dascal, M. (Forthcoming c) *Leibniz and Spinoza – Language and Cognition. Studia Spinozana.*

Dascal, M. & Katriel, t. (1977) *Between semantics and pragmatics: the two types of 'but' – Hebrew 'aval' and 'ela', Theoretical Linguistics*, 4, pp. 143–172.

Dascal, M. & Weizman, E. (1987) *Contextual exploitation of interpretation clues in text understanding: an integrated model*, in: J. Verschueren & M. Bertuccelli-Papi (Eds.), *The Pragmatic Perspective*, pp. 31–46 (Amsterdam/Philadelphia, John Benjamins).

Dascal, M. & Weizman, E. (1990) *The issue of speaker's meaning in literary texts. Hebrew Linguistics* 28/29/30 (Special issue on "Stylistics and the Literary Text"), pp. 9–19 [in Hebrew].

Descartes, René. *Oeuvres de Descartes*, Charles Adam et Paul Tannery (eds.) (Paris, Vrin; Nouvelle Edition, 1967–1975). [= AT].

Descartes, René. *The Philosophical Works of Descartes*, vol. 1 Elizabeth S. Haldane and G.R.T. Ross (translators and editors) (Cambridge, Cambridge University Press, 1967. [= HR].

Descartes, René. *Philosophical Letters*, A. Kenny, translator and editor (Oxford, Clarendon Press, 1970). [= Kenny].

Fodor, J.A. (1975) *The Language of Thought* (New York, Thomas Y. Crowell).

Fodor, J.A. (1980) *Methodological solipsism considered as a research strategy in cognitive psychology*, in *Representations – Philosophical Essays on the Foundations of Cognitive Psychology*, pp. 225–253 (Cambridge, Mass., MIT Press).

Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge, Mass., MIT Press).

Freud, S. (1900) *The Interpretation of Dreams*, transl. J. Strachey (London, Hogarth Press).

Freud, S. (1905) *Jokes and their Relation to the Unconscious*, transl. J. Strachey (London, Hogarth Press).

Hacking, I. (1975) *Why Does Language Matter to Philosophy*? (Cambridge, Cambridge University Press).

Heidegger, M. (1927) *Sein und Zeit* (Tübingen, Max Niemeyer) [Engl. transl. by J. Macquarrie & E. Robinson, Oxford, Blackwell, 1962].

Heidegger, M. (1947) *Brief über den 'Humanismus'*, in: *Platons Lehre von der Warheit*, pp. 53–119 (Bern, Francke Verlag).

Heidegger, M. (1962) *Die Frage nach der Technik*, in: *Die Technik und die Kehre*, pp. 5–36 (Pfullingen, Neske).

Hobbes, Thomas. *Leviathan: or the Matter, Form and Power of a Commonwealth Ecclesiastical and Civil.* London, 1651 (quoted from *The English Philosophers from Bacon to Mill*, Edwin A. Burtt (ed.), New York: The Modern Library, 1939).

Hofstadter, D.R. (1986) *Metamagical Themas: Questing for the Essence of Mind and Pattern* (New York, Bantam Books).

Katriel, T. & Dascal, M. (1989) *Speaker's commitment and involvement*, in: Y. Tobin (Ed.), *From Sign to Text: A Semiotic View of Communication*, pp. 275–295 (Amsterdam/Philadelphia, John Benjamins).

Lakoff, G. (1985) *Metaphor, folk theories, and the possibilities of dialogue*, in: M. Dascal (Ed.), *Dialogue – an Interdisciplinary Approach*, pp. 57–72 (Amsterdam/Philadelphia: John Benjamins).

Lakoff, G. & Johnson, M. (1979) *Metaphors we Live By* (Chicago, The University of Chicago Press).

Leibniz, G.W. *Nouveaux Essais sur l'Entendement Humain*. In *Die philosophischen Schriften von G.W. Leibniz*, vol. 5, C.I. Gerhardt (ed.), Berlin, 1875–1890 (Hildesheim: G. Olms, reprint 1965). [= NE].

Locke, J. *An Essay Concerning Human Understanding*, 2 vol. (ed. by J. Yolton, New York/ London, Everyman's Library).

Morris, C. (1938) *Foundations of the Theory of Signs* (Chicago, The University of Chicago Press).
Newell, A. & Simon, H.A. (1972) *Human Problem Solving* (Englewood Cliffs, NJ, Prentice–Hall).
Rich, E. (1988) *Artificial Intelligence* (Singapore, McGraw–Hill International).
Ryle, G. (1949) *The Concept of Mind* (London, Hutchinson).
Searle, J.R. (1979) *Literal meaning*, in: *Expression and Meaning*, pp. 117–136 (Cambridge, Cambridge University Press).
Sloman, A. (1978) *The Computer Revolution in Philosophy: Philosophy, Science and Models of Mind* (Brighton, Harvester Press).
Turing, A.M. (1950) *Computing machinery and intelligence*, in: E.A. Feigenbaum & J. Feldman, *Computers and Thought*, pp. 11–35 (New York, McGraw–Hill, 1963).
Wason, P.C. (1981) *Understanding and the limits of formal thinking*, in: H. Parret & J. Bouveresse (Eds.), *Meaning and Understanding*, pp. 411–421 (Berlin/New York: Walther de Gruyter).
Wason, P.C. & Evans, J.S. (1975) *Dual process in reasoning?*, *Cognition*, 3, pp. 141–154.
Wilensky, R. (1981) PAM, in: R.C. Schank & C.K. Riesbeck (Eds.), *Inside Computer Understanding* (Hillsdale, N.J., Lawrence Erlbaum).
Winograd, T. & Flores, F. (1987) *Understanding Computers and Cognition – A New foundation for Design* (Reading, Mass., Addinson–Wesley).

Luc Steels

VUB AI Lab.
Brussels (Belgium)

Issues and Open Problems in Subsymbolic Representation

Abstract

The paper puts forward a number of hypotheses concerning the nature of subsymbolic representation and processing and points out areas of future research.

Resumo

Este texto avança com várias hipóteses acerca da natureza da representação subsimbólica e seu processamento e aponta áreas para investigação futura.

1.Introduction

There is a growing feeling among many AI researchers and cognitive scientists that symbolic processing, which is now the dominating paradigm of AI, must be complemented by a *subsymbolic* layer. In this paper we formulate a number of hypotheses concerning what subsymbolic computation may be like and indicate open problems that must be tackled when these hypotheses are adopted.

More specifically, the subsymbolic layer is hypothesised to use *complex dynamics* as computational foundation and thus *massive parallelism*. *Analogical maps*, which represent parts of the problem domain implicitly, are conjectured to play a fundamental role. Mechanisms of self-organisation, analogical world modeling, and analogical reasoning are proposed as major operations taking place over these maps and *schemas* are a possible higher organisation form for subsymbolic computation. They are also the way to link the symbolic and subsymbolic levels. The Piagetian mechanisms of assimilation and accomodation are viewed as primary mechanisms for learning schemas of subsymbolic knowledge.

The paper is in two parts. The first one puts forward a set of hypotheses concerning the subsymbolic layer. The second part discusses open problems.

2.Hypotheses

2.1.Hypothesis 1: The need for subsymbolic computation

The symbolic paradigm

Since its beginning 30 years ago, Artificial Intelligence has made impressive progress based on adopting a *symbolic paradigm* for looking at cognitive phenomena. Symbols are

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposyum, 1989, 79/94
1992, APDC & SPF, Lisboa

viewed as primary units. They can be assembled to form structures which are used to represent parts of the world, intermediate results in problem solving, perceptual input, associative memories, inference rules, etc. Interpreters create structures and operate over them under the control of programs that are themselves represented explicitly.

Some important milestones in the elaboration of the symbolic paradigm were the definition of a symbolic level of computation (McCarthy, 1960), the discovery of the state space search model of problem solving and subsequently of the need for strategic and heuristic knowledge (Newell and Simon, 1972), the translation of linguistic formalisms into computational models (Woods, 1970), the mechanisation of logical inference (Kowalski, 1979), and the realisation that learning is a search process (Winston, 1970). Applications such as expert systems (Feigenbaum, 1977) or natural language interfaces are directly based on these insights.

The symbolic paradigm rests on concepts and theories elaborated mostly in the humanities (philosophy, logic, linguistics and psychology). It is remarkable that it was possible to transform these concepts into structures that are implementable in computational media. Neither were they developed for this purpose, nor was it easy to bridge the gap between mentalistic notions, such as the notion of a goal, and its computational counterparts.

Problems with the symbolic paradigm

Despite the enormous achievements of the symbolic paradigm, there are a large number of issues in Artificial Intelligence that remain unresolved. Consider the following non-exhaustive list:

a.*Flexibility*: No AI system matches the flexibility of the human brain, which tolerates deviations in input (e.g. in pattern recognition), adapts itself easily to slightly different circumstances, shows graceful performance degradation in problem solving, etc. The problems of current AI systems are due to the stringent nature of symbolic representations. Attempts have been made to introduce flexibility by the use of uncertainty calculi (Shortliffe and Buchanan, 1975), fuzzy set theory (Zadeh, 1978), or approximate reasoning (Prade, 1985). Although this research is definitely valuable, it has not lead to widely accepted solutions.

b.*Robustness:* AI systems are not robust, in the sense of being fault-tolerant. Every unit, whether it is a concept, rule, or some other construct, is there to serve a particular function and if the unit malfunctions the whole system fails. This becomes a critical issue in highly parallel systems where the behavior of each component can no longer be guaranteed to function in all circumstances. In contrast, the brain is able to keep functioning even if certain units disappear or if parts are damaged.

c.*Interaction with the world*: Many AI systems now presuppose already a large part of the problem solved, namely the concrete interaction with the world. Problems are posed in terms of symbolic descriptions and solved at the symbolic level. The human brain has to interpret first the world in order to input the description and later interpret the results to translate them back to the world. What is needed however are AI systems that have direct world interaction, i.e. that are autonomous.

d.*Problems in decomposition:* The principle of functional decomposition which underlies all hierarchical systems, including current symbolic computer programs, assumes that a problem can be split up in different subproblems which can be solved independently. A solution for the total system is then obtained by connecting together in an appropriate

way the subsystems that solve each subproblem. This approach has a great number of advantages, for example modularity. But many problems in AI are not decomposable this way. For example, many interpretation tasks (natural language understanding, vision, measurement interpretation, etc.) involve ambiguous data. The data itself does not contain enough information to uniquely interpret it so that knowledge of the situation of general world knowledge must be brought to bear. This leads to a chicken and egg situation: we need to partially interpret the data but to interpret the data we need a partial interpretation. No straight division into subproblems is possible. Similar difficulties arise in planning (e.g. the conjunctive subgoal problem (Sussman, 1974) or learning (to learn something one has to know already a great deal).

e.*Explicit construction*: Most AI programs today are explicitly constructed by hand. This is feasible for small-scale problems, but the complexity barriers are soon reached when realistic problems are tackled. Even for small domains, the difficulties of knowledge extraction, as experienced in the construction of expert systems, are very severe (Steels, 1987). When parallelism is involved, the problem becomes even more serious because issues of synchronisation, protection of shared resources, scheduling, etc. have to be taken into account. It is true that recently there has been substantial progress in the area of machine learning (see Michalski, et.al. (1983) for a representative sample). But the distance to systems that have acquired a complete domain of expertise is still enormous. Moreover learning programs are themselves becoming progressively more complex raising the question how they are formed in the first place.

The current approach to these fundamental problems is to add more and more knowledge to AI systems: knowledge on what to do when the input is not exactly as required, knowledge on flexible control, knowledge on synchronisation and scheduling of parallel reasoning sequences, etc. This exagerates the problems of knowledge extraction and engineering complexity. Sometimes complex bureaucratic systems are proposed even for relatively simple tasks. The reflective systems recurse to a tower of meta-levels in an attempt to deal with issues of flexibility, control of reasoning, etc. (Maes and Nardi, 1988). Although knowledge will remain the key ingredient of any AI system, maybe it is time to look instead for a whole new set of tools and concepts for building knowledge based systems.

Subsymbolic computation.

One possible fundamental hypothesis is that these problems can only be solved by adding a subsymbolic layer to the symbolic one. The subsymbolic layer does not work on the basis of symbols, but on the basis of representations which are presymbolic and closer to the data available from the real world. This helps for solving problems concerning flexibility and real-world interaction. At the same time the subsymbolic level should exhibit properties of self-organisation and distributed processing, thus providing a handle on programming complexity, robustness and difficulties in decomposition.

In an attempt to clarify the notion of subsymbolic computation some further questions need to be addressed:

- What kind of representations are used at the subsymbolic layer.
- What kind of computation takes place at the subsymbolic layer.
- How does the subsymbolic layer interact with the symbolic one.

The following hypotheses give more concrete answers to each of these questions.

2.2.Hypothesis 2: Subsymbolic computation involves analogical MAPS

Symbolic models of intelligence use primarily propositional representations for a variety of reasons. Some of them are technological: they are at the moment easy to implement, even if the symbolic structures are very complex. Other reasons are theoretical: most theories of human knowledge, particularly those inspired by logic, center around propositional representations. In contrast, we hypothesise that the key characteristic of subsymbolic processing is the exploitation of analogical representations or maps.

An analogical map keeps certain key properties of the domain implicit in the representation. (Sloman, 1971) For example, the shape of an object is represented in terms of strength in a particular cell (see fig. 1).

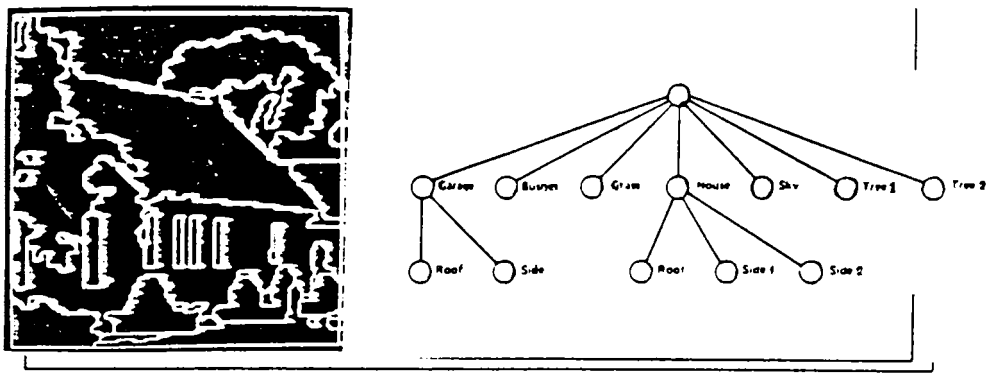


Fig. 1. Analogical vs. symbolic representation

Analogical does not mean that it is like a photographic picture. It could have more or less precision and focus on more or less detail. An analogical map could also be an interpretation of the world in terms of features, but these features are represented as regions of activation, rather than discrete symbolic entities. An example is a 'smell'-system in which a particular smell is internally represented as activation of a particular region on a 'smell'-map, or a 'phonotopic' map of the phonemes, in which each phoneme is a region in this space. (Kohonen, 1988)

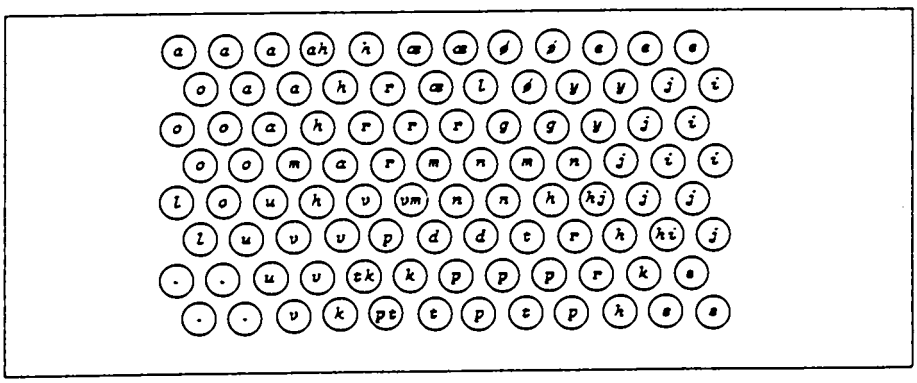


Fig. 2. Phonotopic map of phonemes.

Analogical maps contrast with propositional representations which cut the domain up in distinct objects and use predicates to describe the properties of the domain which are important for problem solving. For example, spatial relations are represented explicity in terms of predicates such as *left-of, on-top-of*, etc.

Not only the representation but also the processing over the representation to solve specific problems constitutes a crucial difference between analogical and propositional approaches. In the propositional approach, symbols are organised in expressions and an inference is defined over these expressions to arrive at new information. Operations over analogical maps are in contrast dynamical operations over the maps themselves. They could be viewed as a form of mental imaging (Kosslyn, 1985).

For example, the behavior of fluids can be modeled by defining operations over analogical maps of fluids (see Gardin and Meltzer, 1989). The fluid is represented as a large collection of molecules which have a behavior expressed as local dynamical laws. This behavior is to a specific level of detail an 'analog' of real physical behavior and allows simulation and prediction. This approach clearly contrasts with the naive physics approach (see e.g. Hayes, 1985) in which fluids are modeled with symbolic descriptions, logical axioms and logical inferences to arrive at new conclusions.

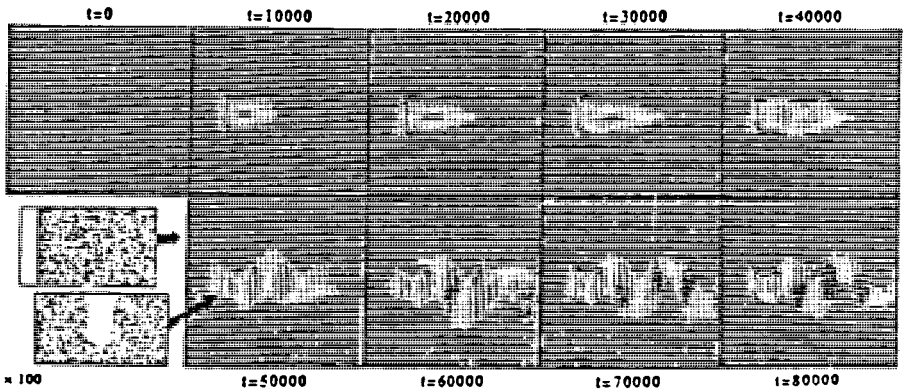


Fig. 3. Behavior of fluid modeled as analogical operations

Similarly, continuous speech can be seen as a dynamic trajectory taking place over an analogical map (fig 4, from Kohonen, 1988). Various phenomena such as coarticulation and assimilation, which are now modeled based on phonological rules operating over distinctive features (i.e. symbolic representations), can be rephrased in terms of dynamical processes over analogical maps.

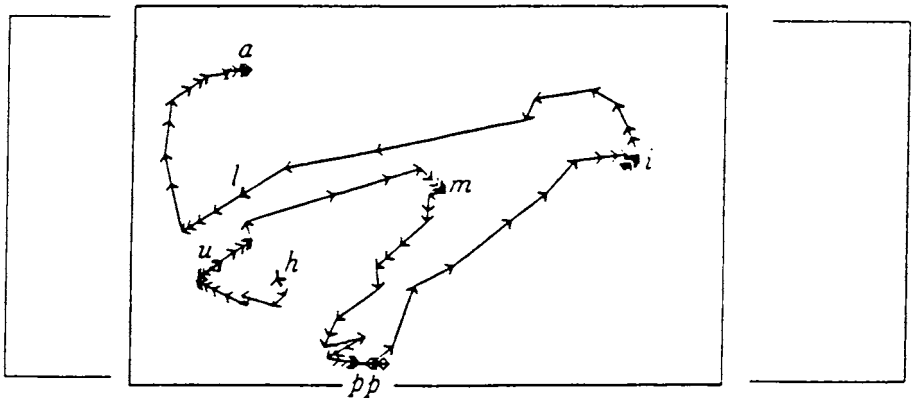


Fig 4. Trajectory over phonotopic speech map for a specific word

Note that analogical maps also contrast with diagrams which have been used in a few AI programs to complement symbolic representations, such as in Gelernter's geometry theorem prover (Gelernter, 1963), or the physics problem solvers of Funt (1980) and Larkin and Simon (1987). Diagrammatic representations decode key properties of the domain in a graph-like representation and translate certain reasoning steps into propagation in graphs. Constructing diagrammatic representations is therefore a nontrivial knowledge-based process. In contrast, analogical maps remain close to the world and do not involve graph-like structures.

Here are some arguments why analogical representations should be the primary form of representation at the subsymbolic level:

1) Uncertainty handling: Symbolic descriptions of reality are very rigid because they force a classification onto the domain. Im many knowledge-based systems (e.g. in expert systems), uncertainty is added later by introducing a certainty factor for each fact. This requires an uncertainty calculus alongside the symbolic representation. Using analogical representations, we avoid in many cases interpretation because the precision of the real world is maintained in the representation. For example, there can be laws that describe the behavior of an obstacle without the need to have absolute certainty of the absolute boundaries of the obstacle or without the need to interpret the obstacle as belonging to a specific symbolic class (wall, person, etc.).

2) Real world interaction: Many of the difficulties in perception come from the need to derive symbolic descriptions. Analog world modeling can operate directly on the data delivered by sensing and thus avoids part of the interpretation problem. For example, the side of the road need not be turned into a concrete detailed description for a mobile robot but can remain at an analogical level.

3) Robustness: Because the analogical world modeling that is proposed is based on molecular fine grained representations, it is more robust. For example, a liquid could be modeled in terms of a large collection of molecules. If some of these do not exactly behave as they should be (e.g. because they were interpreted wrongly, because there is an error in the computation for that molecule, etc.), this does not have a large scale catastrophic effect. In contrast, an error in a symbolic description usually means failure of the complete solution.

4) Programmability: We see in many applications (such as the fluid modeling or the treatment of phonological co-articulation phenomena) that effects follow as an emerging

property of the representation, instead of being explicitly programmed. Also the definition of the maps can be shown to be the result of self-organising, emergent properties. This property reduces substantially the problems in programming complex applications.

5) Time: one of the most important constraints imposed by nature on the cognitive system is time. To avoid untimely situations (e.g.: car accidents) the system has to compute predictions about the changing environment quickly. Therefore a representation which allows for simple and parallel operations for the tasks of interest is advantageous. In analogical representations information can often be organized such that related information is locally available through simple operations.

2.3.Hypothesis 3: Subsymbolic computation is a complex dynamical system

If we want to make progress we will need a model of subsymbolic computation that is as general as the symbolic model embodied in languages like LISP and PROLOG. Symbolic representations are iherently serial. Symbols are allocated and translated to pointers to particular memory cells. If more than one process needs to use the same symbol there needs to be shared memory and thus reduced parallelism. In contrast, analogical maps invite massive parallelism. They can be implemented as a cellular grid or a network of cells in which each cell has a certain strength for a particular feature. Several maps can be overlayed so that there is a potential interaction between the strength of different features. Thus there may be a map for color overlayed with a map for brightness and a map for texture. Operations can be defined as local operations over the nodes in the map. The inherent parallel nature of subsymbolic computation leads to the hypothesis that subsymbolic should be viewed as a complex dynamical system.

What are complex dynamical systems

A *dynamical system* is a system in which a number of properties evolve in time in parallel. The classical and simplest example is a pendulum. Its dynamics can be described in terms of two properties: the angle of elevation θ and the rate of rotation $\theta' = d\theta/dt$. All of these properties follow a specific trajectory made up by the values they take at each point in time. Another example are the gas molecules in a container. Each of these molecules constantly changes position in time.

Dynamical systems abound in nature. Examples include an object falling to earth, evolving animal populations, the growth of an organism, planets moving around the sun, evolution in the stock market, or traffic flow. Even standard computer programs can be viewed as dynamical systems. For example, the program counter, which indicates control flow, continuously changes in time as the computation proceeds.

A distinction can be made between simple and complex dynamical systems, based on the number of properties undergoing change. In the case of the pendulum, the falling object, or the computer program, only very few properties keep changing. These properties can therefore be described and analysed explicitly. In the case of the gas container, there is an extremely large number of properties (e.g. a drop of water contains already billions of molecules). It is therefore no longer feasible to follow the course of each individual property, only statistical or macroscopic statements can be made.

Even more important is the following distinctive feature: Because of the many interactions between the evolving properties in a complex dynamical system, a new set of behaviors emerges. These behaviors are not the consequence of a single element in the

system, but a side effect of the collective behavior of all units. Complex dynamical systems therefore differ from simple dynamical systems because they show *emergent properties*.

These emergent properties are often very remarkable and usually occur suddenly at phase transitions. For example, there are certain reactions (such as the Belousov–Zhabotinsky reaction) that after a while lead to an oscillating behavior (i.e. a chemical clock) or to the formation of patterns. It is as if billions of molecules cooperate on cue, without a supervising global control mechanism, or global communication. (Babloyantz (1986)). Other examples are the formation of a path by an ant society (Deneubourg, et.al. 1983), genetic evolution through natural selection (Dawkins, 1986) or the immune system (Edelman, 1978).

Dynamical computation systems

Although large-scale dynamical systems form the core subject of several scientific disciplines such as thermodynamics, statistical mechanics, population dynamics, sociobiology, etc., it is only recently that they have become relevant for computer science. The main reason is that it is now feasible to construct computational systems that execute a very large number of processes simultaneously. It is now possible to construct computer systems that consist of a large number of processors (at present on the order of 60,000 (Hillis, 1986)) and this number will increase in the future. There are also opportunities for constructing systems based on analogical VLSI circuitry (Mead, 1987), optics (Bell, 1986), or even molecular components (Conrad, 1986), that may raise the level of parallelism even more and bring computers closer to natural complex dynamical systems.

The study of complex dynamical systems reveals that they have a number of powerful properties which could be directly exploited for solving certain computational problems. Recent work on cellular automata (Wolfram, 1983), simulated annealing (Kirckpatrick, et.al., 1983), Hopfield nets (Hopfield, 1982), Boltzmann machines (Hinton and Sejnowski, 1985), neural networks (Rumelhart and McClelland, 1986), genetic algorithms (Holland, 1975), and spin glasses (Kinzel, 1983), has shown that problems which are difficult for the symbolic paradigm, such as constraint satisfaction, associative memories, pattern recognition and pattern classification, can be approached by mapping these problems onto a complex dynamical system which then solves the problem by executing its own dynamics (Steels, 1987c). Of particular interest is the work here of Kohonen (1988) which has shown how maps could be formed through self-organisation.

Most of the work within the dynamics paradigm has not dealt seriously with problem solving or reasoning, which is of course a key ingredient of sensorimotor intelligence and common sense. Although it has been shown how dynamical computation systems can extract maps from the world through sensors, it remains to be shown how operations over analogical maps can lead to make predictions about future states of the world.

Arguments for the use of complex dynamics

Dynamical computation systems show a number of novel properties which explains why they are being studied intensely at this point and why they are relevant to the AI enterprise.

One of them is fault tolerance: a single unit is no longer important and the system can restore itself even if there is damage to some of its parts.

Another one is flexibility. Symbolic computer systems only work when the inputs are

precisely as expected and the situation to be handled falls completely within the class of programmed situations. Dynamical computation systems are capable of handling inexact input and situations that are not completely foreseen because they operate with dynamically varying quantities (e.g. weights on the connection between units) that have certain ranges of operation.

Yet another important property is the capability for self-organisation which is an alternative for explicit programming. Explicit programming is very powerful but quickly reaches complexity barriers for the programmers. Natural phenomena of self-organisation show that elaborate complex structures can be formed. The prime example is genetic evolution which produces complex life forms.

Another reason for the recent interest is that dynamical computation systems are an alternative tool to differential equations for describing and studying natural dynamical systems (Toffoli (1984), Toffoli and Margolis (1987), Wolfram (1983). This is important given the well known difficulties involved in working with non-linear differential equations. The main technique is to simulate at the microscopic level the behavior of the system in an attempt to obtain the same macroscopic properties as the natural system being studied. This assumes however that the computational media implementing these dynamical systems are sufficiently efficient to execute the required number of processes fast enough.

Finally, given that all these features are observed for the brain and that the brain is hardware-wise clearly a complex dynamical system, the study of dynamical computation systems opens up new frontiers for building realistic models of human cognition. Current work in neurology, such as by Edelman (1987), clearly goes in this direction.

All these properties are important for the objective of constructing AI: fault-tolerance, flexibility, self-organisation, simulation of physical phenomena through microscopic laws, and relation to the brain, and therefore justify the adoption of this hypothesis.

Formalism

It has already been pointed out that there are a variety of possible frameworks in which dynamical computation systems can be embedded: neural networks, cellular automata, reaction diffusion systems, spin glasses, etc. Each of these models is equivalent to a certain degree but may have different computational properties which make it harder to study certain issues. This is similar to the different paradigms in programming languages (imperative, functional, logic, object-oriented programming) which are all equivalent to Turing machines but nevertheless allow different styles of programming and are more efficient for certain purposes.

In our own work we have found that the metaphors of reaction-diffusion are well suited to express the many kinds of dynamical processes that are needed over analogical representations. Reaction-diffusion systems are a formalism underlying most of the work on self-organisation in biology and chemistry (Babloyantz, 1986). Diffusion means that material is distributed progressively in space using only local communication (e.g. through cell boundaries). This can be used as a metaphor for the diffusion of information in a network. Reaction means that material may interact to create new material, inhibit the formation of other material, destroy material, etc. This translates to the creation of new information, inhibition or excitation, destruction, etc.

2.4.Hypothesis 4: Subsymbolic computation is schema based

It is obvious that a complex system with many functionalities can only be constructed when there is a higher organisation. One form of organisation for which there is some consensus in AI and psychology is the schema. Schemas have their intellectual ancestors in the schema-concept of Piaget (1962) and Minsky's notion of a frame (1975), particularly as further worked out in the Society of Mind theory (Minsky, 1986).

Today frame-based representations are widely used in all areas of AI as alternatives to logical representations. A wide variety of technological realisations exist, particularly in the form of knowledge representation languages or expert system shells. They are now used on a wide scale in the construction of expert systems and other AI applications.

However schemas so far have been entirely symbolic (although Piaget and Inhelder (1966) discuss schemas for mental imaging). We hypothesise that it is possible to develop subsymbolic schemata and that large parts of sensorimotor intelligence and common sense (particularly mental imagery) can be modeled using such schemas.

The schemas should have a subsymbolic component in conjunction to the symbolic one. They can therefore act as the interface between symbolic and subsymbolic computation. A schema has a number of slots that have to be filled by specific features in specific maps before the schema can be applied. Filling the slots ties the schema to the reality being considered. A schema has a particular functionality: For example, it models a small part of the world such as pushing and pulling, strings, fluids, obstacles. Each schema works with specific analogical representations and defines its own dynamics for working with these representations.

Schemas must be composable with other schemas in a modular way. For example, one schema may come in action after another one to achieve a larger goal, or one schema may come in action when another one failed. Schemas are to be organised in inheritance hierarchies allowing inheritance between different schemas.

We hypothesise that the control over which schema will become active, and the learning of schemata should also be modeled using complex dynamical processes. An understanding of these processes may give new interpretations to the Piagetian notions of equilibration, assimilation and accomodation for which no technical definition yet exist.

New approaches to analogical reasoning

The organisation of sensorimotor intelligence in terms of schemas gives also a new way to handle analogical reasoning. This is the process of using knowledge of the world to solve abstract problems by mapping common sense notions onto the problem. For example, the control flow in a program can be viewed as a flow of water through a set of channels. This model is useful for educational purposes and for envisioning what happens when a particular program is being executed.

There has been a lot of recent work on analogical reasoning and envisioning within the symbolic paradigm (see the overview in Prieditis, 1988). This work involves mostly the construction of complex symbolic representations and the formulation of symbolic mappings between representations. What is proposed here is very different. The mapping does not take place between symbolic representations. Instead the problem is translated into an analogical map onto which common sense schemata can be activated. The activation of these schemata gives a solution which needs be translated back into the original problem domain. Steels (1988a) gives an example of this approach by showing how common sense

schemata for obstacle avoidance and string manipulation can be used for solving abstract problems such as the 8-puzzle.

2.5.Conclusions

In this section four important hypotheses were put forward. Each of these hypotheses questions existing assumptions of AI. The first hypothesis concerned the importance of nonsymbolic or subsymbolic representations as complementary to symbolic representations which are now the dominating paradigm in AI. This hypothesis was refined by three more hypotheses. On the one hand, analogical representations which keep part of their information implicit in the representation are seen as primary at the subsymbolic level. Second, it is assumed that the operations over these representations are massively parallel and can be modeled as a complex dynamical process. Third, it is assumed that knowledge about analogical representations and their corresponding dynamics is organised in schemata.

3.Open Problems

The hypotheses presented in the previous section raise at this point more questions than they give answers, but this is of course why we need to do basic research. For the areas of sensorimotor intelligence, common sense, but also problem solving, they constitute a paradigm shift compared with the standard classical symbolic approach. The methodology we recommend is to pose specific open problems that refine, test, and validate the hypotheses so that concrete experiments could be conducted.

In this section a list of these open problems will be briefly discussed. This list is in a sense the initial research agenda of the new approach advocated in this paper. It is expected that this list will be extended or refined when progress is made. For some of the issues discussed here there are already partial solutions available. Although these solutions will need to be examined critically and novel approaches must be explored.

3.1.Issues of representation

Issues concerning analogical maps

Analogical maps are almost a virgin territory from the viewpoint of knowledge representation research. Here are some of the major issues:

What maps are appropriate for specific important tasks in sensorimotor intelligence and common sense. This question can only be dealt with by considering specific problems and trying to obtain satisfactory solutions. For example, we could take problems concerning the modeling of fluids and develop analogical representations that are capable to solve the problem. Or we could take the problem of inverse kinematics for motor control of robots and use analogical representations of the space through which the arm has to evolve and of the joints that constitute the arm.

How can dynamic maps be represented, derived and used. Dynamic maps are needed for representing time-dependent events. They are a well-known open problem in dynamical computation systems, both in terms of their derivation from input data, and in terms of their usage later. Dynamic maps are important in many areas of sensorimotor intelligence and common sense.

There are a large number of issues concerned with the flexible definition and use of the maps. For example:

- How can maps with different levels of detail be obtained and combined.
- How can maps from different sources (e.g. from different sensors) be combined.
- How can flexible maps be mapped onto other situations for which they do not fit completely.
- What kind of interesting computational tasks can be better solved by analogical representations (e.g. interpolation, smoothing, sharpening of data).

It would be highly desirable to develop a general theory of analogical representation concerning analogical maps. Such a theory would address questions like:

- Are there criteria for deciding which properties need to be made implicit vs. explicit
- Can there be a mathematical definition of the complexity properties of analogical maps.

A large number of open issues concern the relation between analogical maps and real-world data:

- How can maps be obtained by direct input from sensors.
- What kind of low level (parallel) processing is useful for obtaining better analogical maps.
- How can new maps be formed through processes of self-organisation.
- How can maps be translated to the operation of effectors.

Issues concerning schemata

It will be necessary to develop concrete technical definitions of schemas that have an analogical component and to realise these schemata in computational terms.

One of the many open questions in this respect concerns the problem of modularity: is it possible to develop schemas that work relatively independently and how can these schemas be combined in a modular way. Other issues concern the relation between analogical and symbolic representations, such as:

- How can results on maps be interpreted to have an impact on the symbolic layers. This may involve the utilisation of feature detectors to translate a property on the map to a symbolic description, or the direct use of diffuse symbols which have a direct correspondent on an analogical map.
- How can symbolic descriptions have an impact on analogical maps or how can they be used to partially form analogical maps. One set of issues relates to the problem of translating natural language descriptions of scenes into visual images of these scenes and perform deductions important for understanding the rest of the sentence by operations over the analogical maps.

3.2.Issues of learning

The issues of learning fall into two major sections: (i) how are analogical maps and operations over them formed, and (ii) how are schemata and the interaction between them learned.

One major form of learning that will have to be understood better is *self-organisation*. There are already many important results in this area such as the Kohonen framework, but there are also many remaining open problems. For exemple, concerning the formation of

features in time-dependent phenomena, the formation of feature maps with different degrees of precision, etc.

A second form of learning at the subsymbolic level is *induction*. Experiments of this sort have already been carried out by Mohnhaupt and Neumann (1988) who showed that induction processes could lead to more generalized analogical representations of vehicle movements on a street. This work must be critically examined and extended to other domains.

We hypothesised (hypothesis 4) that subsymbolic knowledge is organised into schemata. The learning of these schemata and of how these schemata are combined and hierarchically structured is a completely open problem. There are theories stemming out of psychology (such as those of Piaget), but the computational counterparts of these theories is unknown (see the work of Dresher, 1985, for a beginning). Open problems here are:

- Is it possible to define technically the Piagetian processes of equilibration through assimilation and accomodation in terms of complex dynamical systems operating over schemata?
  What are the basic mechanisms realising these processes, e.g. how can a sensorimotor schema be 'assimilated'?
- In which sense can important evolutionary stages emerge by the large-scale interaction of different schemata?

3.3.Issues of dynamics

One of our key hypotheses is the use of complex dynamics as an alternative for symbolic processing. There are a number of major open problems concerning the kind of complex dynamics that is appropriate, apart from the problem of how this complex dynamics should be used.

A lot of work at the moment is going on using the formalism of neural networks, but other formalisms, for example *reaction-diffusion* systems, may be more appropriate in certain circumstances.

Most current work in the dynamics paradigm assumes an equilibrium dynamics in which the system starts with a given input situation which is left to evolve until a state of rest is reached. A typical example is the Hopfield net in which certain units are clamped and left to evolve until the end result of computation can be read off. (Hopfield, 1982). An important issue in this respect is in how far *non-equilibrium* dynamics could be used, i.e. dynamics which receives a continuous influx from the environment and build up structure in response to this perturbation. Many biological systems are of this sort and selectionism appears to be a primary paradigm for studying these systems (Steels, 1987b). The Kohonen self-organising feature maps, as well as genetic algorithms (Holland, 1975), exhibit selectionist properties.

There are a large number of open theoretical issues which are nevertheless of considerable practical importance to arrive at concrete experiments and technological realisations. Some of these issues are:

- Can a mathematical relation be established between the size of the system (e.g. number nodes in the network or grid) and capacity (e.g. for storage of information)?
- In how far can other attractors than point attractors be used (e.g. limit cycles, spatio-temporal events, or possibly chaotic dynamics)?
- Can there be mathematical predictions about the time needed to reach attractor states?

- How to keep some aspects of the network if the number of nodes are changed?
- Is it possible to develop a theory of the stability of self-organised feature maps which shows, among other things, under which conditions these maps desintegrate?

3.4.Empirical issues

The hypotheses raise also a number of questions in relation to psychological and neurological models.

There are first of all strong relationships with the Piagetian theories of sensorimotor intelligence. Exploring the hypotheses may provide substantially new interpretations of these theories and show their computational validity. Concretely the following questions are important in this respect:

- What is the current state of the art in developmental psychology concerning the acquisition of sensorimotor intelligence and common sense?
- Can these data be related to the hypotheses?
- Can the Piagetian experiments be interpreted in computational terms?

The question of mental imagery has also been a controversial subject in psychology with strong proponents and strong opponents (see the discussion in Block, 1981). Also within the framework of Piagetian developmental psychology experiments were already conducted on mental imagery (Piaget and Inhelder, 1966). It is expected that further research may throw light on the issues by providing technical models that can be shown to perform mental imagery through operations over analogical maps. Major issues in this respect are:

- Are these computational experiments in any sense capable to settle psychological disputes concerning the use of mental imagery in humans?
  What kind of predictions can be formulated about mental imagery that could be settled by psychological experimentation?

Finally, there is a growing neurological literature on maps and map formation. In what sense can the data contained therein be brought in relation to the theoretical and computational models put forward here. In how far can selectionist principles, recently put forward by Edelman (1987) as a major hypothesis on brain function, be validated.

3.5.Computational issues

Subsymbolic computation as envisaged here relies heavily on massively parallel computation. This field is at the moment still in great flux with controverses ranging about adequate architectures and programming models. Some of the issues here are:

- Complexity of the nodes: Some architectures (such as the Connection Machine, Hillis, 1987) assume simple nodes but assume a large number of them. Other architectures (such as those based on the transputer) assume complex nodes with heavy communication between nodes. There are important trade offs depending on the kind of activity that needs to be done at each node.
  Data-level vs. process level parallelism. Data-level parallelism assumes that the same operation is carried in parallel over a large number of data. Process parallelism assumes that there are many independent processes. What is the most appropriate model for subsymbolic computation? It seems that data level parallelism is definitely more appropriate for operations over analogical maps, but what about the coordination of schemata.

. Special vs. general purpose. Complex dynamical computation systems consume vast amounts of memory and computing power. One important approach to make their usage feasible is by translating specific solutions into hardware of into representations that are closer to hardware. The use of silicon compilers or dedicated hardware that supports data level parallelism directly could help in this respect, but it is an open question which approach is most cost effective for the problems considered.

By being a user with specific demands, experimentation along the lines advocated in this paper can considerably help to settle some of these issues in parallel computing and push the state of the art in a particular direction.

4.References

Babloyantz, A. (1986) *Molecules, Dynamics and Life. An Introduction to Self-Organization of Matter*. Wiley-Interscience. New York.

Bell, T.E. (1986) *Optical computing: a field in flux*. p. 34–57. IEEE Spectrum. August 1986.

Conrad, M. (1985) *On design principles for a molecular computer*. pp. 464–480. in Comm ACM. Vol. 28, nr 5.

Dawkins, R. (1986) *The blind watchmaker*. Longman Scientific and Technical Pub. Harlow.

Deneubourg, J., J. Pasteels and J. Verhaeghe (1983) *Probabilistic Behaviour in Ants: A Strategy of Errors?* In J. Theor. Biol. (1983) 105, 259–271.

Dresher, G. (1986) *Genetic AI. Translating Piaget into LISP*. MIT AI Memo. 890.

Edelman, G. (1987) *Neural Darwinism*. Basic Books. New York.

Feigenbaum, E. (1977) *The Art of Artificial Intelligence: Themes and Case Studies in Knowledge Engineering*. 5th IJCAI, Cambridge. MA.

Gardin, F. and B. Meltzer (1989) *Analogical representations of Naive Physics*. AI Journal. Vol 38, nr 2, march 1989.

Gerlernter, H., J. Hansen and D. Loveland. (1963) *Empirical Explorations of the Geometry Theorem Machine*. in E. Feigenbaum and J. Feldman. (eds) *Computers and Thought*. McGraw-Hill, New York. Hayes, P. (1985) *Naive Physics I: Ontology for Liquids. pp. 71-107*. in J. Hobbs and R. Moore (eds.) *Formal theories of the commonsense world*. Ablex Publishing, Norwood. New Jersey.

Hillis, D. (1986) *The Connection Machine*. MIT Press. Cambridge, Ma.

Holland, J. (1975) *Adaptation in Natural and Artificial Systems*. University of Michigan Press. Ann Arbor.

Kinzel, W. (1985) *Learning and Pattern Recognition in Spin Glass Models*. In: Condensed Matter. Zetischrift fur Physik B. 60, 205–213. Springer-Verlag. Berlin.

Kirkpatrick, S., C. Gelatt, M. Vecchi (1983) *Optimization by Simulated Annealing*. Science, 220, 671–680.

Kohonen, T. (1988) *Self-organisation and associative memories*. Springer-Verlag. Berlin.

Kosslyn, S.M. (1980) *Image and Mind*, Cambridge Mass. Harvard University Press.

Kowalski, R. (1979) *Logic for Problem Solving*, North-Holland, Amsterdam.

Larkin, J. and H. Simon (1987) *Why a Diagram is (Sometimes) Worth Ten Thousand Words*, Cognitive Science 11, pp. 65–99.

Maes, P. and D. Nardi (1988) *Meta-level architectures and reflection*. Amsterdam, North-Holland Pub.

McCarthy, J. (1960) *Recursive Functions of Symbolic Expressions and Their Computation by Machine*. ACM, p. 184–195, 1960.

Mead, C. (1987) *The artificial retina*. Proceedings of the 1st IEEE conference on neural networks. San Diego Ca.

Michalski, T., J. Carbonell and T. Mitchell (eds) (1983) *Machine Learning – An Artificial Intelligence Approach*. Tioga Pub. Cy. Palo Alto Ca.

Minsky, M. (1975) *A framework for representing knowledge*. in *The Psychology of Computer Vision*. P. Winston (ed.) McGraw–Hill Book Cy, New York.

Minsky, M. (1986) *The Society of Mind*. Basic Books. New York.

Mohnhaupt, M. and B. Neumann (1988) *Some aspects of learning and reorganisation in an analogical representation*, Proc. International workshop on knowledge representation and knowledge (re)organisation in machine learning, K. Morik (Ed.), to appear.

Newell, A. and H. Simon (1972) *Human Problem Solving*. Prentice–Hall, Englewood–Cliffs, New Jersey.

Piaget, J. (1962) *The stages of the intellectual development of the child*. in Wason and Johnson–Laird (eds.) *Thinking and Reasoning*. Harmondsworth, Middx: Penguin. 1968.

Piaget, J. and B. Inhelder (1966) *L'Image mental chez l'enfant*. Paris, P.U.F.

Prade, H. (1985) *A computational approach to approximate reasoning and plausible reasoning with applications to expert systems*, IEEE Trans. Patt. Anal. Mach. Intell. PAMI–7 (3), 260–283.

Prieditis, A. (1988) *Analogica*. London, Pitman.

Shortliffe, E. and B. Buchanan (1975) *A model of inexact reasoning in medicine*. Math. Biosci, 23, 351–379.

Sloman. A. (1971) *Interactions between philosophy and AI – the role of intuition and non–logical reasoning in intelligence*. Proceedings 2nd IJCAI, London.

Steels, L. (1987) *The deepening of expert systems*. AI Communications. Vol 0, 0. North–Holland Pub. Co. Amsterdam.

Steels, L. (1987b) *A selectionist approach to self–organisation*. in Proceedings of the first IEEE Conf on neural networks. San Diego, Cal.

Steels, L. (1987c) *Artificial Intelligence and Complex Dynamics*. IEEE Workshop on advanced architectures for knowledge representation. Mount Fuji, Japan.

Steels, L. (1988a) *Steps towards common sense*. in ECAI–88 Munchen, also VUB AI Lab. Memo 88–3. Brussels.

Steels, L. (1988b) *A reaction–diffusion model of computation*. VUB AI Lab. Memo. 88–5 Brussels.

Toffoli, T. (1984) *Cellular Automata as an alternative to (rather than an approximation of) differential equations in modeling physics*. Physica 10D (1984) 117–127. North–Holland. Amsterdam.

Toffoli, T. and N. Margolus (1987) *Cellular Automata Machines. A new environment for modeling*. MIT Press, Cambridge Ma.

Turing, A. (1952) *The chemical basis for morphogenesis*. Philos. Trans. R. Soc. London (B), 237, 37–72.

Winston, P. (1970) *Learning structural descriptions from examples* MIT Techn. Rep AI–lab. Cambridge, Ma.

Wolfram, S. (1983) *Statistical mechanics of cellular automata. pp. 601–643*. In Rev. Mod. Phys., Vol 55, nº 3. July 1983.

Woods, W. (1970) *Transition Network Grammars for Natural Language Analysis*. Communications of the ACM, vol 13, nº 10.

Zadeh, L. (1978) *Fuzzy sets as a basis for a theory of possibility*. Fuzzy Sets Sys. 1, 3–228.

Daniel Andler

CREA (Centre de Recherche en Épistemologie Appliquée), École Polytéchnique
Paris (France)

Intelligence Artificielle, Philosophie et Sciences Cognitives*

Résumé

Plus de trente ans après sa naissance officielle, l'IA se trouve aujourd'hui prise dans des vents contraires. Est-elle l'une des sciences de la cognition, ou bien est-elle avec ces sciences dans un rapport d'extériorité et d'échange? Est-elle destinée à devenir la *science de la cognition, ou au contraire à se fondre dans une nouvelle discipline?*

L'IA ne demeure pas sans voix, au contraire, d'autant que certaines des attaques dont elle fait l'objet s'adressent globalement au "cognitivisme"; en se portant au secours du cognitivisme ses défenseurs defendent aussi l'IA. On peut néanmoins se demander s'il faut prendre au sérieux la question du rôle de l'IA dans l'étude de la cognition alors que l'échec de ses ambitions sur le plan théorique serait avéré et qu'elle serait donc destinée à disparaitre. Il s'agit là d'un faux problème: l'IA n'est que partiellement liée au cognitivisme, si bien qu'il est prématuré de parler de son échec et que la question de ses rapports aux sciences cognitives mérite d'être posée à nouveau aujourd'hui.

Resumo

Mais de trinta anos após o nascimento oficial, a IA encontra-se hoje envolvida por ventos contrários. Trata-se de uma das ciências da cognição, ou mantém com estas ciências uma relação de exterioridade e intercâmbio? Está destinada a tornar-se a ciência *da cognição, ou, pelo contrário, a fundir-se em uma nova disciplina?*

A IA não ficou sem voz no debate, até porque muitos dos ataques que sofre são dirigidos globalmente contra o cognitivismo; saindo em socorro do cognitivismo, os seus defensores defendem também a IA. Pode, no entanto, duvidar-se se vale a pena tomar a sério o papel da IA no estudo da cognição quando, sendo patente o fracasso das suas ambições teóricas, estaria destinada ao desaparecimento. É um falso problema, este: a ligação da IA ao cognitivismo é somente parcial, pelo que é prematuro falar de fracasso e a questão das suas relações com as ciências cognitivas merece ser recolocada na actualidade.

L'intelligence artificielle est-elle aujourd'hui l'une des sciences de la cognition? Ou bien est-elle avec ces sciences dans un rapport d'extériorité et d'échange, équilibré ou non? Est-elle destinée à devenir la science de la cognition? Ou n'aura-t-elle été au contraire qu'une forme transitoire vouée à se fondre dans une nouvelle discipline? Plus de trente ans après la naissance officielle de l'IA, ces questions prennent un relief particulier: non seulement elles demeurent sans réponse nette, mais leur pertinence est mise en cause par

* Este estudo que, no essencial, versa a matéria da conferência de Daniel Andler no Simpósio de Lisboa, foi publicado na *Revue Internationale de Philosophie*, n.º 172 (n.º dedicado à Inteligência Artificial), 1/1990, pp. 62-86, sob o título "Quelle est la place de l'Intelligence Artificielle dans l'étude de la cognition?". Agradece-se ao seu director, Prof. Michel Meyer, a autorização concedida para a republicação nesta colectanea.

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposyum, 1989,95/111
1992, APDC & SPF, Lisboa

une nouvelle série d'attaques contre la spécialité, alors même qu'elle a acquis une importance, au moins quantitative, et une autonomie institutionelle qui devraient la mettre à l'abri de toute attente[1].

Après avoir essuyé, sans trop de dommages finalement, les critiques d'un Dreyfus[2] et d'un Searle[3], l'IA se trouve en effet aujourd'hui prise dans des vents contraires. Parmi les nombreux facteurs contribuant à jeter le doute sur son avenir, citons le jugement entièrement négatif de Hilary Putnam[4], considéré pendant longtemps comme un allié de poids par l'IA, le développement du connexionnisme[5] et le ralliement à ses thèses (ou à d'autres formes d'opposition) de certains des meilleurs chercheurs de l'IA[6], et enfin l'émergence des sciences cognitives en tant que fédération autonome de disciplines[7].

Certes, l'IA ne demeure pas sans voix, au contraire, d'autant que certaines des attaques dont elle fait l'objet s'adressent globalement au "cognitivisme", doctrine qu'elle partage avec bon nombre de programmes de recherche dans les sciences cognitives[8]. En se portant au secours du cognitivisme, ses défenseurs, tels Jerry Fodor et Zeno Pylyshyn[9] défendent aussi l'IA. On peut néanmoins être tenté de se demander s'il faut prendre au sérieux la question du rôle de l'IA dans l'étude de la cognition alors que son échec serait avéré et qu'elle serait donc destinée à disparaître.

Nous verrons qu'il s'agit là d'un faux problème; que l'intelligence artificielle n'est que partiellement liée au cognitivisme, si bien qu'il est prématuré de parler de son échec et que la question de ses rapports aux sciences cognitives mérite d'être posée à nouveau aujourd'hui.

1. Les ambitions de l'IA

L'un des principaux obstacles que rencontre celui qui se penche sur l'IA, qu'il tente d'évaluer ses chances de succès ou de saisir son articulation avec les sciences cognitives[10], est que l'IA parle par mille bouches; protéiforme, elle ne se laisse saisir que difficilement, et n'hésite pas à reformuler son projet si celui-ci semble l'engager dans une impasse. Si opportune que soit une telle stratégie, elle ne relève pas de l'opportunisme: l'IA hésite sur sa véritable vocation, et ceux qui la pratiquent, comme ceux qui l'observent, se font d'elle les conceptions les plus variées. Quels sont donc les principaux visages de l'IA?

a. Technologie, science, doctrine. L'IA se présente souvent comme l'art de rendre les ordinateurs intelligents, c'est-à-dire ici capables d'exécuter, seuls ou en équipe avec un opérateur humain, des tâches exigeant de l'intelligence. Elle est alors une branche de l'informatique, et dans la mesure où elle se donne pour objectif de réaliser des applications préspécifiées, en s'appuyant sur tout corpus scientifique qu'elle jugera utile, elle est une technologie. La quasi-totalité des travaux actuels en IA relèvent soit exclusivement, soit par certains de leurs aspects, de cette orientation. Ils ne prétendent donc pas accroître nos connaissance théoriques, mais notre savoir-faire.

Cependant, dès l'origine, et encore aujourd'hui dans ses discours officiels à destination externe, comme sous la plume de certains de ses praticiens s'adressant à leurs collègues. l'IA se pose en science. Son objet? L'intelligence, la cognition, les systèmes de traitement de l'information — les avis sont partagés. En tous cas, c'est l'IA-science qui revendique une place parmi les sciences cognitives, et parfois la première.

Enfin, l'IA s'est également longtemps identifiée à une doctrine, à une thèse concernant la nature de l'esprit et de ses capacités. En effet, le projet d'où lui vient son nom et qui lui a permis de se forger une identité, de se créer une image, de mobiliser attention et ressources, était de fabriquer une machine intelligente. Mais il demeurait vide de contenu

tant qu'il ne s'accompagnait pas d'une proposition de méthode. En effet, depuis la magie jusqu'à la procréation, en passant par cent idées inspirées par la mécanique, la thermodynamique, la biologie ou encore, dans le passé de plus proche, la cybernétique[11], les méthodes n'ont pas manqué. L'IA devait donc afficher d'entrée de jeu une conception propre de son objet: si elle s'estimait capable de réussir, c'était en raison de la nature du phénomène à expliquer et à simuler (l'intelligence), laquelle dictait à l'IA sa méthode et marquait sa spécificité en tant que programme de recherche. Ainsi naissait une doctrine qui se confond avec le cognitivisme, ou en constitue une variante informatique[12].

b. Intelligence, esprit, cognition, connaissance. Si l'IA visait d'abord l'intelligence, elle ne doutait pas qu'elle atteindrait l'esprit (*mind*). En effet, l'esprit est selon elle le dépositaire, voire simplement l'hypostase de l'intelligence: l'esprit est ce qui est capable d'intelligence, sans pour autant être l'encéphale: c'est donc le terme collectif qui désigne l'ensemble des capacités du cerveau manifestant cette qualité qu'est l'intelligence, et que notre organe partagera un jour, si du moins l'IA parvient à ses fins, avec d'autres systèmes matériels dits de traitement de l'information. Des origines à nos jours, on pourrait multiplier indéfiniment les citations illustrant l'emploi.

Certains ont cependant commencé à s'interroger sur cette confusion. Observant que l'on pouvait, sans trop forcer le sens du mot, estimer que l'IA s'approchait de l'intelligence, tandis que l'esprit, lié qu'il est aux notions de compréhension et de conscience, demeurait toujours aussi éloigné, ils ont estimé préférable de dissocier les deux termes, ou les deux concepts. C'est ainsi que B. Chandrasekaran propose de considérer l'intelligence comme un *outil* de l'esprit[13] — l'IA se retirerait ainsi, au moins provisoirement, d'un champ où elle rencontre la philosophie. Se cantonnant à la recherche de l'explication et de la production d'*effets d'intelligence*, elle s'exposerait à moins de critiques. Cette prudence est celle de la majorité des praticiens. Pour nombre des porte-parole de la discipline, elle n'est pas acceptable.

Au contraire, en désignant souvent désormais son objet par le terme "cognition" (ou "phénomènes cognitifs"), l'IA élargit officiellement son territoire, qui devient co-extensif à celui des sciences cognitives. Elle étend ainsi sa compétence aux capacités et dispositions de l'esprit humain qui ne font pas intervenir, du moins en première analyse, l'intelligence: la perception, le contrôle moteur, ainsi que des phénomènes tels que ceux qui sont sous-jacents ou antérieurs à l'emploi du langage et au raisonnement: reconnaissance des caractères et des phonèmes, règles grammaticales internalisées, etc.

Enfin, la place grandissante prise par les "connaissances"[14] dans la conception que l'IA se fait de l'intelligence, et dans les simulations qu'elle réalise, a conduit certains à accentuer la différence entre la composante "épistémique" et la composante "heuristique" de l'intelligence[15]. La première, susceptible d'une caractérisation quasi logique, est indépendante de la seconde, qui ne concerne que les techniques de mise en oeuvre des contenus épistémiques. La mission de l'IA se rapproche ainsi de l'une des tâches principales de la philosophie, à savoir décrire et classer les connaissances de l'homme.

c. Absorption, substitution, collaboration. L'attitude de l'IA vis à vis des autres disciplines a varié, à mesure que se modifiait la conception qu'elle se faisait de sa tâche, et surtout que les sciences cognitives commençaient à se dégager de leurs disciplines d'origine pour s'orienter vers la constitution d'une nouvelle fédération, et que la philosophie se découvrait une curiosité nouvelle pour les choses de la cognition.

Au commencement, l'IA n'était que le rassemblement d'une poignée de chercheurs parmi les quelques dizaines qui, dans différentes disciplines, pressentaient l'avènement de ce qui allait devenir les sciences cognitives[16]. Il était sans doute assez naturel alors

d'envisager la création d'une discipline unifiée, qui s'appuyant sur la doctrine "computo-représentationnelle" de l'esprit, reprendrait à nouveaux frais les projets disparates de la psychologie, de la linguistique, de l'anthropologie, de la philosophie. De cette période datent les premières déclarations d'intention "annexionniste". Certaines envisageaient l'absorption des diverses disciplines au sein d'une nouvelle science de l'esprit, qu'il convenait d'appeler intelligence artificielle en raison du *leadership* exercé par le programme de recherche ainsi nommé en 1956. D'autres expliquaient la nécessité pour l'IA, véritable science de l'esprit, à se substituer aux disciplines faillies, trahies par leur défaut de scientificité. C'est ainsi que l'IA allait constituer la première "philosophie expérimentale" — la seule qui vaille — parce que les programmes (informatiques) dans lesquels ses propositions s'expriment sont susceptibles de vérification[17].

Plus tard, quand les sciences cognitives se sont constituées en tant que telles, et que la philosophie a repris ceux des problèmes soulevés par l'IA qu'elle se sentait en mesure d'aborder, l'IA a hésité entre deux attitudes: elle a revendiqué une place parmi les sciences cognitives, à égalité avec chacune d'elles; mais l'instabilité de ce rapport, qui comporte pour elle le risque d'être à son tour absorbée, morceau par morceau, par les sciences cognitives, le résidu retournant en informatique, l'amène périodiquement à revenir à son ambition d'origine: constituer le coeur, sinon la totalité, d'une science unifiée de la cognition.

d. Buts proclamés, buts poursuivis. L'IA a affiché au cours de sa brève histoire maintes ambitions. Mais il est plus facile d'inscrire publiquement un objectif à l'ordre du jour que d'en ôter un que l'on juge hors d'atteinte, ou mal formulé. Aussi s'est-il creusé un écart inquiétant entre les buts proclamés et les buts effectivement poursuivis par l'IA.

Or ce sont les buts proclamés qui façonnent en grande partie la perception que l'IA a d'elle-même et offre aux observateurs, car il est difficile, sinon impossible, de déterminer objectivement et synthétiquement ce que sont les buts qu'elle poursuit, pour la raison habituelle que toute action peut répondre à une infinité d'intentions, à laquelle s'ajoutent ici une extraordinaire variété des recherches et des méthodologies, et le caractère particulièrement évanescent du but ultime — l'intelligence. Pourtant, il serait regrettable de prononcer sur l'IA un jugement négatif fondé sur le seul écart entre les buts proclamés (ici ou là) et les buts atteints. Un bilan plus équilibré exige de comparer ceux-ci aux buts poursuivis[18].

Pour nous qui nous interrogeons sur la contribution qu'apporte ou peut apporter l'IA à l'étude de la cognition, il importe de distinguer ce que Putnam appelle sa tâche "notionnelle" de sa tâche "réelle"[19]. C'est à quoi l'examen des productions de l'IA nous aidera.

2. Les fruits de l'IA

Face à cet éventail toujours changeant d'ambitions, il serait déjà malaisé de décider si elles sont en voie d'être réalisées, ou semblent au contraire hors d'atteinte. Mais la saisie synthétique des accomplissements de l'IA, avant même toute comparaison avec des objectifs affichés, est d'une difficulté notoire. Cela tient en partie à la méthodologie et au style d'exposition de la discipline, en partie au caractère disparate des travaux qui s'en réclament. Ces traits n'ont du reste rien de contingent, et entrent eux-mêmes dans tout bilan qu'on pourrait entreprendre.

Ce n'est pas à un tel bilan qu'il est possible ici de s'atteler. On tentera seulement de dresser une sorte d'inventaire préliminaire, en distinguant les diverses espèces de biens

produits par l'IA, selon leur nature et les ingrédients entrant dans leur composition. Il deviendra alors possible de s'interroger sur la nature des liens — identité, rivalité, complémentarité... — qui unissent l'IA aux autres disciplines.

a. Méthode et forme. Examinons d'abord les "sources de connaissance"[20] auxquelles puise l'IA.

La première, et la principale, est l'informatique. C'est elle qui fournit le cadre de référence fondamental, même s'il n'est pas toujous visible. Le rappeler, c'est s'en tenir à une banalité, malgré les arguments de certains chercheurs qui veulent s'affranchir de l'ordinateur. Certes, l'IA se conçoit parfois comme l'étude de l'intelligence *humaine*[21], et ne doit considérer alors *en principe* l'ordinateur que comme un modèle, et comme un outil, parmi d'autres. Certes, l'accent mis sur la composante épistémique raproche l'IA de la logique et l'eloigne de l'informatique. Certes, nombre de textes ne font aucune référence à l'ordinateur, nombre de chercheurs s'en tiennent aussi éloignés que possible dans leur pratique, et certains prônent une abstinence méthodologique[22] à son égard. Il demeure que l'informatique fournit le cadre institutionnel des recherches et la formation de la plupart des chercheurs, le vocabulaire et le medium formel (les langages Lisp et Prolog, les structures de représentation) et surtout les critères préalable et ultime de succès: toute explication que l'on propose pour une capacité cognitive donnée doit être clairement compatible avec les exigences de l'ordinateur, et doit conduire en principe à un programme conférant à l'ordinateur la capacité visée.

Ces manifestations d'indépendance à l'égard de l'informatique témoignent en revanche de l'existence d'autres sources auxquelles l'IA puise. L'une est constituée des mathématiques et de la logique au sens le plus large — les sciences formelles. Le calcul des probabilités, la statistique, les logiques classiques de divers ordres et les logiques non classiques, l'analyse harmonique, la topologie et la géométrie sont quelques-unes des théories auxquelles l'IA fait appel régulièrement. On pourrait ranger sous la même rubrique des théories relevant de l'informatique théorique, plus proches de la logique ou de l'algèbre que du reste de l'informatique.

Les sciences cognitives sont pour l'IA une autre source de connaissances — de concepts, de méthodes, de résultats. Cette évidence passe parfois inaperçue, et parfois est masquée par la tendance de l'IA à modifier, soit substantiellement soit terminologiquement, ses emprunts, par son ambition de retrouver par ses propres moyens, et sous les formes qui lui conviennent, les voies tracées par la linguistique ou la psychologie, par la "bi-appartenance" de certains de ses chercheurs — linguistes, psychologues ou mathématiciens de formation recrutés par des laboratoires d'IA —, et enfin par le recours à la synecdoque consistant à désigner le tout de l'entreprise cognitive par le nom d'une de ses parties, l'IA.

La dette de l'IA à l'égard de la philosophie, discipline aux contours flous et changeants, est moins facile à déterminer. S'il est certain que des notions telles que représentation, symbole, système formel, structure, règle, schéma, inférence, voire connaissance, sont d'origine philosophique, il serait difficile de démontrer que l'IA les a toujours trouvés sous la plume de philosophes. Elles appartiennent pour la plupart à un vocabulaire partagé par bon nombre de sciences, ainsi que par le sens commun.

C'est justement dans le sens commun, dans l'intuition du chercheur que réside la dernière des sources de l'IA, et non la moindre. Formé le plus souvent par l'informatique, la logique ou les sciences de l'ingénieur, parfois encore par la physique, le chercheur en IA aborde son problème de modélisation avec un mélange d'expérience (sur le plan informatique) et d'innocence (quant au phénomène à simuler). Pour le spécialiste formé en philosophie ou dans une science humaine, il y a là un motif de perplexité. Il ne conteste

certes pas que la frontière entre sens commun et science est floue. Il reconnaît également que l'intuition n'est pas une donnée invariable, et qu'elle se modifie avec l'expérience. Il n'empêche, selon lui, que nombre de phénomènes ne sont susceptibles d'une description naturelle et exacte que dans le vocabulaire de la science dont il relève. L'élaboration d'une telle description, puis dans un deuxième temps la mise au jour de régularités, de connexions ou de liens causaux rattachant le phénomène étudié à d'autres, ou reliant ses divers aspects, voilà les tâches dont il est investi, voilà celles qui occupent son temps et son esprit. Dès lors, il ne peut que s'interroger sur l'intérêt et sur les chances de succès des investigations de son collègue d'IA. Il est tenté de ne voir pour lui qu'une alternative entre l'amateurisme et le ralliement à sa propre discipline — entre l'impasse et l'emprunt.

Cette conclusion semble prématurée. Nous devons d'abord nous interroger sur la forme que prennent les productions de l'IA, puis considérer ses contributions réelles. Trois types de textes émanent d'elle. L'immense majorité est composée, du moins au premier regard, de programmes. En fait, ce qui est diffusé est invariablement une description plus ou moins complète d'un programme ou d'une famille de programmes, accompagnée de commentaires sur les motivations de son constructeur; il arrive souvent que de plus le programme n'existe nulle part sous forme achevée, que seules certaines parties, ou qu'un modèle réduit, soient réalisés. Cependant, c'est bien le programme idéalement achevé qui est présenté comme le véritable résultat de la recherche.

Une seconde catégorie de documents est constituée de théories formelles. Elles portent soit sur les algorithmes utilisés en IA, soit, plus rarement, sur la structure d'ensembles abstraits de données, ou sur l'acquisition de données. Par leur forme, ces travaux sont très proches de la logique.

Enfin, l'IA produit une grande quantité de textes programmatiques, méthodologiques, critiques, dans lesquels à des considérations intuitives sur l'objet de l'IA en général, ou sur une catégorie particulière de phénomènes à modéliser, se mêlent une évaluation des recherches, des spéculations prospectives et des réflexions sur le rôle et l'avenir de l'IA, sur ses méthodes, sur ses bases institutionnelles, sur ses effets sociaux, sur les rapports entre science et applications, etc. Ces textes jouent un rôle essentiel dans la diffusion des idées et de l'image de l'IA, non seulement à l'extérieur, mais aussi à l'intérieur. En effect, le faible taux de théorisations de l'IA et l'éparpillement de ses programmes de recherche ne lui permettent pas de fonder sa culture sur un corpus théorique ou un objectif opératoire commun (l'objectif de doter l'ordinateur d'intelligence demeurant purement "notionnel"). Aussi est-ce le commentaire et la réflexion métathéorique qui donnent à l'IA l'occasion de consolider une unité problématique.

b. Contenus. Venons-en à l'évaluation des contributions substantielles de l'IA. Elle dépend de la perspective choisie: l'IA prétend en effet tantôt apporter des solutions aux problèmes que se posent les sciences cognitives et la philosophie, tantôt progresser dans un domaine qu'elle possède en propre.

Dans la première perspective, celle d'une problématique partagée avec les autres sciences, nous prendrons le risque d'affirmer que le bilan est dans l'ensemble négatif.

Commençons en effet par les grandes questions — celles qui portent sur la nature de l'esprit, de l'intelligence, de la connaissance, du langage, de la communication..., et auxquelles se consacrent d'une part les sciences cognitives expérimentales, d'autre part la philosophie. Deux positions semblent défendables. La première est que les progrès, quelle qu'en soit l'origine, restent modestes et fragiles. La seconde est qu'ils sont considérables, et qu'ils sont le fait de toutes les disciplines concernées, sauf précisément de l'intelligence artificielle. Le fait que le modèle de l'ordinateur ait joué un rôle important, ou que

l'entreprise de l'IA ait pu amener telle ou telle discipline à repenser sa stratégie ne doit pas nous induire en erreur: s'il s'agit de recenser les conceptions ou les thèses importantes (concernant, rappelons-le, les grandes questions) actuellement actives dans les sciences de la cognition ou en philosophie, aucune ne provient de l'IA. Naturellement, une telle affirmation soulève nécessairement des objections auxquelles on ne cherchera pas à répondre d'avance. On se contentera de témoigner que selon maints observateurs qualifiés, une inspection des grandes courants, problèmes et corps de doctrine jouant un rôle actif dans le domaine ne révèle aucune influence majeure de l'IA en tant que telle[23]. C'est ainsi que Putnam a pris au sérieux l'ambition affichée par l'IA de faire la lumière sur les grands problèmes de la connaissance (de l'*epistemology*)[24], et n'a cru pouvoir mettre en évidence qu'une ignorance de ces problèmes, ou une incapacité à en prendre la vraie mesure. En linguistique, le diagnostic est similaire; en psychologie il est sans doute moins tranché, car la position uniformément prudente ou pessimiste, selon laquelle les progrès majeurs sont de toute manière incertains, y est plus répandue qu'ailleurs.

Examinons en second lieu les questions de portée limitée, celles qui font l'objet, en IA ou dans les sciences cognitives expérimentales, de recherches spécialisées. La situation est toute différente: les candidats au titre de contribution positive de l'IA, loin d'être invisibles, pullulent. Chaque programme, chaque commentaire ou suggestion susceptible de conduire à un programme permettant à l'ordinateur de simuler une capacité cognitive constituent *ipso facto* des contributions *potentielles* à la question de la nature et de l'explication de cette capacité. Pourtant, ce ne sont pas des contributions *effectives*.

La raison, dans certains cas, est simple et reconnue. Il en est ainsi lorsque la capacité étudiée est réalisée par l'ordinateur avec des moyens qui diffèrent manifestement de ceux que l'esprit peut concevablement mettre en oeuvre. Il n'est personne aujourd'hui pour penser que la mémoire d'une machine de von Neumann, qu'une calculatrice électronique ou qu'un système d'aide à la conception nous apprennent quoi que ce soit sur la mémoire, l'aptitude arithmétique ou le processus de conception chez l'homme.

Mais le plus souvent la raison est ailleurs. La plupart des programmes sont en effet fondés, en partie du moins, sur l'imitation de l'homme (faute d'une caractérisation indépendante du phénomène cognitif étudié). Aussi se présentent-ils comme des modèles, des schématisations ou idéalisations de phénomènes humains, et sont-ils des candidats sérieux à la dignité de contribution à la psychologie ou à la linguistique. S'il convient cependant, me semble-t-il, de leur dénier cette dignité, c'est qu'il leur manque à la fois le caractère de théorie et la mise à l'épreuve par l'expérience.

Les programmes peuvent-ils être considérés comme des théories? On l'a prétendu[25], ajoutant parfois qu'ils ont sur les théories ordinaires l'avantage d'être nécessairement rigoureux, car écrits dans un langage entièrement formalisé, et d'être directement et automatiquement vérifiables (ou du moins falsifiables): il suffit de faire "tourner" le programme; tant qu'il obtient le résultat souhaité, il est corroboré. Il est certain que, s'il est effectivement achevé et en mesure de piloter l'ordinateur ou le robot, le programme comporte une garantie de vérification que les théories n'ont généralement pas. Cette garantie est toutefois limitée: pas plus que n'importe quel programme, ceux de l'IA ne comportent de garantie de *correction*: rien ne permet d'affirmer que leur comportement sera celui qui leur est demandé (en termes formels, rien ne permet de s'assurer à l'avance que la fonction calculée est celle que le programme est censée calculer). Il est donc impossible[26] de remonter de la "performance" de la machine à sa "compétence", c'est-à-dire à la définition intensionelle de ses capacités.

D'autre part, l'acceptation d'un programme *en tant que théorie* dépend de sa capacité

à donner entièrement satisfaction; or à ce jour, aucun programme ne s'approche même du but de simuler, dans un domaine naturel et suffisamment étendu de variation des conditions, une fonction cognitive complexe. Il conviendrait donc, semble-t-il, de rejeter tous les programmes écrits à ce jour. Si l'on s'en abstient, c'est pour deux raisons. La première est que l'intuition demeure que l'*intention* qui a présidé à la rédaction du programme peut comporter des éléments de vérité qu'il faut conserver. La seconde est que cette intention a pu être trahie par l'une des multiples décisions d'intendance qu'il a fallu prendre pour permettre au programme de tourner, ou par une combinaison d'entre elles -- décisions qui ne sont dictées par aucune considération liée à la théorie sous-jacente du phenomène à simuler, mais par des nécessités algorithmiques. Ces choix, arbitraires au regard de la théorie viennent pallier son *incomplétude*. On voit donc se retourner l'argument de l'effectivité: la précision complète du programme n'est obtenue qu'au prix de choix dépourvus de pertinence théorique — lesquels font obstacle à l'évaluation des programmes toujours imparfaits que l'IA a produits à ce jour.

Nous venons de voir que les programmes existants ne peuvent fournir directement aucune théorie, d'une part parce qu'ils sont en pratique défectueux, et qu'il est impossible d'en extraire systématiquement la part de vérité; d'autre part parce que, même parfaits, ils sont voués à demeurer implicites, en un sens que l'on a précisé. (Qu'un programme puisse *suggérer* une théorie n'est pas ce qu'on a voulu contester; il s'agit là d'une évidence: presque n'importe quoi peut en faire autant.) Une voie reste cependant ouverte ao partisan de la thèse des programmes comme théories. Il peut arguer en effet d'abord que rien n'exclut la possibilité de programmes entièrement satisfaisants, ensuite que de tels programmes, qui verront peut-être le jour demain, devront être acceptés tels quels, et non pas après avoir subi une traduction leur conférant la forme d'une théorie traditionnelle. La force de sa thèse, peut-on penser, est d'accepter l'impossibilité théorique ou pratique, dans le domaine de la cognition, de théories au sens traditionnel, et de leur substituer un nouveau type d'explication qui ne s'y ramène effectivement pas.

Il s'agit là d'un de ces appels à une réforme fondamentale de méthode qui ponctuent régulièrement l'histoire de l'IA. Si l'on hésite à y répondre, c'est qu'il paraît difficile de renoncer, en matière d'explication scientifique ou philosophique, à certains critères — en l'espèce, ceux d'intelligibilité, de pertinence et de cumulativité. Or un programme est littéralement illisible pour l'oeil non exercé, et pour l'oeil entraîné exige un patient et incertain déchiffrage; son *intelligibilité* est donc conditionnée soit par le développement d'une nouvelle intuition, résultat d'une longue évolution culturelle qu'il nous est impossible de prévoir; soit par la présence de commentaires, à défaut d'une traduction. La *pertinence* du programme pour l'explication du phénomène étudié est également tributaire d'un commentaire: non seulement rien n'indique que le processus concrétisé par le programme est identique au phémomène, ou en constitue du moins un modèle, une idéalisation, mais il se peut aussi que l'essentiel de l'efficacité apparente du programme soit due à des choix opportuns de primitives ou d'agencement — on a alors affaire à une véritable pétition de principe, le phénomène à expliquer étant introduit clandestinement dans l'explication même. Certes, les théories classiques ne comportent pas elles non plus de garantie de pertinence et de non-circularité! Du moins sont-elles dans l'obligation d'offrir des méta-arguments en leur faveur. C'est cette fonction argumentative qui fait défaut aux programmes.

Ce n'est par conséquent, nous le voyons, qu'accompagné d'un commentaire que le programme peut répondre à deux des exigences cardinales de l'explication théorique. Qu'en est-il de la capacité de s'agréger, de s'accumuler et de constituer ainsi graduellement

un savoir? La réponse est ici moins nette. L'une des vertus des programmes est en effet, contrairement à tant de théories, en particulier dans le domaine des sciences de l'homme et en philosophie, de pouvoir en principe s'emboîter ou s'''appeler'' à l'infini, y compris récursivement. La difficulté est que c'est précisément dans l'agencement des programmes que pourrait résider l'essentiel de l'explication des grands phénomènes simulés (à l'extrême, l'intelligence elle-même). Or cet agencement demeure informulé, que ce soit sous forme de théorie ou sous forme de programme. La théorie ''architecturale'' unifiée qu'appelle de ses voeux Allen Newell[27] aurai précisément pour fonction de montrer ce qu'on peut attendre de la concaténation adaptée des programmes (ou des théories), et de dégager les principes et le contenu théorique d'une telle concaténation. Il semble cependant que sur ce plan tout reste à faire.

La seconde grande raison pour laquelle les programmes ne peuvent faire office de théories est liée à la première. C'est qu'ils ne se laissent pas facilement évaluer expérimentalement. N'ayant pas la forme de théories, ils sont mal adaptés aux procédures de mise à l'épreuve expérimentale, et le travail de l'IA se sépare ainsi de fait de celui des sciences cognitives expérimentales. Il est possible qu'à mesure que se développeront les collaborations interdisciplinaires, cette situation vienne à se modifier suffisamment pour que l'écriture de programmes se substitue en partie à l'édification de théories comme pôle du mouvement d'aller-retour entre expérimentation et théorisation caractéristique de la démarche des sciences expérimentales. Encore ce mouvement ne se déroulera-t-il sans doute pas à l'intérieur de l'IA, mais s'insèrera dans un commerce entre l'IA et les sciences cognitives concernées.

Il nous reste à examiner les contributions à l'élucidation de questions précises que l'on pourrait espérer retirer des autres textes produits par l'IA. Réservant le cas des théories formelles, qui relèvent soit des exceptions dont il sera question dans un instant, soit du développement interne de l'IA qui sera évoqué ensuite, considérons les textes qui accompagnent les programmes, ou appartiennent à la troisième espèce définie plus haut: textes programmatiques, méthodologiques, évaluatifs, prospectifs etc. La raison pour laquelle ils ne sont généralement pas retenus comme textes scientifiques ou philosophiques est double. D'une part ils n'observent, dans le choix de leurs prémisses et dans leur argumentation, aucune méthodologie précise. D'autre part, ils ne font pas appel à l'expérimentation. Ils encourent ainsi les deux principaux reproches que l'on aime à se lancer d'un pôle à l'autre de l'univers des sciences cognitives: ils ont le caractère spéculatif de la ''philosophie dans son fauteuil'' (*armchair philosophizing*), sans en avoir la rigueur et l'ossature théorique qui font aussi défaut à tant de recherches empiriques. Surprenant retournement d'un argument que l'IA avance en sa propre faveur, qui est d'être à la fois expérimentale[28] et rigoureuse!

Sans quitter encore le terrain que l'IA veut partager avec les sciences cognitives, venons-en à une série importante de travaux qui font exception au jugement négatif porté jusqu'à présent. Il s'agit d'une part de contributions à des recherches interdisciplinaires, d'autre part de recherches qui ne semblent relever de l'IA que pour des raisons purement contingentes, d'ordre institutionnel notamment. Les premières échappent par définition, à l'heure actuelle, à la fois dans leur nature et dans leurs résultats, à toute revendication par une discipline. Aussi ne plaident-elles ni pour ni contre la thèse d'une contribution spécifique de l'IA en tant que telle aux sciences de la cognition: que l'IA participe à l'effort collectif, c'est une évidence que personne ne devrait chercher à nier; qu'elle y apporte des concepts et des méthodes propres, qu'elle façonne les résultats à son image est précisément ce qui dans les travaux interdisciplinaires reste indéterminé.

Tout autre est le cas des travaux — ils portent principalement sur le langage naturel, la logique et les théories formelles de l'apprentissage, de la décision ou des jeux — dont rien n'indique l'appartenance à l'IA, sinon le rattachement de leurs auteurs et la source de leur financement. Rien en effet ne s'opposerait à ce que ces travaux se rangent en logique, en linguistique, en philosophie, en économie ou en psychologie théorique. Les auteurs ont bien souvent reçu leur formation initiale dans ces domaines, ou bien, provenant de l'informatique ou des sciences de l'ingénieur, ils s'y sont formés ultérieurement. Les échanges les plus significatifs ont lieu avec des collègues rattachés à ces domaines, et bien souvent dans des publications spécialisées. C'est donc de manière quelque peu clandestine que l'IA produit en son sein des contributions directes aux sciences cognitives et aux disciplines de base. Il n'y a rien là qu'il faille déplorer, rien non plus qui vienne étayer l'hypothèse d'une contribution spécifique de l'IA.

Il est temps de se pencher sur les contributions de l'IA à ce qui est sa problématique ou son terrain propre, quelle que soit la difficulté qu'il y ait à en donner une définition préalable. Trois axes se laissent discerner dans le développement scientifique et technologique de l'IA[29].

Les plus important quantitativement, et le plus facile à caractériser est celui que l'on pourrait désigner par l'appellation de "IA au sens strict ou étroit", ou par celle de "génie cognitif au sens large" (c'est-à-dire non limité au *knowledge engineering* et aux techniques des systèmes experts). Il s'agit de l'ensemble des efforts destinés à étendre les capacités des ordinateurs et des robots. Sur ce terrain, nul ne songe à contester le rôle clé, sinon exclusif, de l'IA. Quant à ses progrès, ils font l'objet d'appréciations divergentes, même de la part de ses spécialistes[30]. La principale raison de s'interroger sur la compétence de l'IA est fournie par les succès du connexionnisme, lequel peut être vu comme un concurrent de l'IA classique et offrir une réfutation de ses présupposés. Mais si l'on convient de replacer le connexionnisme au sein l'IA, ce à quoi on est amené lorsqu'on dissocie l'IA au sens strict de la doctrine cognitiviste à laquelle l'IA au sens large s'est longtemps identifiée, on n'a plus aucun motif sérieux de mettre en question la légitimité de l'entreprise ni d'en prédire l'échec définitif.

A cet axe premier et principal se rattache un domaine qui pour plusieurs raissons constitue désormais un second axe à lui tout seul. Il s'agit de la vision artificielle[31], qui mériterait le nom de théorie abstraite de la vision si elle n'incluait pas également les aspects strictement technologiques, notamment informatiques et électroniques, du projet. Si, tout en demeurant intégré à l'IA, cette nouvelle spécialité en constitue une branche à part, c'est parce qu'elle est conceptuellement largement indépendante du reste de l'IA, et a acquis une cohérence propre, qui explique sa force scientifique et institutionnelle.

Le troisième axe propre à l'IA est le plus problématique. Il prend appui sur la notion de traitement abstrait de l'information. L'hypothèse originelle de l'IA, selon laquelle le niveau principal d'explication, en ce qui concerne l'intelligence ou la cognition, se situe non seulement au-dessus de celui de la neurobiologie et de l'électronique, mais aussi de la psychologie et de l'informatique, reste active. David Marr a su lui donner une formulation d'autant plus convaincante qu'il s'est montré acerbe dans ses critiques de certaines faiblesses méthodologiques et conceptuelles de l'IA[32]. Ni les échecs dans certaines réalisations, ni les arguments des philosophes, ni les attaques de certains partisans du connexionnisme assimilé à une neuroscience cognitive ou abstraite[33] ne l'ont définitivement réfutée. Tout au plus est-on davantage enclin à admettre, en IA et généralement dans les sciences cognitives de tendance théorique, que ce niveau de traitement de l'information n'est peut-être pas le seul à être essentiel. La possibilité

demeure de caractériser à ce niveau des régularités importantes. Qu'il s'agisse encore d'une promesse, ou plus exactement que les observateurs soient encore loin d'être persuadés que l'IA a su dégager ce niveau abstrait d'explication et à y inscrire un véritable commencement de théorie ne compromet pas radicalement l'entreprise. Nul doute que l'avenir de l'IA comme science indépendante ne se joue largement là.

3. Le rôle présent et futur de l'IA

Il paraît possible dès maintenant de formuler certaines conclusions concernant le rôle que l'IA jouera à l'avenir, en se fondant sur l'observation du rôle qu'elle a effectivement joué jusqu'à présent.

a. Science fondamentale ou melting pot? L'IA a parfois exprimé l'ambition d'être *la* science fondamentale de la cognition. Pourvue d'un langage universel, celui des algorithmes ou celui des programmes, et d'une méthodologie analytico-synthétique irréprochable, elle unifierait les résultats dispersés obtenus dans les différentes sciences et érigerait une théorie générale de l'esprit et de l'intelligence, fournissant ainsi tant la base d'explications et de prédictions détaillées des phénomènes cognitifs naturels, que celle de simulations par l'ordinateur de ces mêmes phénomènes.

Rien ne permet, semble-t-il, de penser que cette ambition soit en passe de se réaliser. D'autant moins que l'unité de l'IA elle-même cette "interdiscipline" dont parle Newell[34] est plus problématique que jamais. Si l'IA a cependant bel et bien joué un rôle singulier et notamment fédérateur dans le développement des recherches sur la cognition, c'est pour des raisons d'ordre sociologique. Dépourvue de traditions bien établies, elle présente, face aux sciences humaines lourdes de souvenirs, une aire de liberté conceptuelle. N'ayant pas accumulé les tentatives inabouties, ni constitué à chaque occasion — échec, nouveau départ — des clans fortement opposés les uns aux autres, elle autorise le chercheur, souvent frais émoulu d'études scientifiques (au sens restreint du terme), à tracer librement son chemin entre théories constituées, intuitions personnelles et procédés de programmation. Les chercheurs déjà formés dans les sciences humaines trouvent quant à eux dans les laboratoires d'IA à la fois la stimulation et l'absence de contraintes propices aux idées nouvelles. Ils y croisent des collègues d'autres disciplines, et autour d'un projet commun nouent un dialogue que les institutions traditionnelles ne favorisent pas. Bref, l'IA a quelque chose d'un Nouveau Monde, à la fois terre d'accueil, *melting pot* et nouvelle frontière. On s'y sent libre de définir de nouvelles règles de méthode, de nouveaux critères de sérieux et de réussite. On s'y sent à l'abri de la censure qu'exercent, plus ou moins ouvertement, les courants majoritaires dans les disciplines établies — que ce soit dans le choix des sujets de recherche ou dans le style d'exposition. Cela favorise même des travaux de type traditionnel; un exemple frappant en est donné par les recherches sur les logiques non classiques, qui connaissent depuis quelques années un développement considérable sans devoir à l'IA davantage qu'un cadre accueillant — laboratoires, revues, congrès... — que celle-ci lui fournit, il est vrai, en raison de ses propres besoins: c'est donc aussi en leur assignant un but que l'IA encourage ce genre de travaux.

L'IA procure donc, du moins dans certains cas[35], un cadre et un climat propice à maintes recherches relevant des domaines les plus variés. Peut-elle à l'avenir contribuer d'une autre manière à créer une communauté d'intérêt, un langage commun? Il est trop tôt pour le dire; mais si c'est le cas, ce sera sans doute davantage en imposant un certain ton, un certain style d'approche qu'en fournissant un cadre conceptuel.

b. Contributions aux sciences particulières. Une seconde ambition de l'IA est de

fournir des solutions aux problèmes que se posent la philosophie et les diverses sciences cognitives, séparément ou collectivement, et plus généralement de contribuer à ces disciplines.

Or, nous l'avons vu, les apports de l'IA ne sont guère perçus par les disciplines en question comme des solutions — sauf dans certains cas où ont prospéré, au sein de l'IA, des programmes de recherche féconds mais sans rapport évident, sinon institutionnel, avec l'IA. Deux questions demeurent donc ouvertes: que vont devenir ces recherches spéciales? Préfigurent-elles la forme que pourrait prendre demain la contribution de l'IA aux problèmes centraux qu'elle partage avec les sciences cognitives?

S'il est difficile de prédire l'évolution des institutions (et notamment des étiquettes), on peut dès à présent discerner une tendance à la constitution, au sein de l'IA, d'essaims centrés autour de recherches spécialisées telles que la vision artificielle, la logique, le traitement du langage naturel ou le connexionnisme. Ces essaims portent d'ailleurs un nom, celui de "*community*", et leur réalité n'est guère contestable. Mais on peut se demander s'ils ne sont pas simplement des branches de l'IA, destinées à en demeurer, ou en devenir, en raison même de leur solidité, les piliers. On se risquera à affirmer que ce n'est pas le cas, pour les deux raixons suivantes: d'une part, ils ne sont que très partiellement liés à l'histoire et aux choix stratégiques de l'IA; d'autre part, ils ont à l'inverse des liens très étroits avec d'autres sciences. Ils se distinguent clairement en cela d'autres branches telles que le *knowledge engineering* (systèmes experts ou "à base de connaissances"), le *machine learning* ou la résolution heuristique de problèmes. Dès lors, on peut raisonnablement s'attendre à ce qu'ils s'établissent à leur compte dans la galaxie des sciences cognitives, à égale distance de l'IA et d'autres disciplines.

Quant à savoir si demain l'IA pourrait apporter des éléments de solution à certains problèmes centraux des sciences cognitives, c'est évidemment fort difficile à dire. On peut cependant s'interroger sur les raisons de la relative stérilité dont elle a sur ce plan fait preuve jusqu'à présent. Ne faut-il pas la chercher dans son insistance pour trouver des solutions positives et des confirmations, alors que le progrès scientifique et philosophique se mesure en partie à l'attention prêtée aux impasses, aux limitations, aux contradictions, aux infirmations de toutes sortes? L'IA n'a ni le temps ni la volonté de ce travail sur le négatif. Son tempérament est celui des constructeurs d'objets, d'édifices, de dispositifs concrets: c'est à réussir, plus qu'à décrire et comprendre, qu'elle tend naturellement.

Ce décalage n'est nulle part plus manifeste qu'en matière philosophique, au point que l'on peut se demander, avec Putnam, ce que l'IA fournit à la philosophie, et avec Newell ce que celle-ci apporte à celle-là. Contentons-nous d'observer que dès qu'elle pose le problème de l'intelligence, celui de la mémoire, du savoir, du raisonnement, l'IA se place sur le terrain de la philosophie. Dès lors les notions qu'elle met en oeuvre, les modèles qu'elle construit, les programmes par lesquels elle les met à l'épreuve, et qui sont, comme le dit Dennett, autant d'expériences de pensée, mais suffisamment complexes pour conduire à des résultats surprenants[36] — tout cela constitue à la fois un matériau (l'équivalent de données empiriques) pour les théories philosophiques de l'esprit, du langage etc., et un objet d'étude pour l'épistémologie.

c. Modèle ou modèles. La force de l'IA, contrairement à ce que pensent certains de ses fondateurs, n'est pas de proposer ou d'illustrer une théorie de l'esprit (de l'intelligence, du savoir, du langage etc.) particulière, bien moins encore de prouver sa justesse, même partielle. L'IA n'a rien à gagner, et beaucoup à perdre, à confondre sa cause avec celle d'un modèle particulier, et ses bases avec la doctrine cognitiviste. C'est ce que montre d'abondance l'épisode connexionniste, au cours duquel l'IA, après avoir voulu éliminer le

nouveau venu parce qu'il met en cause, lui semblait-elle, ses fondements mêmes, se décide à l'adopter parce qu'au contraire, comme elle s'en avise, il partage ses objectifs et lui apporte des idées prometteuses. Ce que l'IA offre aux sciences cognitives, ce ne sont ni des solutions, ni des théories, ni des preuves, ni un modèle unique (qui vaudrait solution, théorie et preuve), ce sont *des* modèles.

***

Nous voici parvenus au point où, pour en dire plus sur l'IA, il faudrait en savoir plus sur l'esprit. Si la question de l'objet, de la place et de la contribution de l'IA est si difficile, c'est que les réponses aux grandes interrogations restent pour l'essentiel sans réponse. L'esprit est-il susceptible d'un savoir comparable à la physique, organisé autour de grands principes peu nombreux, ou est-il au contraire un entrelacs de minuscules dispositifs, accumulés et combinés au cours de millions d'années d'évolution naturelle, échappant à toute description synthétique?

Telle est la grande question dont certains veulent faire dépendre l'avenir de l'IA. Il faut d'abord la préciser. Ce ne sont pas deux, mais quatre possibilités qui se présentent. La première est que l'esprit est fondamentalement simple. Dès lors, ses lois existent et finiront par être dégagées; ce seront des principes eux-mêmes simples et peu nombreux. Tel était le pari de l'IA à ses débuts; elle était en quête de l'essence, du noyau, du mécanisme clé. Or non seulement cette quête a été vaine (ce qu'on ne saurait reprocher à l'IA), mais les progrès de l'IA et des sciences cognitives sur presque tous les fronts ont été marqués par une grande lenteur. Aussi l'hypothèse de la simplicité fondamentale paraît-elle aujourd'hui peu plausible[37].

La seconde possibilité est que l'esprit est *complexe*, c'est-à-dire que tout en comportant un grand nombre d'unités et de parties fonctionnelles, et en impliquant de longues chaînes d'interaction, il n'en est pas moins susceptible à un niveau approprié d'une description intelligible, s'ordonnant autour de principes organisateurs en nombre restreint. Cette conception, qui est aujourd'hui aussi bien celle des sciences physiques que des sciences du vivant, nourrit la pensée et l'espoir de nombreux chercheurs dans les sciences cognitives, tels Chomsky.

La troisième possibilité est que l'esprit est *compliqué*, et ne présente aucune régularité d'ordre supérieur susceptible de simplifier sa description. Chaotique, il ne laisse aucune prise à l'explication, et autorise seulement la description, et éventuellement la simulation, unité par unité, ou encore l'évocation globale, par des moyens non scientifiques, littéraires par exemple. C'est à cette hypothèse que se rallient les sceptiques, tels H. Dreyfus.

La quatrième et dernière possibilité, sur laquelle insiste Dennett, est que la vérité combine les trois hypothèses précédentes: l'esprit résulterait d'un agencement complexe d'éléments simples, donnant lieu à des descriptions mêlant des dispositifs disparates et arbitraires à des principes d'organisation. L'esprit serait une affaire d'*architecture*, l'intelligence naturelle le résultat d'une Nature évolutionnaire faisant oeuvre d'ingénieur, et la méthode de l'IA consisterait à parcourir dans l'ordre inverse les opérations d'agencement de l'évolution.

Chomsky estime que l'IA n'a d'intérêt (pour l'étude de la cognition) que dans la troisième hypothèse — et qu'alors ce sont les sciences cognitives telles qu'il les conçoit, que les préfigure la linguistique théorique, qui sont vouées à l'impasse. Ces dernières ont au contraire une chance d'aboutir dans la seconde hypothèse. Dennett, comme Newell, estiment qu'à défaut de la première ou de la seconde, la quatrième hypothèse préserve à la

fois les chances et l'intérêt de l'IA. Mais est-il bien vrai que le sort de l'IA soit lié à la vérité de l'une de ces hypothèses? Ce n'est pas sur. Qu'il s'agisse de mettre au jour des lois, des niveaux de description simple, des agencements ou l'absence de tout principe, ce sont les sciences cognitives (éventuellement réunies en *une* science de la cognition) qui ont pour vocation de dire ce que la science peut dire de l'esprit (dans sa définition restrictive d'organe cognitif). Le rôle de l'IA est tout autre, si plate que puisse sembler cette conclusion au terme d'une trop longue discussion: produire dans la machine des effets d'intelligence. C'est là ce à quoi elle s'emploie effectivement, c'est ce qu'elle se donne les moyens de réussir. L'organisme est-il simple, complexe, chaotique, ou bien combine-t-il principes et "bricolage"? De la réponse dépend la forme que prendront finalement les sciences biologiques, mais non leur existence, ni d'autre part celle de la médecine et des autres arts du corps. A côté de la physique et de la chimie, qui ont leur raison d'être quelle que soit la structure fondamentale de l'univers matériel, les arts de l'ingénieur prospèrent.

Il est certes concevable que l'esprit, ou l'intelligence, ne puissent jamais être constitués en objets de science. Dans ce cas, les sciences cognitives n'auront connu qu'une existence ephémère, et la question de la contribution que leur apporterait l'IA se révèlera vide de contenu. (Il est vrai qu'alors l'IA pourrait se révéler l'unique discipline durable ayant l'intelligence pour objet — mais elle n'en serait pas la science.) Dans le cas contraire, il est certain que la nature exacte des rapports entre sciences et "ingéniérie" de la cognition dépendra de la nature de la cognition. Mais ces rapports seront ceux, dissymétriques, qui lient une technologie aux sciences fondamentales sur lesquelles elle s'appuie. L'IA, loin de se substituer aux sciences cognitives ou à la philosophie, leur empruntera sans doute davantage qu'elle ne leur apporte. Ce qui n'ôtera rien à son intérêt ni à sa dignité: l'architecture n'est pas diminuée par sa dette à l'égard de la physique, ni celle-ci par ses emprunts aux mathématiques. La distribution des rôles n'est pas un véritable enjeu pour les institutions. Les hommes peuvent choisir leur discipline, ou leur cheminement entre les disciplines. Ce sont eux qui font les découvertes, et les découvertes qui font, refont et parfois même défont les disciplines.

Notes

(1) Si Allen Newell ne pense pas que l'IA soit destinée à devenir "la nouvelle reine des sciences", c'est qu'il estime que nous vivons une époque bien trop démocratique pour cela! (*Cf.* Bobrow & Hayes 1985, p. 379).

(2) *Cf.* Dreyfus 1972/1979 et Dreyfus & Dreyfus 1986.

(3) *Cf.* Searle 1980.

(4) *Cf.* "Much ado about nothing much", in Graubard 1988. Putnam a également donné en 1988 une série de conférences portant sur l'IA; le titre de l'une d'elle est assez éloquent: "AI: A budget of problems".

(5) *Cf.* Rumelhart & McClelland 1986, Smolensky 1988, Graubard 1988.

(6) Après J. Weizenbaum, passé à l'opposition dès les années 1970, T. Winograd et D. Waltz sont les plus connus: cf. Weizenbaum (1976), Flores & Winograd (1986), Waltz (1988).

(7) *Cf.* Andler & Livet 1989.

(8) *Cf.* Haugeland 1981, Dennett 1986, Andler 1986, Andler 1989.

(9) *Cf.* Pinker & Mehler 1988.

(10) La locution "sciences cognitives" est ambigue. Elle désigne tantôt toutes les approches scientifiques de la cognition, dans un but soit d'explication soit de simulation ou de spécification, et subsume dans ce cas l'intelligence artificielle; tantôt les seules disciplines dont l'objet est la cognition naturelle (cas auquel on leur adjoint ou non les neurosciences). (*Cf.* par exemple Andler 1987 ou Andler & Livet 1989.) Nous nous interrogeons dans le présent article

sur les rapports que l'IA entretient avec les (autres) sciences cognitives — psychologie, linguistique, neurosciences — et avec la philosophie (autre discipline dont la place est ici incertaine), sans préjuger la question de savoir si ces rapports sont internes à un vaste ensemble de disciplines, ou bien s'établissent de part et d'autre d'une frontière marquée.

(11) *Cf.* McCulloch 1965/1988, Dupuy 1985.

(12) On se réfère plus volontiers en IA à l'''hypothèse du système symbolique matériel'' de A. Newell et H. Simon: cf. Newell & Simon 1976. Rappelons que selon cette hypothèse, la cognition s'explique comme un *calcul* sur des représentations symboliques effectué par un système — naturel ou artificiel — de traitement de l'information.

(13) *Cf.* Chandrasekaran 1987.

(14) Ce terme prend en IA (et souvent dans les sciences cognitives) une acception particulière: il recouvre toute espèce de représentation interne de faits et de règles auxquels le système de traitement de l'information a recours pour accomplir sa tâche. Les règles grammaticales postulées pas Chomsky sont donc en ce sens des connaissances (Chomsky, en réponse à certaines critiques, dit d'elles que le locuteur les ''*cognize*'', par opposition à ''*know*''. *Cf.* par exemple Chomsky 1976.) Les notions d'accès conscient, d'évaluation et de libre disposition qui sont associées à l'usage courant sont donc souvent absentes ici.

(15) *Cf.* McCarthy & Hayes 1969. Récemment, D. Michie déclarait que c'est dans le domaine des connaissances, plutôt que dans celui de l'intelligence, que les progrès sont le plus nets (''... *Knowledge, rather than intelligence, is where today's action is*''); in Bobrow & Hayes 1985, p. 378.

(16) *Cf.* Gardner 1986/1988, Miller 1979.

(17) Ce projet fait écho à celui d'une ''épistémologie expérimentale'' cher à certains cybernéticiens; mais ceux-ci ne prétendaient pas substituer leur science à la philosophie traditionnelle — ils ne disposaient pas de la ''différence absolue'' que constitue — selon les partisans d'une IA ''reine des sciences'' — l'ordinateur.

(18) C'est d'ailleurs l'une des raisons des nombreux appels à la modestie qui se sont fait entendre dans la profession: l'IA craint d'être plus sévèrement jugée qu'elle ne le mérite. Mais il est aussi de sa responsabilité de redéfinir adéquatement ses objectifs et de reformuler les problèmes auxquels s'appliquent les solutions qu'elle a dégagées, lorsqu'il devient clair qu'elles ne résolvent pas les problèmes initialement posés, et pour lesquels elles avaient été élaborées.

(19) *Loc. cit.* note 4 p. 1.

(20) C'est une expression courante en IA. Pour D. Lenat, les méthodes de l'IA, au sens le plus général, sont ses ''sources d'énergie'' (*sources of power*); *cf.* Lenat 1984.

(21) L'avocat le plus éloquent de cette thèse est Roger Schank.

(22) *Cf.* par exemple Doyle (1984).

(23) A titre de test préliminaire, et certes non concluant, on pourra suivre Gardner (1985/1987) dans le parcours historique et analytique des sciences cognitives qu'il propose.

(24) *Loc. cit.* note 4 p.1. Si je m'appuie sur l'analyse critique de Putnam pour avancer une réponse à la question précise que nous considérons, je ne partage pas pour autant son jugement négatif sur l'entreprise de l'IA dans son ensemble. Daniel Dennett, dans une réponse à Putnam (Dennett 1988, p. 290–292) conteste vigoureusement son jugement quant à la qualité philosophique de l'IA. Selon lui, d'une part il y aurait bel et bien chez les chercheurs de l'IA une conscience aigue des grands problèmes (l'induction, l'analogie, la pertinence, le contexte etc.), et d'autre part on ne saurait reprocher à l'IA de piétiner depuis un quart de siècle face à des obstacles que deux millénaires n'ont pas permis à la philosophie de franchir. Le désaccord n'est pas bien profond: sur le premier point, on évoquera la métaphore favorite de Newell, celle du verre mi-vide mi-plein (les chercheurs sont parfois avertis, parfois inconscients, souvent à mi-chemin); le second est une affaire de présentation: c'est dans la mesure où ils prétendent avoir résolu ou dissout les grands problèmes, ou encore être sur le point de le faire, que les porte-parole de l'IA provoquent l'agacement des philosophes. Sur

le fond de la question, Dennett et Putnam s'accordent pour estimer que les contributions positives de l'IA restent à venir.

(25) *Cf.* par exemple McCarthy 1979.

(26) A l'heure actuelle: on sait en effet que les logiciens sont aujourd'hui à la recherche d'un langage de programmation "transparent", dont la vérification soit triviale, comme l'est celle du langage mathématique formalisé (l'idée est d'ailleurs d'assimiler les *démonstrations* à des *programmes;* les programmes constitueraient ainsi, inversement, la démonstration de leur propre correction). *Cf.* par exemple Parigot 1988. Lorsqu'elle disposera d'un tel langage, l'IA sera en voie de surmonter l'une de ses principales limitations théoriques. Encore faudra-t-il qu'elle puisse préserver la transparence au moment d'intégrer au programme sa dimension cognitive (une fonction "nue" n'est pas en soi l'abstraction d'une capacité cognitive; il faut qu'elle opère sur des représentations) — ce qui pose un problème supplémentaire redoutable. Il y a là néanmoins l'espoir d'un progrès décisif.

(27) Dans ses conférences William James de 1987. *Cf.* Newell (à paraître).

(28) Voir le titre même de Newell & Simon 1976.

(29) La distinction entre science et technologie est particulièrement incertaine dans le cas de l'IA. Quoiqu'elle ait été récemment instaurée par la société savante la plus influente du domaine (l'American Association for Artificial Intelligence), elle ne semble obéir à aucun critère précis. On a plutôt l'impression d'avoir affaire à un continuum. Les raisons de cette situation mériteraient d'être analysées, mais ne le peuvent ici.

(30) *Cf.* Bobrow & Hayes 1985.

(31) *Cf.* Marr 1982, Hurlbert & Poggio 1988.

(32) *Cf.* Marr 1982 (ou un extrait pertinent dans Haugeland 1981, p. 129–142).

(33) *Cf.* par exemple Churchland & Sejnowski 1987, qui critiquent le "rêve de Marr".

(34) *Cf.* Newell 1983.

(35) Nous adoptons ici la perspective la plus favorable à l'IA. Il est certain que l'ouverture et la liberté d'esprit dont il est fait état ne sont pas universelles. L'IA n'est pas exempte de rigidités et de préventions, mais ce n'est pas ce qui nous intéresse ici: il s'agit d'expliquer pourquoi elle joue le rôle particulier qui est le sien, non de souligner qu'elle ne le joue qu'imparfaitement.

(36) En quoi elles se distinguent, ajoute-t-il, de celles du philosophe (*op. cit.* p. 289).

(37) Roger Schank estimait en 1985 que le principal progrès de l'IA était d'avoir pris la mesure de la complexité de la pensée (*cf.* Bobrow & Hayes 1985, p. 383). Jerry Fodor, quant à lui, formulait vers la même époque un diagnostic pessimiste sur les chances d'une science cognitive des processus centraux (Fodor 1983: p. 162–166 de la traduction française).

Bibliographie

Andler, Daniel (1986). *"Studying cognition today"*, rapport à l'European Science Foundation; et Eidos **5**, p. 177–225; avec deux appendices in H. Sinding-Larsen, ed., *Artificial Intelligence and Language*. Oslo: Tano, p. 201–246.

Andler, Daniel (1987). *"Progrès en situation d'incertitude"*, *Le Débat* **47**, p. 213–234.

Andler, Daniel (1989). *"Cognitives (sciences)"*, Paris: Encyclopaedia Universalis, nouv. ed.

Andler, Daniel & Pierre Livet (1989). *"Les sciences cognitives: l'émergence d'une nouvelle constellation"*, *Universalia 1989*. Paris: Encyclopaedia Universalis.

Bobrow, Daniel G. & Patrick J. Hayes, eds (1985). *"Artificial intelligence — Where are we?"*, *Artificial Intelligence* **25**, p. 375–415.

Chandrasekaran, B. (1987). *"What kind of information processing is intelligence? A perspective on AI paradigms and a proposal"*, preprint, à paraître dans D. Partridge & Y. Wilks, eds, *Source Book on the Foundations of AI*. Cambridge: Cambridge University Press.

Chomsky, Noam (1976). *Reflections on Language*. New York: Pantheon Books. Trad. fr. Paris: Maspéro, 1977.

Churchland, Patricia Smith & Terrence J. Sejnowski (1989). *"Neural representation and neural computation"*, in L. Nadel, ed. *Neural Connections, Mental Computation.* Cambridge, Mass.: MIT Press.

Dennett, Daniel C. (1988). *"When philosophers encounter artificial intelligence"*, in Graubard 1988, p. 283–295.

Doyle, Jon (1984). *"Expert systems without computers, or Theory and trust in Artificial Intelligence"*, *AI Magazine* (Summer 1984), p. 59–63.

Dreyfus, Hubert L. & Stuart E. Dreyfus (1986). *Mind Over Machine*. New York: Macmillan.

Dreyfus, Hubert L. (1972/1979). *What Computers Can't Do*. New York: Harper & Row. Trad. fr. *Intelligence artificielle: mythes et limites,* Paris: Flammarion, 1984.

Dupuy, Jean–Pierre (1985). *"L'essor de la première cybernétique (1943–1953)"*, *Cahiers du* CREA **7** (Ecole polytechnique, Paris).

Flores, Fernando & Terry Winograd (1986). *Understanding Computers and Cognition: A New Foundation for Design.* Hillsdale, NJ: Ablex.

Fodor, Jerry A. (1983).*The Modularity of Mind.* Cambridge, Mass.: MIT Press. Trad. fr. *La Modularité de l'esprit,* Paris: Minuit, 1986.

Gardner, Howard (1985/1987). *The Mind's New Science: A History of the Cognitive Revolution.* New York: Basic Books.

Graubard, Stephen R. (1988), ed. *The Artificial Intelligence Debate. False Starts, Real Foundation.* Cambridge, Mass.: MIT Press, et numéro spécial de Daedalus **117** 1 (Winter 1988).

Haugeland, John (1981), ed. *Mind Design*. Cambridge, Mass.: MIT Press.

Hurlbert, Anya & Tomaso Poggio (1988). *"Making machines (and artificial intelligence) see"*, in Graubard (1988).

Lenat, Douglas B. (1984). *"Computer software for intelligent systems"*, *Scientific American* **251** 3 (Sept. 1984), p. 152–160.

Marr, David (1982). *Vision*. San Francisco: Freeman.

McCarthy, John & Patrick Hayes (1969). *"Some philosophical problems from the standpoint of artificial intelligence"*, in B. Meltzer & D. Michie, eds, *Machine Intelligence* **4**. Edinburgh: Edinburgh University Press.

McCarthy, Thomas (1979). *"Programs as theories"*, in M. Ringle, ed., *Philosophical Perspectives in Artificial Intelligence*. Atlantic Highlands, NJ: Humanities Press.

McCulloch, Warren S. (1965/1988). *Embodiments of Mind*. Cambridge, Mass.: MIT Press.

Newell, Allen & Herbert Simon (1976). *"Computer science as empirical enquiry: symbols and search"*, *Communications of the Association for Computing Machinery* **19**, p. 113–126. Repris dans Haugeland 1981.

Newell, Allan (1983). *"Intellectual issues in the history of artificial intelligence"*, in F. Machlup & U. Mansfield, eds, *The Study of Information: Interdisciplinary Messages*. New York: Wiley.

Newell, Allan (à paraître). *Unified theories of cognition.*

Parigot, Michel (1988). *"Preuves et programmes: les mathématiques comme langage de programmation"*, *Le Courrier du CNRS*, supplément au nº 69 "Images des mathématiques", p. 40–47.

Pinker, Steven & Jacques Mehler (1988). *Connections and Symbols.* Cambridge, Mass.: MIT Press.

Rumelhart, David E. & James L. McClelland (1986), eds. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, Mass.: MIT Press.

Searle, John R. (1980). *"Minds, brains, and programs"*, *The Behavioral and Brain Sciences* **3**, p. 417–457. Reproduit dans Haugeland (1981).

Smolensky, Paul (1988). *"On the proper treatment of connectionism"*, *The Behavioral and Brain Sciences* **11**, p. 1–74.

Waltz, David L. (1988). *"The prospects for building truly intelligent machines"*, in Graubard (1988).

Weizenbaum, Joseph (1976). *Computer Power and Human Reason.* San Francisco: Freeman.

Avron Polakow

Philosophy Dept., Ben Gurion Univ. of the Negev,
Beer Sheva (Israel)

ASS ("Artificial Stupidity S.")

Abstract

The aim of this paper is to argue that there is no issue of intelligence in "AI". The issues about "artificial intelligence" refers to ontology (what a computer is), not to epistemology (what a computer can know) and the question of whether a computer is intelligent goes against some of our fundamental beliefs about what exists, what the world is, and what range of meanings we attribute to the term "exist".

Resumo

O propósito deste estudo é sustentar que não há qualquer perspectiva de inteligência na "IA". As perspectivas da "inteligência artificial" referem-se à ontologia (o que é o computador), não à epistemologia (o que pode conhecer o computador), e questionar se o computador é inteligente vai contra convicções fundamentais sobre o que existe, sobre o que é o mundo e sobre o elenco de significações que atribuímos ao termo "existir".

I - An End: To the Rational Animal?

> It is so easy not to be of age. If I have a book which understands for me, a pastor who has a conscience for me, a physician who decides my diet, and so forth, I need not trouble myself. I need not think, if I can only pay — others will readily undertake the irksome work for me.(1)
>
> But I hear shouting on all sides, "Do not reason!" The officer says: "Do not reason but drill!" The tax collector: "Do not reason but pay!" The cleric: "Do not reason but believe!" Only one prince in the world says, "Use your reason as much as you will, and about what you will, but obey!"(2)
>
> Immanuel Kant

Given their *intelligence* are computers intelligent, or instead, are they *stupid*? Is artificial intelligence really a species of intelligence, or is it merely a species of unintelligence?

My aim here is to show that there is no issue of intelligence in "artificial intelligence." The issues about artificial intelligence boil down to ontology, not epistemology: they are not about what a computer can *know*, but about what a computer *is*. I shall indicate that the question of whether a computer is intelligent goes against some of our fundamental beliefs about what exists, what the world is, and what range of meanings we attribute to the term "exist".

This view is contrasted with the one maintained by Turing, that of criterialism, which claims that there is *no criterion* which could be used to distinguish my behavior from that of a computer's. So everything (thinking in particular) which can be attributed to me can be attributed to a computer as well.

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 112/123
1992, APDC & SPF, Lisboa

Oddly enough, despite the impression that criterialism is an argument about concrete computers, it is actually an argument about *abstract* machines; not about performance, but about competence. We are not to seek intelligence in the performance of this or that company's latest model (despite what the ads. claim!). We shall pursue this matter presently.

Doubtless the question of whether computers are intelligent would not have arisen were it not for the very impressive achievements of modern computers in executing human tasks, the *invisible* human tasks. The man-in-the-street has more respect for their ability than the expert, because in ways unbeknown to him, computers do those oh-so-human things; they play better than average chess, they spell perfectly, they calculate far better than us, they prove odd logical theorems of which few of us have any knowledge, and the dialect of these *dialectical machines* is now no further from our own language than that of some foreigner who speaks English. Their expert advice in some fields exceeds that of our own human experts. If this is what they are *now* capable of, what more is necessary to convince us of their intelligence? They do all this, without consciousness, without oedipal complexes, without Super-egos, Ids, repression, defence-mechanisms and what have you, although these can always be added if missed/necessary.

These abilities directly overthrow basic ontological preconceptions, and if taken seriously, are tantamount not to a mere reform of ontology, but to a literal revolution. The Aristotlean *differentia* of the human species from that of other animals *qua* rational, is wholly undermined. Descartes' extension of Aristotle's views is even more threatened: since he equated the body with the mechanical, a thoughtless, extended machine. Thought, for him, needs no mental mechanic; it is governed by the *unmechanical.*

Computers, putatively, undermine the Aristotlean-Cartesian conception, for now we seem to be distinguished within the *species* of *rationalia*, by the *differentia animalia*. We have lost our ontological loneliness with the membership of the new *rationalia*. This is the new ontological revolution — a complete categorical exchange: *differentia* and species have swapped roles. Natural kinds are extinguished by *new artificial kinds*, the family of *rationalia*. The natural essences will be obliterated forever with the birth of these new nominal essences. We as the subject under discussion have switched ontologocial allegiances from the *animalia* to the *rationalia.*

An impulsive but over-reactive response to this revolution might be to retreat into irrationality, to the person of emotion, with his anxiety, hopes, pleasures, — or, in short, to the *human* being. Despite its appeal, this view is insufficient to counter the challenge of technology. In fact it opens up a new flank for attack: instead of artificial intelligence, it will be called on to deal with *artificial humaneness.*

I intend to defend the *conservative* view without any repeal to humaneness. Perhaps this might give some weight to the "other" debate, the moral one.

II. Is Thought Intelligence?

The words "thought" and "intelligence" are often used interchangeably with respect to computers. Computers, supposedly, both think and have intelligence. Perhaps the thought is artificial but the intelligence is real.

One way to pinpoint the locus of issues between intelligence and thought is to ask: Does intelligence require — as a *necessary* condition — thought? In other words, does thought in any way *determine* intelligence; i.e. where there is no thought there is no intelligence.

"Thought" and its cognates is context-rich, bending to different connotations in different contexts. We think *about* different things; we *have* different thoughts; we *sink* into our thoughts.

"Intelligence" shows much paucity of context. We *don't* intelligent, and certainly *not* about something. *No* one has *sunk* into intelligence, and our *thoughts* have no parallel, *intelligibilia*(3), objects of intelligence, which only require that we *perceive* them so as to be intelligent. Intelligence is an ability and as such differs from thought. Some of us have more of it and some of us less. It comes in degrees. It is not easy to be intelligent. How easy it is — in contrast — to think. We simply cannot avoid it. We are constantly bombarded by our thoughts. In some senses of the term "thought", no ability accompanies its reference. It is universal, and no prizes are awarded for exhibiting it.

There is no simple transition from thought to intelligence. No *set* of thoughts, as large as we choose, nor the possession of such thoughts contributes as such to intelligence. The web of belief in which our thoughts are embedded does not grant us any ability. No *particular* thought is in itself the expression of any ability, any more than any other particular thought.

The more we think, the more we alter our thoughts; it is not *that* thinking nor *that* attention which will make us intelligent. Intelligence is a very special *achievement*, an achievement with a singular *side-benefit*, that of *problem-solving*. It would be absurd to claim that a particular solution to a problem is *intelligent but incorrect*. Intelligence is a commitment. Thought, in contrast, is not binding. We take it, *as it comes*, successful or not, as a solution to a problem, or not as a solution to a problem, or even as a mistaken solution to a problem. It is generally not about problems at all.

Perhaps the *necessity* of thought for intelligence, — the view that thought does not determine intelligence, — lies in the history of the terms "thought" and "intelligence". "Intelligence" is a new term, made current by the use of IQ-tests. It supplanted the more classical term "rationality." We don't speak of "artificial rationality." The founding fathers of IQ-tests did not think of quantifying rationality. It would have been odd had Descartes claimed: I think 75.3% therefore I exist 75.3%; and even odder if this were phrased in the third person. ("EQ-tests" viz. "Existent Quotient tests").

IQ-tests made a detour around processes, dealing directly with *results*. They treated only the taxonomy of achievements; and did not try to go from the activity of thought to our overt acts. They preferred external descriptions to explanation, as if the results were *impromptu*. The success of this aproach is more consumer-related than supply-related. IQ-tests have been used as indicators of further achievements.

In a way computers have turned back the clock, simply because they do *not* have any *impromptu* results. Much labour is unlost in searching for intelligent means to complex problems. All the energy is invested in the *process* towards the achievement. In particular, computers renew interest in the relation between process and result.

I wish to realign this relation so as to reach a new reflective equilibrium between thought and intelligence. Before this, though we must still consider whether intelligence is necessary for thought. No conceptual analysis will settle the necessary conditions between intelligence and thought. At most we must seek some goodness of fit between these concepts.

For some time now computers have been exhibiting "higher" human activities, but the uniqueness and impressiveness of such activities lies in the field of non-mathematical, non-numeric activities. There are the "soft" activities, the "invisible" ones, like our thought. As the computer "fundis" have it: the number-crunching era is over. Now it is

the thought–munching era.

Number–crunching leaves us epistemologically cold (although our Greek predecessors would have been extra–ordinarily impressed by this!). But thought–munching is a different question altogether. Here computers are trespassing on the core of our epistemological property. They are kneading the same epistemological dough as we do, without a whit of knowledge thereof. Is thought–crunching the same thing as intelligence? I would like to put aside two common objections against thoughtless intelligence:

Firstly there is the *trip–wire* objection, the objection that there will always be *some* human activities displayed by our intelligent behavior which computers will not be able to do. There are specific human abilities which are simply beyond the reaches of the computer. Here it is *conceded* that some of the activities of the computer are intelligent, but that there is a *limit* to what computers can do. This is a type of racism against computers: they are *inherently* inferior. Frankenstein stories are like the typical racist myths — they instil fear while inspiring superiority. But the fear is always *individual* rather than collective fear. *They* are never a threat to our *"race"* (the "intelligent" or the "white," etc.).

This, though, is too optimistic. It depends on an *a priori* conviction that what we can do but computers cannot as yet do, they will never be able to do. This view has been refuted many times in the past and there is no reason to assume that it will not be refuted in the future as well. To claim otherwise one must have a theory of computers which would exclude certain activities. But no such theory exists. The trip–wire objection pits the past against the future, i.e., despite past achievements there are things in the future which computers will never be able to do. The past is always sacrosanct, the future never. Computers have a *negative disposition*, a certain *inability* to perform some human activity. But how could any theory give an *a priori* reason for this? Nothing in the abstract account of machines gives grounds for this view.

Secondly, we have the *whip–lash* objection which emphasizes the *present*: all activities of the computer are *granted*. It is merely denied that any of these are intelligent, on the grounds that the computer does *not* give as a *result* what we give. When a computer prints out "The sum of the angles of a triangle is 180°" it is supposed to be doing something entirely different from what we do when we write "The sum of the angles of a triangle is 180°". Dretske emphasizes this point:

> For machines, even the best of them, don't have an IQ. They don't *do* what we do — at least none of the things that, when we do them, exhibit intelligence. And it's not just that they don't do them the way we do them or as well as we do them. They don't do them at all. They don't solve problems, play games, prove theorems, recognize patterns, let alone think, see and remember. They don't even add and subtract.[4]

So in what, then, does the difference between ourselves and the computers lie? In achievements? Certainly not in the fact that the computer prints whereas we write. Were that the reason we would never even use a simple abacus, on the grounds that moving the beads cannot comprise an arithmetic operation.

Dretske[5] compares the relationship between ourselves and the computer to that of a tool to its purpose, much the same as our own relationship to a key. We open doors by means of keys, but keys do not–except in a figure of speech — open doors. We know this fact about keys.

Yet this analogy of Dretske's is incorrect because we cannot *discover* that a computer is a *tool*. Do you know, by mere examination of this text, whether it is the work of a man or of a computer? Keys always *tell* us their tale. The computer, however, is not independentally individuated like a key. Is this, here and now, an untale being told, or an

untold tale? The instrumentalist analogy does not suffice, mainly, I suspect, because computers are only tools in an *invisible* way, by exemplifying some or other *abstract* machine.

They don't have a function without a program, like toasters which toast whenever they are switched on. Like us, at turn-on-time, they are *tabulae ragae.*

The whip-lash objection, unlike the trip-wire objection, does not instil any fear of the computer's achievements — admitted as such. It is the fear that computers actually can and do *behave* (and don't merely exhibit activity) which threatens us, as if they were about to burst onto the scene of nature, the human biological scene. Whip-lashists wish to retain the view that they achieve results without behavior. *Results are not a sort of behavior.*

This may sound surprising, as problem-solving in the human domain appears to be a case of behavior *par excellence*. What precisely is the role of behavior in problem-solving? In solving a problem it is clear that the *medium* in which the solution is expressed is irrelevant to its correctness. It is the result which counts, not the gesture by which it is expressed. Neither my style of writing, nor that of the computer is considered in evaluating a result. We could coin a neology: *Results are alienated from their means of expression.* A spit in the eye could express the theory of relativity as could Beethoven's fifth symphony. Were spits and symphonies our natural language we would have no trouble understanding them. The whip-lash objection denies this not by racism, but by sadism: if we cannot win the race against the computers, then beat them down. Unfortunately the results cannot be beaten away. They require an explanation.

Results cannot be argued away as *nothing*, a sort of *unbehavior*. Results are not miracles, appearing *ex nihilo*. For human beings, results are *premeditated outcomes*. The premeditation could be a mere spontaneous passing whim, but without it the results cannot be understood. A murderer's escaping from a prison cell is not a *mere* event but an outcome of some prior action on the murderer's part, whether planned or contingent. When I move a chess piece, this is a result of a prior strategy, whether sophisticated or superficial. It is not a mere pushing of the piece. Pushing a piece by mistake only counts as a move in the game if incorporated (even retroactively) into an intention. Strategies first, results later! That is why results are outcomes.

What are the results of a computer supposed to be outcomes of? In the case of humans we attribute results to *thoughts*. Results, intelligent or not, are assumed to be products of thought. There are no random results, even if we treat some results as intuitive or insightful. Results don't appear as an *unpart* of us, like bolts out of the blue. Mystics may think such, but then mysticism is not problem-solving behavior (more likely it is problem-*creating* behavior).

Functionalism is the one theory which claims that results are *not* outcomes. A functionalist account of intelligence deals with some property of *abstract* machines which can be realized in an infinite variety of concrete machines. An abstract machine describes the guidelines of solving a problem in a computational manner. However, no concrete machine is essential to the abstract machine. Since we human and electronic machines are clearly made of different materials, it is quite clear that our physical equipment in solving problems differs, even if our ways of solving a problem might be identical.

Functionalism deals with the *manner* of a solution and not with its means. Were intelligence in computers a question of means then intelligence would underdetermine its means. There are an indefinite number of ways to reach intelligent results. The moral of this story is that intelligence is not to be sought in the physical, concrete computer.

Functionalism defines "results" by way of an abstract machine, say a Turing machine

(the archetype of all possible computers) that unrealized unthinking model. All transitions of the abstract machine given by the machine-table, determine which results the machine will reach. Physical realization of the abstract machine table is irrelevant. If this is true for intelligence in general then functionalism holds that:

> Intelligence is no more the outcome of thought than it is the outcome of a concrete computer. Intelligence is well-defined by the machine-table of an Abstract Turing Machine.

How are we to interpret such a view? The machine-table itself — even if supplemented by different cases of input and output as they are run through — cannot be intelligent-in-itself. A machine table *per se* is no more nor less intelligent than the numbers etched on a coin would be valuable. Neither the set of machine states, nor any particular machine state, can be called "intelligent". No particular machine state is any more intelligent than any other. It is a category-mistake to say "intelligent machine state."

Perhaps we are supposed to attribute the intelligence not to states but to the transitions between them. The elegance of transitions in solving a problem is what attracts our attention. The breakdown of a problem into minute steps and its step-after-step solution, impresses us as intelligent.

Despite impressions, this is not intelligence. *Firstly*, all programs are step-by-step transitions, and some such programs are *meaningless*. The transitions, clever and complex as they might be, are not guarantees of anything. *Secondly*, and more importantly, even if we were to agree for a moment that certain transitions were "intelligent", *those very transitions can be remapped onto interpretations which are wholly unintelligent*. There is nothing in a machine-table which determines its interpretation.

Machine-tables are no different from musical scores. The same score can be interpreted as a melody, as a counter-melody, as any sequence of notes given suitable transpositions, or even as secret correspondence between intelligence agents who had adopted musical notation as their secret code. There is nothing inherently musical about a score, and there is nothing inherently intelligent about a machine-table.

Intelligence is radically underdetermined by machine-tables, in just the same way as reference and translation are underdetermined by language.

Could we treat machine-states as beliefs and treat intelligence as if it were a result of our beliefs? Instead of intelligence let's return a moment to the more restricted topic of problem-solving. How is it that machines solve problems? Problem-solving behavior can be mimicked by transition function between one state and another in the machine-table. It is some data, input or output, or some random action, which pushes a machine from state to state.

Belief does not operate this way. Human beliefs are not defined as discrete entities. It is not at all clear what the boundaries of any particular belief are. The more liberal-minded obscure boundaries; the more dogmatic-minded try to bring them into sharp relief. Right now the important question for us lies in the transition from one belief to another. It is important to note that beliefs *don't have the function* of solving problems as such. Their use — if at all — in problem-solving, is more the exception than the rule. It would be absurd if, to the question, "How are you solving the problem...?" we were to answer, "From one belief to the next", just as machines solve problems from one state to the next. Beliefs are not machine-states, for at least the reason that there are no transition functions between them. There is no data, input or output, which, together with a given appropriate and correct belief, would lead us on to another belief by some function.

Some light can be shed, in this very special context, on why beliefs are *opaque*

whereas machine-states are necessarily *transparent*. Machine transitions must follow transition *functions* (even in the case of nondeterministic automata, where "functions" are replaced by "relations"). The machine-table is *all* that the machine is. The machine-states are well-ordered. It is an abuse of the term "state" to include in its extension both natal states and machine states. Beliefs cannot be determinate states, if only because they are not individuated in time. The criterion for transparency is substitution *salva veritate*. The transition functions of a machine guarantee that machine-states are substitutible *salva veritate*.

How do humans solve problems? How do we go from belief to belief? Could it be by transition functions? Without *believing in a function* how could we get out of a current belief? But if our transitions are beliefs, then *we move from belief to belief by belief*!! In fact, we give very partial accounts of how we solve problems. Whatever the *process* of how this takes place, this process is *not one of our beliefs. However, if we don't have a belief about the correctness* of a move from one belief to another, then such a move will be invalid. Yet, believe as much as we will, no belief pushes us in our problem-solving activity. Later we will see that belief is not a competence whereas problem-solving is. Even if we are sypplied with the computer program to solve a problem, we would not, by using it, have acquired a solution to a problem. Explicit instructions for solving problems do not count, for us, as the solutions to the problem.

Sceptics might intercede here and claim that there is no intelligence required in problem-solving. All solutions are equal, whether intelligent or unintelligent.

Alternatively, some might seek refuge in rationality: Intelligence is rationality, a form of complex order. Descartes thought anyone could solve any problem, given correct beliefs and sufficient training. Mathematical problems could be solved by anyone who thought rationally. Rationality was juxtaposed to primitiveness. Yet rationality is *justification*, intelligence is *not*. But it is not the justificationary aspect of rationality, that everything can be given a *reason*, which leads us to how problems have solutions. We are quite rational even without having a whit of problem-solving ability.

Yet intelligence is not *impromptu*. We need to invest many human resources, — even though it is not clear which — in order to become/remain intelligent. But even given an intelligent result, the means to its achievement is not necessarily intelligent. Imagine for a moment, *a present-day prophet* who has devoted his life to probing into the secret meanings of telephone books, and who, lo and behold, discovers a recurring number in all telephone books standing for the letters j, e, h, o, v, a, h. This *telephone-book Cabalist* would not be considered intelligent, however sophisticated the mathematical equipment he employed.

A minimal requirement for intelligent activity is that the *result* be considered intelligent. Intelligence is a purpose-concept. Without a particular attitude to the results we will not attribute intelligence to the activities which bring about such results. In this sense intelligence is more akin to thinking rather than to problem-solving. There are priorities as far as goals are concerned, some rated as intelligent, some not.

This is a key problem for all advocates of artificial intelligence: AI unlinks the process from the result. It is neutral as to whether the result is intelligent or not. No amount of rule manipulation, machine-states, complexity, order, functions, or whatever, can make the unintelligent intelligent. Of course, given an intelligent result much research is invested in how to reach it.

This is the excitement department of AI. But given the rules, *qua* rules, one can reinterpret them *stupidly* just as much as one can interpret them *intelligently*. Here's a brief

thought–experiment to show how:

> While in operation my computer–printer fouls up and begins tearing up the paper while continuing to print. So the printer, with its sort of phantom–limb printer, continues merrily on while I drink coffee in the canteen. On my return, I collect the torn bits of paper and wonder: "Did or did not my printer print my article?" There seems to be a way to check this. If I can put together the puzzle of bits of paper then I will tend to claim that it had. Were this impossible, I would tend to claim that it hadn't. That being the case, the result of the thought–experiment is that the word "print" has its meaning determined by what interpretation I can put on the pieces of paper. If deciphering is a criterion of intelligence, then processes are only of derived importance as to whether something is intelligent.

III. Do Computers Think?

What is it that computers are supposed to be thinking about? About the world order, about the nature of substances, about the theater of the absurd, about power, about the indeterminacy of translations, about general relativity, about Henry Moore and Francis Bacon, about retributive or utilitarian theories of punishment, about contrapunctual music? Perhaps they are supposed to be thinking without thinking *about* anything. They have no ideas, opinions or beliefs. They are *Erkentnissfrei*. They are analytical, not epistemological, machines.

In fact it is quite ridiculous to even ask what they are thinking *about*. They are not considering thought at all. To ask whether *computers* think is a totally different question from asking whether *we* think. You might say that they have the *knack of thought*, whereas we have the *key to thought*. Quite often we have no way of describing how we think, but they have no knowledge *about* what they are thinking.

Their thinking is slavishly devoted to problem–solving. They have no excuse to think without a problem to solve. They think when they have the *ability* to solve a problem. Solutions are embedded in the systematic connections of the program and the data–base. This combination is all that there could be in the computer solution of a problem.

Holding an opinion or thought is not an ability. We think even when *not* solving problems. Thoughts are embedded in the web of belief, in theories and/or in language–games. We may pursue the consequences of our thoughts within the web, if our knowledge is sufficiently explicit. For human beings, thought is at least the *acceptance* of a belief–system.

Following up thoughts in the web of belief is not a problem–solving ability. Thought is not primarily an ability. Even holding onto a web of belief is not an ability. The transition from belief to belief need not be "correct" in order to count as thought. It is the overall picture which is more important. How the beliefs accumulate together. No training improves our "ability" to believe, because beliefs have no need of training. We don't measure thoughts by goodness of their fit into the web of belief. Were thinking the ability to follow up thoughts within the web of belief it is doubtful if we could ever think. It would be like a person trying to understand the speech of someone by first giving a grammatical analysis of that speech. It would bog us down immediately.

It is the *use* of thought which requires ability. Computers, unlike us, cannot *possess* any specific thoughts. As far as computers are concerned, *unused* thoughts are *unpossessed* thoughts. Their possession is in their use. The mere registration in a data–base of some fact is not a thought.

What, then, constitutes a *problem* and how is a problem to be *solved*? It is difficult to

define these. True enough problems are expressed in language, and as the language is more scientifically sophisticated, so the problems are expressed more explicitly. Nonetheless many problems have turned out to be pseudo-problems even though they were well-formulated within formal systems. In contrast, many apparently problem-free, clear situations are in fact problem-ridden. Even given a problem, there is no *a priori* method that tells us which means we are to use for its solution. There is no method for defining problems.

Human thought is not specifically oriented towards problem-solving. This is a use of our thought. Are all our activities — not only thinking — transposable into problem-solving activities? Can feeling be transposed into the problem of *how* we feel? What are we to make of pleasure as a problem-solving activity? A reductionist AI program would have us believe that this is possible.

One of the successful cases of reductions is that of language-understanding. Few of us can describe our understanding of the speech of others as a problem-solving activity. We don't have explicit knowledge of the rules involved. Our understanding is immediate. Nonetheless, in many cases there are rule-bound analyses which, given speech production, yield understanding of the speech productions. Such analyses *solve* the problem of understanding a given speech *production*. The language-users do not conceive speech understanding as a problem-solving activity. This move has forced us into reconceiving conceptual divisions.

IV. Ontology

It might be asked:

Why not acknowledge that, if computers solve problems, then at least in *so* doing they are thinking? As I claimed at the outset of this article, I believe the answer to this question is one of ontology, and does not really bear on computers. I shall now proceed to justify this claim. Turing's[6] famous criterion was supposed to solve this problem. He saw behavior as the final arbitrator, so that humans had no additional ability beyond any other thinkers.

One criticism against this view was that it didn't distinguish identity from mimickry. An adroit parrot could be called intelligent. There is a deeper criticism, though, against criterialism. Imagine a computer, with very advanced software and programs, designed to produce rubbish-heaps. By the Turing-test such a computer doesn't think, because only its output counts, when deciding if it thinks. All depends on our human judgement. An errant genius may one day collect all the rubbish, claim that it is directly related to the program, and that in fact what looks like trash is actually a solution to how Quantum Theory can be reconciled with Relativity Theory. Taken on its own, the output appears to be rubbish. Taken *as a whole*, output together with the program, it can be interpreted successfully. Our observational data is theory-dependent: the output is program-dependent for *its meaning*.

Two ontological theories have been offered to rebut the view that computers think *intentionalism* and *realism*.

Searle[7] advocates intentionalism. His Chinese Box experiment, where Chinese tokens are successfully traded by following certain rules in the box and without any knowledge whatsoever of Chinese, is meant to show that intentionality is the missing link in understanding Chinese. Without intending the meaning of the symbol, Chinese cannot be understood.

However, this theory is ambiguous. We might well attribute intentionality to computers, if the use of the term is derived from third-person uses. We attribute intentionality not to explain a past action but to *predict* a future one. If prediction fails we deny intentionality, or claim counter-intentionality. Fulfillment of predicition is the criteria of intentionality, not the converse. If intention is our best predictor, why deny it in the case of computers? Dennet[8] argues against Searle that intentionality is an attitude, or what he has called a "stance", towards behavior. He rejects Searle's original intentionality. The ambiguity of whether intentionality is original or just a stance, is not solved by Searle's Chinese Box experiment.

Searle's *original* intentionality, supposedly, is derived from the *first-person* use of intentionality. But this leads to an undermining of his experiment. According to the first-person account of intentionality, mental acts are bi-polar, the subjective pole I, and the objective pole, the content of my mental act. I understand Chinese if I *intend* the meanings of the Chinese symbols to *myself*. I'm not just playing around with Chinese poker tokens, I am asking (and answering) what the tokens mean. I'm not just a token-pusher. Searle, though, is in error here. If I don't push my tokens correctly, however much I intend meanings for the tokens, I will fail to understand Chinese. It *is* the symbol-manipulation which counts, and if successful, no extras are needed.

Dretske advocates a different ontological tack, *realism*. It is not the subject-symbol relation which indicates that we think and computers do not. It is the relation between symbols and the world which counts. It's not the symbol-manipulation which is thought, but the understanding that symbols are about the world. Without realism, there is no thought.

> What we have so far required of any aspiring symbol-manipulator is, in effect, that *some* of its symbols be actual signs of the conditions they signify, that there *be* some system-to-world correlations that confer on these symbols an intrinsic meaning, a meaning they do not derive wholly from the purposes and intentions of their users. This puts the symbol-manipulator *in the world* in a way it would not otherwise be.[9]

This is in contrast to Searle's intentionality. Thought which is not ultimately about something external to itself is not thought. Metaphysics is the ground for thought. No Realism, no Thought! I suppose there is at least one realist proposition excluded from the computer, the proposition 'I am a computer.' A computer cannot be realist because it could not know that it was a computer. This is precluded because computers cannot determine the reference of the term 'computer'. For the computers this is a mere symbol, and not a sign. It might be open to rule-manipulation but it is not open to a reference-relation.

Dretske even withdraws the use of the terms "symbol" and "information" in regard to computers. "Information" is an epistemic term, something beyond the reach of computers.

Dretske is a realist, not an indeterminist, and is of the opinion that without realism, we would fall into the trap of acknowledging that computers think. Indeterminism, like Putnam's latest variety,[10] leads, in Dretske's view, to the conclusion that computers think. Because reference is indeterminate, what remains is sheer syntactical manipulation, something by which I will not differ from the computer. For the Indeterminist the computer is just another philosopher, just another theory-holder like myself.

V. Knowledge and Cartesian Engines

Here's another metaphysical set of puzzles. Computers are not idle abstractions. They

are physically constructed even if their programs and solutions can be set down on paper. But why do abstract machines have to be made concrete? If we have an abstract machine to deal with a problem, why do we need to construct a concrete one? Turing's description of ideal abstract computers antedated their actual construction by a few years. Here is an enigmatic presentation of why concrete machines are necessary. Imagine we are asked to solve the following computer-problem:

Construct an abstract machine to solve the Cartesian
Problem that my existence follows from my thought.

Now, some may offer the following solution:

This is a machine-table.
Machine-tables solve problems.
I solve a problem.
ergo, I exist.

This solution is unacceptable, but not because of the indexical "I", which is purely figurative here. 'This' would be a sufficient replacement, if we preferred less poetic language. The solution is unacceptabe because it is self-refuting. Were the solution correct, then the machine-table would have to include conditions of strings which are accepted/rejected by it. It is an insufficient condition to qualify as a machine-table that in the table it is written "this is a machine-table".

Would a concrete machine do any better? Do we need a Cartesian Engine to draw our coaches into existence? Imagine if we realized the abstract machine as a Cartesian Engine in the following manner:

This is a realized program.
Realized programs solve problems.
I solve the Cartesian problem.
ergo, I exist.

This is equally sterile. *No physical process can report that it is solving a problem.* Reporting is not a physical quality. Yet the concrete computer has no non-physical, abstract, residue once made concrete. Computers may spill ink which we take as the signs "I solve a problem." But ink-spilling is no more and no less than ink-spilling. That does not in itself comprise evidence for problem-solving. Truth and falsity of propositions are beyond the computer's ability. In addition, no problem-solving behavior could even decide "I am now solving a problem."

Searle[11] thinks that intentionality is biological, and computers, not having organic properties, cannot have intention. But the paradox of the Cartesian Engine also undermines Searle's theory. If knowledge that something is a biological process is encoded into those biological processes themselves, we would have a biological version of the Cartesian Engine. It is fallacious to claim that we *know* we are biological organisms because we are biological organisms. Biology is not a case of Searlean introspection.

Why the need to *realize* abstract machines? Machine-tables include everything necessary for the moves in a concrete computer. A series of *passes* written down on paper alongside an abstract machine-table give all that the concrete computer will give us. We can spatialize time of the real computer by writing down its execution passes on a page of paper. Paper is a purely spatial, *static*, media. The inefficiency of this, our idleness or incompetence to do this, may be the only reason we have to hand things over to concrete computers.

So neither as abstract or as concrete machine can computers prove their existence. They cannot even deduce that they are problem-solvers. So how are abstract machine-tables to be realized? No machine-table has any property that it needs concrete realization.

Computers present a very unique relation between the technological theory and the machine built according to it. All other machines have their purpose built into their *concrete constructions*. No abstract designs for washing machines have been known to wash laundry. Only *actual* washing machines do that. In contrast to washing machines, computers already have their function given by their abstract design. They *operate in the abstract*! In total contrast to washing machines, the actual purpose of a computer, even in the case of a specific problem, is understood only by interpreting the computer program. A computer which multiplies without a screen or printer cannot communicate its activities to us, but by knowing how it is programmed we can state what it is doing. Of course our best interpretation of its abstract design is to add a concrete screen or computer. But this doesn't change the problem it was programmed to solve.

How, it may be asked, do we manage to actualize a theoretical, abstract description of a computer in a concrete construction? Two conditions are required:

(1) Setting up a correspondence between the *machine-states* of the abstract computer and physical *events*.

(2) Setting up a correspondence between the *transition-functions* in a machine-table and physical *processes*.

Of course, technologically speaking, we must be able to construct and control such events and processes. Whatever is necessary, no mere inspection and description of concrete physical equipment is sufficient to disclose what the function of a computer is. Of course, there are other necessary conditions to construct computers.

Concretization of abstract computers leads neither to existence, to thought, nor to intelligence. But equally, machine-tables are not intelligent, nor are their realizations. 'Thinking' is ambiguous between 'performance' and 'knowledge'. Performance is not knowledge, and knowledge is not performance.

References

(1) I. Kant: *An answer to question: "What is enlightenment?"*, in *Kant on History*, pp. 3–5. The translations have been amended slightly. "Beantwortung der Frage: Was ist Aufklarung", A481–2, *5. Dezemb. 1783*. Kant Werke in Zwolf Bande, Theorie–Werkausgabe, Suhrkamp Verlag, Frankfurt am Main, 1964, Seite 53–61.

(2) *Ibid.*, A485.

(3) The term "intelligibilia", but not the concept, is borrowed from "The Inaugural Dissertation: On the form and principles of the sensible and intelligible world" (1770), in *Kant's Precritical Writings*, OUP, Oxford, 1968.

(4) F. Dretske: *"Machines and the Mental"*, Proceedings and Addresses of the American Philosophical Association, September 1987, 59:1, p. 24.

(5) *Ibid.*, p. 25.

(6) A. Turing: *"Computing Machinery and Intelligence"*, in *The Mind's Eye*, eds. D.R. Hofstadter and D.C. Dennett, Penguin 1981.

(7) J.R. Searle: *"Minds, Brains and Programs"*, in *The Mind's Eye*, eds. D.R. Hofstadter and D.C. Dennett, Penguin 1981, pp. 353–373.

(8) D.C. Dennett: *Brainstorms*, Ch. 2, "Intentional Systems" pp. 3–22, The Harvester Press, 1981.

(9) Dretske, *op.cit.*, p. 29.

(10) H. Putnam: *Language, Truth and History*, ch. 1, "Brains in a Vat", Cambridge University Press, 1981.

(11) Searle, *op. cit.*, p. 372.

2

LÓGICA & RACIONALIDADE

LOGIC & RATIONALITY

Newton C.A. da Costa

Dep. Filosofia, Univ. São Paulo
São Paulo, SP (Brasil)

Lógica e Inteligência Artificial

Resumo

O autor analisa os contributos da IA para a melhor compreensão do actual estado de desenvolvimento da Lógica. São consideradas, em particular, as relações entre a IA e uma nova modalidade de lógica, denominada Lógica Paraconsistente.

Abstract

The author analyses the contribution of AI for the better understanding of the present state of development of logic. In particular, the relations between AI and a new kind of logic, called Paraconsistent Logic, are considered.

Introdução

A lógica, até bem pouco tempo, era uma ciência una, que progredia linearmente, mesmo após sua matematização por matemáticos, lógicos e filósofos como Boole, Peano, Frege, Russell e Whitehead. Os desenvolvimentos revolucionários, efetuados na década de 30, como os devidos a Gödel e Tarski, ainda se situam dentro do que podemos denominar de lógica clássica ou tradicional.

Não obstante todos os avanços da lógica tradicional, uma outra revolução paralela ocorreu no domínio da ciência criada por Aristóteles, de índole muito diversa. Referimo-nos à instituição das lógicas não-clássicas. Elas produziram, como no caso da geometria não-euclidiana, uma transformação de natureza profunda no âmbito científico, cujas consequências, de caráter filosófico, ainda não se investigaram de maneira sistemática e integral.

Designamos por lógica clássica ou tradicional, essencialmente, o estudo do cálculo de predicados de primeira ordem, com ou sem igualdade, bem como de certos de seus subsistemas, como o cálculo proposicional clássico, e de algumas de suas ampliações, por exemplo a lógica tradicional de ordem superior (teoria de tipos) e os sistemas comuns de teoria dos conjuntos (Zermelo-Fraenkel, von Neumann-Bernays-Gödel, Kelley-Morse, NF de Quine-Rosser, ML de Quine-Wang, ...). A lógica em apreço baseia-se em sintaxe e semântica bem estabelecidas; assim, a semântica usual do cálculo de predicados funda-se no conceito de verdade de Tarski.

As lógicas não-clássicas caracterizam-se por amplificar, de algum modo, a lógica tradicional ou por infringir ou limitar seus princípios ou pressupostos nucleares.

Dentre as primeiras, ditas lógicas complementares da clássica, lembraremos as lógicas

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 127/132
1992, APDC & SPF, Lisboa

tradicionais das modalidades aléticas, das modalidades deônticas, dos operadores epistêmicos e dos operadores temporais. Dentre as segundas, chamadas de heterodoxas ou rivais da clássica, citaremos as lógicas paraconsistentes, as paracompletas e as intuicionistas sem negação (Griss, Gilmore, ...). Para detalhes sobre essas lógicas, pode-se consultar, por exemplo, da Costa [4] e [5], Haack [12] e [13] e Rescher [15].

A lógica, devemos salientar, é muito mais do que a disciplina das formas válidas de inferência. Seria difícil enquadrarmos, v.g., a teoria dos modelos, em sua forma atual, e a teoria da recursão em uma lógica assim definida. No entanto, para os fins deste artigo, podemos identificar a lógica (dedutiva) como a disciplina que se ocupa especialmente das formas válidas de inferência (ou de raciocínio).

Por outro lado, a cada lógica dedutiva L acha-se, normalmente, associada uma lógica indutiva L' que, em certas condições, indica de que maneira inferências inválidas segundo L podem ainda ser utilizadas. Os padrões para que isto seja legítimo se codificam em L'. (Ver Mikenberg et alii [14] e da Costa e French [6]). As lógicas indutivas situam-se, perfeitamente, entre as lógicas não-clássicas (talvez como *suplementares* das correspondentes lógicas dedutivas).

A Inteligência Artificial (IA) contribuiu para o progresso de várias lógicas novas (lógicas não-monotônicas, lógicas de 'default', lógicas paraconsistentes em geral,...). Isto se dá pelo fato de, particularmente no caso dos sistemas especialistas, necessitarmos de formas não tradicionais de inferência. A paraconsistência, v.g., se impõe, pois regularmente se trabalha com conjuntos inconsistentes (contraditórios) de informações.

Neste artigo procuraremos delinear a grande relevância que a IA está adquirindo, no tocante à compreensão profunda do significado mesmo da logicidade (e, indiretamente, para o próprio entendimento da razão, sua estrutura, limites e formas de aplicação). Para tanto, concentrar-nos-emos, apenas, no caso da lógica paraconsistente, com segurança uma das mais heterodoxas dentre as lógicas heterodoxas.

A lógica paraconsistente

Uma teoria (dedutiva) T, assentada na lógica L, diz-se consistente se entre seus teoremas não existem dois, tais que um seja a negação do outro; em hipótese contrária, T denomina-se inconsistente. A teoria T chama-se trivial se todas as sentenças (fórmulas fechadas) de sua linguagem forem teoremas; se isto não ocorrer, T é não-trivial.

Se L for uma das lógicas comuns, como a clássica e a intuicionista de Brouwer-Heyting, T é trivial se, e somente se, for inconsistente. Em outras palavras, lógicas como essas não *separam* os conceitos de inconsistência e de trivialidade.

L denomina-se *paraconsistente* se puder funcionar como fundamento de teorias inconsistentes e não triviais. Dito de outro modo, uma lógica paraconsistente mostra-se capaz de manipular sistemas inconsistentes de informações sem o perigo de trivialização. (Somente em certas circunstâncias específicas a presença de contradição implica em trivialização).

Os precursores da lógica paraconsistente foram o lógico polonês J. Lukasiewicz e o filósofo russo N.A. Vasilev. Simultaneamente, por volta de 1910, embora de maneira independente, eles ventilaram a possibilidade de uma lógica paraconsistente que restringiria, por exemplo, o princípio de contradição, quando formulado do seguinte modo: Dadas duas proposições contraditórias, isto é, uma das quais é a negação da outra, então uma das proposições é falsa. Vasilev chegou, mesmo, a articular determinada lógica paraconsistente, que ele batizou de *imaginária*, modificando a silogística aristotélica. Nenhum deles tinha,

na época, uma visão ampla da lógica clássica, tal como hoje a encaramos; eles a tratavam mais ou menos do prisma de Aristóteles, de conformidade com as tendências então dominantes na área.

O primeiro lógico a estruturar um cálculo proposicional paraconsistente foi o polonês S. Jáskowski, discípulo de Lukasiewicz. Em 1948 ele publicou suas idéias sobre lógica e contradição, mostrando como se poderia construir um cálculo sentencial paraconsistente, possuindo motivação conveniente. O sistema de Jáskowski, nomeado por ele de *lógica discursiva*, desenvolveu-se posteriormente (a partir de 1968) em virtude das obras de autores como J. Kotas, L. Furmanowski, L. Dubikajtis, N.C.A. da Costa e C. Pinter. Assim, chegou-se a edificar uma verdadeira lógica discursiva, englobando um cálculo de predicados de primeira ordem e uma lógica de ordem superior (há, inclusive, teorias discursivas de conjuntos, intrinsecamente ligadas à teoria de atributos, baseada no cálculo S5 de Lewis).

Os sistemas iniciais de lógica paraconsistente, contendo todos os níveis lógicos, envolvendo, pois, cálculos proposicionais de predicados e de descrições, bem como lógicas de ordem superior, devem-se a N.C.A. da Costa (1954 em diante). Isto se realizou de maneira independente das perquirições dos autores já mencionados.

Hoje em dia existem, inclusive, sistemas paraconsistentes de teorias de conjuntos, estritamente mais fortes do que os clássicos, porquanto os contêm como subsistemas estritos, e matemáticas paraconsistentes. Aliás, estas matemáticas encontram-se aparentadas com a matemática *difusa* (fuzzy) que, sob certo ponto de vista, encaixa-se no rol daquelas.

Ficou provado, em decorrência da elaboração da lógica paraconsistente, que se torna possível manipular sistemas informacionais inconsistentes e extraordinariamente fortes, sem ser preciso que se eliminem as contradições e sem se cair na trivialização.

Digno de menção é o fato de que a lógica paraconsistente nasceu de considerações puramente teóricas, tanto lógico-matemáticas, como filosóficas. As primeiras se referem, para exemplificar, a problemas relativos ao conceito de verdade, aos paradoxos da teoria dos conjuntos e à vaguidade inerente não somente à linguagem natural, como na científica. As segundas, por seu turno, encontram-se correlacionadas com temas tais como os fundamentos da dialética, as noções de racionalidade e de logicidade, e a aceitação de teorias científicas.

Algumas das consequências da estruturação da lógica paraconsistente, que podemos classificar em duas categorias, as 'positivas' e as 'negativas', são as seguintes: Positivas: 1) Melhor elucidação de alguns conceitos centrais da lógica, tais como os de negação e de contradição, bem como o papel do esquema da abstração em teorias dos conjuntos (antinomias conjuntistas); 2) Compreensão mais profunda de determinadas teoria filosóficas, especialmente a dialética e a teoria dos objetos de Meinong; 3) Comprovação da possibilidade de teorias fortes e inconsistentes, embora não triviais (os paradoxos comuns podem ser tratados de uma nova perspectiva); 4) Organização de esquemas ontológicos diferentes dos da ontologia tradicional. Negativos: 1) Demonstração de que certas críticas à dialética afiguram-se infundadas (por exemplo, os conhecidos reparos de Popper); 2) Prova de que as exigências metodológicas impostas às teorias científicas se mostram demasiadamente restritivas e mereceriam ser liberalizadas; 3) Evidência de que a concepção usual da verdade como correspondência, à la Tarski, não acarreta as leis da lógica clássica, sem suposições extras, geralmente mantidas implícitas.

Detalhes sobre a lógica paraconsistente acham-se em trabalhos como os seguintes: Arruda [1], da Costa [4] e [5], da Costa e Marconi [7] e Grana [11].

Hodiernamente, em IA, precisamos manipular sistemas inconsistentes de informação;

mais ainda, necessitamos *processar* semelhantes sistemas, via uma programação paraconsistente. Tentar transformar esses sistemas em outros consistentes seria não apenas impraticável, mas, sobretudo, teoricamente fora de propósito. Logo, a IA constitui um campo onde a lógica paraconsistente topa naturalmente com aplicações de suma importância. Assim, a informática em geral está intimamente unida à paraconsistência. Para detalhes, o leitor pode consultar, entre outros, os seguintes artigos: Blair e Subrahamanian [3], Subrahmanian [16] e [17] e Batens [2]. Sob certo ângulo, as próprias lógicas não monotônicas e do 'default' incluem-se na classe das paraconsistentes (em acepção ampla).

IA e paraconsistência

Em conexão com a exposição precedente, eis alguns problemas filosoficamente significativos: a) As lógicas não-clássicas são realmente lógicas? b) Pode mesmo existir lógicas rivais da clássica? c) Em última instância, não seriam as lógicas chamadas de rivais apenas complementares da clássica? d) Que relação há entre racionalidade e logicidade? e) A razão pode se expressar por meio de lógicas distintas, incompatíveis entre si?

Obviamente, nos limites do presente artigo não podemos tratar de todas essas indagações, nem mesmo de forma resumida. (Aliás, já discorremos sobre elas, com detalhes, em da Costa [4]).

Porém, adotando-se posição 'operacional', se sistema lógico denotar uma espécie de *organon* de inferência, a IA contribui para nos conduzir, inescapavelmente, à conclusão de que há várias lógicas, clássicas e não-clássicas, e entre as últimas, complementares e rivais da clássica.

Ademais, a IA corrobora a possibilidade e a relevância prática de lógicas da categoria das paraconsistentes, tão afastadas dos padrões até há pouco tempo fixados para a logicidade. Isto constitui, sem dúvida, algo surpreendente para quem não está acostumado com os últimos avanços da informática.

Convém lembrar aqui que existem numerosos argumentos enfraquecedores da posição dos que defendem a tese do caráter absoluto da lógica clássica. Arrolamos, a seguir, quatro desses argumentos:

1) Qualquer contexto racional dado é compatível com infinitas lógicas capazes de figurarem como lógicas a ele subjacentes.

2) Os conceitos lógicos fundamentais, como o de negação, têm que ser vistos como a 'family resemblance' no sentido de Wittgenstein. Não há nenhum motivo especial para se recusar, digamos, à negação paraconsistente a dignidade de negação; se alguém assim proceder, deveria, também, sustentar que as retas das geometrias não euclidianas não são, efetivamente, retas ...

3) A semântica comum, por exemplo do cálculo restrito de predicados, baseia-se na teoria de conjuntos. Como há diversas teorias de conjuntos (clássicas), há numerosas interpretações possíveis, não equivalentes entre si, de tal semântica. Por conseguinte, o referido cálculo não se mostra tão bem determinado como parece à primeira vista.

4) Para a lógica tradicional de segunda ordem (e de ordem superior) não existe axiomatização correta e completa. Ela escapa, pois, à axiomatização (recursiva).

Assim, as contestações às indagações *a* e *b* são afirmativas. Uma resposta simples à questão *c* parece difícil: no fundo, trata-se, em grande parte, de um problema terminológico. Todavia, em princípio, em virtude da discussão anterior, nada impede que se aceite que há lógicas rivais, que não se incluem no rol das complementares da tradicional. Finalmente, sobre *d* e *e*, frisaremos que se possui excelentes argumentos para se demonstrar que a razão

permanece razão mesmo quando se manifesta por meio de lógicas não-clássicas. (A própria lógica clássica não constitui um sistema bem definido).

Do exposto, pensamos que as conclusões que se impõem são suscetíveis de resumo, como se segue:

"De fato, como insiste Enriques [Enriques [8], [9] e [10]], a ciência aparece a todo momento como imperfeita, em qualquer de suas partes, desenvolvendo-se por auto-correção e auto-integração. Não constitui aquisição hirta, à qual se vão juntando pouco a pouco outras; há permanente vaivém desde os fundamentos até as teorias mais complexas, corrigindo-se erros aqui, eliminando-se inconsistências acolá. Mas a história comprova que toda teoria científica encerra algo de verdadeiro: a mecânica newtoniana, embora superada pela de Einstein, evidentemente contém traços de verdade; restringindo-se de maneira conveniente o seu campo de aplicação, ela funciona, prevê, e, portanto, *tem* que conter uma parcela de verdade. Mas a verdadeira verdade é um caminhar constante para a verdade — eis o ensinamento da história, fora de qualquer dúvida séria. A ciência é mais luta, avanço, do que etapa adquirida, conquistada, e as categorias científicas fundamentais modificam-se no decorrer do tempo." (da Costa [4], p. 231).

Mais ainda:

"A lógica se vai constituindo através da história e não parece possível prever-se as vicissitudes de sua evolução. Um especialista do início do século, ainda que familiar com as obras de Frege, Russell e Peano, dificilmente poderia prever as transformações que se processariam na lógica nos últimos quarenta anos. Não se trata somente de progresso em extensão; o próprio conceito de logicidade se modificou. Hoje, as lógicas heterodoxas entraram em cena com grande ímpeto: ninguém conseguiria predizer onde nos levarão as lógicas polivalente, relevante e paraconsistente. Talvez nos próximos anos nos esteja reservada uma nova alteração da idéia de logicidade, impossível de se imaginar no momento." (da Costa [4], p. 232).

"Razão, como se definiu ..., é a faculdade de conceber, julgar e raciocinar. Conceber e raciocinar constituem patrimônio exclusivo da razão; mas julgar, no sentido preciso da palavra, também é atividade racional. Mesmo quando a inspiração, a sensibilidade ou alguma forma primitiva de intuição não racional fornecem a base para o juízo, é a razão que julga, dado que somente ela manipula e combina conceitos. Aliás, a maioria dos usos comuns da palavra 'razão' derivam-se de razão conceituada como a faculdade de conceber, julgar e raciocinar; assim, para discernir bem e adotar normas racionais de vida, há de se ter em conta a razão na acepção definida. Além disso, existe um conjunto de regras e de princípios regulando o uso da razão, especialmente como ela se manifesta nos contextos racionais. Também é lícito chamar de razão a esse conjunto de regras e de princípios. Pois bem, quando se indaga se a razão se transforma ou mantém-se invariante, torna-se mais conveniente, sem dúvida, interpretar a questão como se referindo à razão qual conjunto de regras e de princípios, e não como faculdade. Assim formulado, o problema tem resposta imediata: a razão modificou-se no decurso do tempo. Por exemplo, as categorias racionais subjacentes à física aristotélica, à newtoniana e à hodierna, divergem profundamente; *ipso facto*, os princípios que regem essas categorias variam, donde se conclui que a própria razão se transformou." (da Costa [4], p. 233).

Exprimindo-se por intermédio de lógicas as mais variadas, a razão não deixa de ser razão...

A IA é, atualmente, um dos pilares nos quais se apoiam as ponderações que acabam de ser feitas. Então, ela não possui apenas valor prático, de aplicação tecnológica, como também valor teórico, contribuindo para o melhor equacionamento dos problemas da

lógica, da razão e, portanto, da própria cultura.

Bibliografia

[1] ARRUDA, A.I., *'A survey of paraconsistent logic'*, em *Mathematical Logic in Latin America*, editado por A.I. Arruda, R. Chuaqui e N.C.A. da Costa, North-Holland, Amsterdão, 1980, pp. 1-41.

[2] BATENS, D., *'Static and dynamic paraconsistent logics: a plea for their use in expert systems'*, CCAI 3 (1986), nº 1-2, 33-50.

[3] BLAIR, H.A. e V.S. Subrahmanian, *'Paraconsistent foundations for logic programming'*, Pré-publicação da School of Computer and Information Science, Syracuse University, 1988.

[4] DA COSTA, N.C.A., *Ensaio sobre os Fundamentos da Lógica*, Hucitec, São Paulo, 1980.

[5] DA COSTA, N.C.A., *'The philosophical import of paraconsistent logic'*, Journal of Non-Classical Logic 1 (1982), 1-19.

[6] DA COSTA, N.C.A. e S. French, *'Pragmatic truth and the logic of induction'*, a aparecer no British Journal for the Philosophy of Science (1989).

[7] DA COSTA, N.C.A. e Marconi, *'An overview of paraconsistent logic'*, Pré-publicação da Sociedade Paranaense de Matemática, 1987, a aparecer em Logica Nova.

[8] ENRIQUES, F. e G. Santillana, *Storia del Pensiero Scientifico*, vol. 1, Zanichelli, Bolonha, 1932.

[9] ENRIQUES, F., *Per la Storia della Logica*, Zanichelli, Bolonha, 1922.

[10] ENRIQUES, F. e G. Santillana, *Compendio di Storia del Pensiero Scientifico*, Zanichelli, Bolonha, 1936.

[11] GRANA, N., *Logica Paraconsistente*, Loffredo, Nápoles, 1983.

[12] HAACK, S., *Deviant Logic*, Cambridge University Press, Cambridge, 1977.

[13] HAACK, S., *Philosophy of Logics*, Cambridge University Press, Cambridge, 1978.

[14] MIKENBERG, I., N.C.A. da Costa e R. Chuaqui, *'Pragmatic truth and approximation to truth'*, Journal of Symbolic Logic 51 (1986), 201-221.

[15] RESCHER, N., *Topics in Philosophical Logic*, Reidel, Dordrecht, 1968.

[16] SUBRAHMANIAN, V.S., *'Query processing in quantitative logic programming'*, em *Proc. 9th. Int. Conference on Automated Deduction*, editado por E. Lusk e R. Overbeek, Springer, Berlim, 1988, pp. 81-100.

[17] SUBRAHMANIAN, V.S., *'Mechanical proof procedures for many-valued lattice-based logic programming'*, Pré-publicação da School of Computer and Information Science, Syracuse University, 1988.

Antonio A. Martino

Univ. de Pisa, Pisa & Inst. para a Documentação Jurídica, CNR, Florença (Italia)

Carlos E. Alchourrón

Inst. de Filosofia da Fac. de Direito, Univ. de Buenos Aires, Buenos Aires (Argentina)

Lógica sem Verdade

Sumário

1. O dilema de Jørgensen. 2. A solução proposta. 3. Breve história de um preconceito filosófico. 4. Normas sem lógica (ou os preços ontológicos). 4.1. Imperativos sem lógica. 4.2. Normas e proposições normativas. 4.3. Proposições normativas e normas verdadeiras. 4.4. Mundos possíveis. 4.5. Validade e invalidade. 4.6. A racionalidade do legislador. 5. A noção abstracta, sintáctica e semântica de consequência. 5.1. A noção abstracta de consequência. 5.2. A noção sintáctica de consequência. 5.3. A noção semântica de consequência. 5.4. O sentido dado por regras de uso num contexto. 5.5. Que operadores para a lógica. 5.6. Só num contexto a parte ganha sentido. 6. A lógica deôntica. 7. Por que razão a solução do dilema interessa a toda a lógica. 8. Consequências para informáticos.

Summary

1. The dillema of Jørgensen. 2. The solution we advocate. 3. Brief history of a philosophical prejudice. 4. Norms without logic (or the ontologic costs). 4.1. Imperatives without logic. 4.2. Norms and normative propositions. 4.3. Normative propositions and true norms. 4.4. Possible worlds. 4.5. Validity and invalidity. 4.6. The legislator's rationality. 5. The abstract, syntactic and semantic notion of consequence. 5.1. The abstract notion of consequence. 5.2. The syntactic notion of consequence. 5.3. The semantic notion of consequence. 5.4. Meaning as provided by rules of use within a context. 5.5. Which operators for Logic? 5.6. The part cannot earn meaning but in context. 6. The deontic Logic. 7. Why the dillema concerns all the Logic. 8. Consequences for computer experts.

1. O Dilema de Jørgensen

Pode dizer-se que o conteudo fundamental da lógica está na obra de Aristóteles , na parte do *"Organon"* que são os *Primeiros Analíticos*, onde se trata do silogismo entendido como uma forma na qual a partir de um conjunto de proposições se segue necessariamente alguma outra.

Toda a história da lógica consistiu em esclarecer essa relação de consequência entre as premissas e a conclusão.

A partir desta relação típica que é a de consequência podem definir-se as duas propriedades relevantes que são a incompatibilidade e a consistência. Uma vez definido este trio, é fácil caracterizar o sentido dos conectivos lógicos.

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 133/150
1992, APDC & SPF, Lisboa

Na lógica standard a noção de consequência caracteriza-se a partir de e com relação aos enunciados verdadeiros ou falsos, de tal modo que dizer que a um enunciado se segue um conjunto de enunciados, significa afirmar que em todas as interpretações possíveis em que resultem verdadeiros todos os enunciados que figuravam como premissas, é verdadeiro o enunciado que figura como conclusão. Também o sentido dos conectivos lógicos está caracterizado pelo valor de verdade que recebem os compostos proposicionais a partir dos valores de verdade das proposições componentes.

Muitas normas, e particularmente as que mais interessam a juristas, políticos e moralistas, expressam-se com enunciados que carecem de valores de verdade. Contudo é usual constatar que juristas, moralistas e politologos inferem normas de outras normas, declaram que certas normas são incompatíveis entre si, assim como realizam operações com conectivos lógicos aplicados às normas para dizer que uma implica a outra, ou que duas obrigações conjuntivas devem ser ambas satisfeitas, etc.

Se a aplicação lógica tem lugar entre enunciados verdadeiros ou falsos e as directivas não têm valores de verdade terá de se concluir que uma lógica de normas não é possível. Portanto, ou as relações lógicas se dão entre enunciados susceptíveis de verdade ou falsidade e então não existe uma lógica normativa, ou as relações lógicas se aplicam também às normas, mas então a lógica não se ocupa só de inferências entre enunciados verdadeiros ou falsos.

Dito de outro modo, partindo da pressuposição de que as normas carecem de valores de verdade, de duas uma: 1. ou a noção de inferência e os conectivos proposicionais se caracterizam em relação às noções de verdade e falsidade e não há relação lógica entre as normas nem se lhes podem aplicar conectivos lógicos; 2. ou há uma lógica das normas mas então a noção de inferência não pode ser caracterizada pelas noções de verdade e falsidade nem os conectivos lógicos recebem o seu sentido da dependência do valor de verdade dos seus compostos.

Tal é o dilema de Jørgensen [1].

Até E. Mally publicar o seu *"Grundgesetze des Sollens"* [2], em 1926, a lógica não se ocupou das normas mas apenas das relações entre proposições verdadeiras ou falsas. O dilema de Jørgensen consiste em considerar estes "encadeamentos de enunciados como raciocínios e modificar, em tal caso, a concepção tradicional da lógica assim como várias noções lógicas (as de negação proposicional, de implicação, de equivalência, etc.) ou salvaguardar a noção vigente de lógica negando o carácter de raciocínios destes encadeamentos de proposições" [3].

A assunção do dilema, já o dissemos, consiste em negar valor de verdade às normas.

A primeira alternativa oferece uma solução radical ao dizer que, dado que não têm valores de verdade ou falsidade, as normas nada têm a ver com a lógica. Isto comporta graves consequências.

As teorias menos dramáticas preferem dizer que, embora as normas não tenham valores de verdade, as suas relações lógicas podem ser explicadas por meio de proposições normativas que descrevem as normas e que pertencem ao discurso indicativo [4], ou então sustentar que as normas não têm valor de verdade mas são susceptíveis de outros dois valores, incompatíveis entre si e necessariamente atribuíveis a cada norma. Esta posição, que é a mais difundida, identifica os dois valores com validade e invalidade [5].

Há uma terceira posição que, reconhecendo a alogicidade do discurso deôntico, atribui capacidades a um "legislador racional" de que se socorre para caracterizar os comportamentos deônticos adequados.

Uma quarta possibilidade consiste em negar a afirmação segundo a qual as normas não

são susceptíveis de verdade ou falsidade, o que impede o enunciado da primeira parte do dilema. Isto implica que todas as normas são susceptíveis de verdade ou falsidade, e sendo uma característica da teoria da correspondência que a verdade ou a falsidade dependem dos factos, deve-se admitir que existem factos normativos [6].

Uma quinta possibilidade consiste em desenvolver a teoria semântica dos mundos possíveis para explicar em que mundos as normas podem dizer-se verdadeiras ou falsas, no sentido específico em que se fala de verdade ou falsidade com respeito a um mundo.

2. A Solução Proposta

Noutra alternativa opta-se pela segunda haste do dilema: aceitando-se que as normas, apesar de não serem verdadeiras ou falsas, têm uma lógica, explicam-na não a partir dos conceitos de verdade ou falsidade mas tomando como primitiva a noção de consequência. O sentido dos conectivos lógicos, deônticos, ou o que quer que sejam, explicita-se pelo uso que se faz deles num contexto de dedução.

Nesta alternativa nega-se que a caracterização semântica das relações e propriedades lógicas seja imprescindível, e daí o nome de "lógica sem verdade".

Se se prescinde da noção semântica de consequência, a lógica deôntica deixa de enfrentar obstáculos filosóficos.

Estamos convencidos de que esta é uma solução muito razoável, sem grandes novidades teóricas mas que um velho preconceito filosófico impediu de ver no passado.

Mais do que novidade teórica o que a solução apresenta são consequências importantes: a) ampliando-se o âmbito da lógica para fora do juizo descritivo não só se pode falar de lógica de normas, mas também noutros campos menos tradicionais; b) dado que os computadores recebem somente instruções sintácticas, os informáticos vêm o seu âmbito de actuação garantido e reforçado.

Por último, embora acreditemos firmemente na razão do que enunciamos, pensamos que as discussões filosóficas sobre os fundamentos da lógica não devem paralisar o desenvolvimento técnico e científico, mas, antes pelo contrário, estimulá-lo.

3. Breve História de um Preconceito Filosófico

Hao Wang conta que Gödel percebera que só o preconceito do finitismo impediu Skolem de ver que os seus descobrimentos levavam ao teorema da completude semântica [7]. Cremos que um preconceito semelhante impediu de ver que os desenvolvimentos de lógicas de sequências tornaram não indispensáveis as pressuposições semânticas para aclarar a noção de consequência.

Aristóteles teve em conta as dificuldades que apresentavam os juizos não indicativos e manifestou expressamente que a lógica da qual se ia ocupar só era aplicável a expressões susceptíveis de verdade ou falsidade, deixando à retórica ou à poesia o estudo das outras expressões com sentido [8].

Esta ideia percorre toda a história da lógica e chega até Tarski. Tarski irá sustentar uma posição similar à de Aristóteles reservando só para os juizos descritivos o campo da lógica e construindo uma versão semântica da verdade como "correspondência" da realidade. A noção de consequência será considerada como uma função de valores de verdade dos enunciados expressos nas premissas e na conclusão [9].

A ideia da verdade como correspondência manifesta-se no facto de um enunciado satisfazer a condição de Tarski. Esta condição pode ser enunciada assim: um enunciado é

verdadeiro, se, e só se, o que diz o enunciado acontece. É a idéia segundo a qual se diz que a lógica não é arbitrária porque depende das correlações semânticas da linguagem que usamos para nos referirmos ao mundo. Esta ideia pode-se encontrar no *Tractatus* [10]. É a ideia que define os conectivos em função dos valores de verdade das proposições que afectam e por isso não é difícil compreender que seja no mesmo *Tractatus* que as tábuas de verdade aparecem pela primeira vez [11].

Ocupando-se a lógica basicamente das noções de consequência, incompatibilidade e consistência, é sumamente importante caracterizar estes conceitos, e o modo clássico consistiu no recurso às noções de verdade e falsidade como primitivas para se ter noções unívocas, não arbitrárias. Até Carnap, que apresenta uma versão "convencionalista" da lógica, permitindo qualquer tipo de eleição arbitrária de regras de dedução sempre que não se tenha dado anteriormente uma interpretação, sustenta que qualquer interpretação deve ser justificada em termos de verdadeiro ou falso, e assim a liberdade sintáctica resulta condicionada pelas noções semânticas [12].

Tarski e Carnap deram-se conta da arbitrariedade histórico-convencional da redução da lógica ao estudo de um número muito restrito de operadores. Chegam inclusivé a dizer que não há uma boa explicação para isso, mas, no fundo, temem a expansão de tais conectivos porque poderia pôr em perigo a primazia da preferência da condição semântica de Tarski. Verdade e falsidade continuam a significar para eles, como para Quine, a única referência evidente. Temem, também, que a extensão dos operadores possa levar a lógica para o caminho, ameaçado por Wittgenstein, de se ocupar não só dos conectivos, como de ter também de aclarar todos os termos da linguagem.

Ninguém usaria a lógica numa extensão tão vasta, já que, a ocupar-se de tudo o que é analítico, se exorbitaria da lógica clássica.

Vincularmo-nos ao que se chama efectivamente lógica, é vincularmo-nos à elaboração do sentido de certas palavras e muito poucas; palavras que, além disso, devem ter já um significado dado com relação aos valores de verdade ou falsidade das proposições que afectam. É o que sustenta Prior de modo lacónico, elegante e contundente [13], negando que o sentido dos conectivos possa ser definido em termos de deducibilidade, mas sim que é necessário, afirma ele, recorrer às tábuas de verdade, isto é, a uma explicação semântica. A sua argumentação consiste em inventar um conectivo arbitrário 'tonk' caracterizado por regras de inferência que permitem "demonstrar" um absurdo.

Aristóteles e Tarski dão-se perfeitamente conta de que estão a deixar de lado uma parte dos juizos, todos os que não sejam enunciados asseverativos; consideram que as dificuldades contidas na superação da linhagem semântica justifica esta amputação.

Eis uma boa razão para se explicar a razoabilidade e durabilidade do preconceito, baseado na linhagem filosófica que desde sempre explicou a validade de um raciocínio com base nos valores de verdade das partes componentes. Outra boa razão são os resultados excelentes a que a lógica chegou e o medo de os pôr em perigo. Uma terceira razão, a não desdenhar, é que, quando se pensa em lógica, se pensa numa série de operações sintácticas que esperam interpretação, e essa interpretação consiste em atribuir valores de verdade.

Esta linhagem esqueceu a outra vertente da história da lógica que parte da ideia global do raciocínio como inteligível e explica o sentido das partes, inclusivé o dos conectivos, pela função que cumprem no raciocínio.

Esta ideia é esclarecida através de uma apresentação axiomática da noção de deducibilidade e por este motivo os seus melhores desenvolvimentos são relativamente recentes, se bem que a sua presença possa ser rastreada desde Platão.

O preconceito da insuperabilidade das dificuldades suscitadas pelo discurso não

descritivo continuou a não deixar ver que o próprio Tarski tinha oferecido uma noção de consequência muito mais abstracta do que a noção de implicação lógica baseada na verdade e falsidade e que, apoiando-se naquela noção, Gentzen construira um sistema que não necessita de qualquer referência semântica [14].

Este preconceito impede que os lógicos se ocupem saudavelmente das lógicas de alguns discursos, como o discurso normativo, prescindindo de questionar previamente os valores de verdade dos enunciados componentes, e antes tratando da identificação clara e precisa das funções dos conectivos tipicos dessas outras lógicas.

Preconceito que obriga alguns autores a verdadeiros saltos mortais para darem algum sentido às operações lógicas com normas, e outros, a interrogarem-se por vezes sobre o valor dos próprios esforços, porventura destinados ao fracasso.

Vamos tratar de ilustrar sumariamente, e por tipos, tais desventuras da lógica deôntica e demonstrar num capítulo à parte a sem razão do preconceito.

4. Normas sem Lógica (ou os preços ontológicos)

Perante o dilema de Jørgensen a posição mais generalizada consiste em sustentar a validade da primeira haste do dilema, isto é: dado que as relações e as propriedades lógicas têm que ver com verdade e falsidade, e que as normas não são nem verdadeiras nem falsas, as normas não têm que ver com a lógica.

Perante as consequências contraintuitivas que tal posição provoca, podem distinguir-se várias posições que, com diferentes manejos, tentam salvar algumas operações "lógicas" que se efectuam com normas e dentro das normas. Vamos examiná-las muito brevemente, dando de barato que se trata de classificar posições e, portanto, que a referência a autores é simplesmente exemplificativa, sem qualquer pretensão histórica.

4.1 Imperativos sem Lógica

Esta é sem dúvida a posição mais simples e radical. Perante o dilema de Jørgensen aceita que as normas não têm a ver com a lógica.

Às normas não podem ser aplicadas nem as propriedades nem as relações lógicas. Particularmente não pode ser aplicada às normas a noção de consequência, já que não há afirmação verdadeira nem falsa sobre a qual se possa construir a noção de consequência.

Todo o empirismo lógico pode ser um bom exemplo ao considerar a linguagem normativa como discurso emotivo, e o mais categórico é Ayer [15].

Outro exemplo ilustre, se bem que menos claro, é o Kelsen da *Teoria Geral das Normas*. Kelsen assevera que interrogarmos-nos sobre se existe uma relação entre condição e consequência significa interrogarmos-nos sobre se as asserções envolvidas são verdadeiras ou falsas, e, obviamente, conclui que o princípio da não-contradição e a relação de consequência não são aplicáveis às normas [16].

Dentro desta posição pode distinguir-se a daqueles autores, como Hare por exemplo, que embora irredutíveis com respeito aos imperativos, começam a estabelecer diferenças dentro dos enunciados normativos e distinguem a parte descritiva (frástico) e a parte prescritiva das normas (néustico), dizendo que a primeira se comporta como uma proposição descritiva que, portanto, admite o cálculo lógico, enquanto a segunda não admite qualquer classe de operações lógicas por ser um imperativo [17].

Os próprios defensores desta posição resistiram a admitir todas as desagradáveis consequências que a mesma acarreta. Se não é possível aplicar às normas a noção de

consequência lógica, também não é possível aplicar as propriedades da incompatibilidade e da consistência, que são interdefiníveis com a primeira.

Mas, além disso, tão-pouco podem ser aplicados os conectivos proposicionais, que também se definem em termos de verdade e falsidade; e faltando os conectivos é difícil fazer qualquer discurso com normas, já que não se podem ligar os átomos em enunciados moleculares.

Mais ainda, se não é possível formular juizos de inconsistência entre as normas, como é possível que advogados e juristas discutam sobre tais conflitos, a ponto de terem inventado critérios de solução como *lex posterior derogat priori?*

4.2 Normas e Proposições Normativas

Perante a gravidade das consequências de se aceitar, sem mais considerações, a primeira haste do dilema, uma posição mais conciliadora consiste em distinguir entre normas e proposições normativas. Parte-se da noção genérica de enunciado deôntico e distingue-se uma primeira interpretação prescritiva, não susceptível de comércio lógico, e uma segunda interpretação, esta descritiva, capaz de aceitar as categorias de verdadeiro e falso, dado que são asserções sobre normas.

Supondo-se a diferença entre uma linguagem e a realidade, é possível notar que na primeira existe uma oração que num certo uso expressa uma norma e noutro uso expressa uma proposição normativa. Quando expressa uma norma o seu tema é normar acerca do que acontece no mundo, quando expressa uma proposição normativa o seu tema é a norma que norma sobre o mundo.

Os enunciados deônticos podem, portanto, servir seja para normar sobre a realidade, seja para fazer afirmações acerca das normas (neste caso tais enunciados são metalinguísticos com respeito a estas) que regulam na realidade.

Os enunciados deônticos admitem assim duas interpretações: a prescritiva, na qual as normas são expressadas, que não consente valores de verdade, e a descritiva, na qual se expressam proposições normativas que têm valores de verdade. Esta distinção serve para sustentar que a lógica deôntica é basicamente a teoria da interpretação descritiva e por essa razão lhe são aplicáveis as noções lógicas clássicas.

Um exemplo célebre e complexo desta posição encontra-se em *Norma e Acção* de Von Wright, descrito pelo próprio [18]: "Em *Norm and Action* (1963) fiz uma distinção tripartida, que considero útil, entre normas, formulações normativas e proposições normativas. As formulações normativas têm uma característica "ambiguidade": a mesma e única expressão pode ser usada seja *prescritivamente* para enunciar uma norma ou uma regra de conduta, seja *descritivamente* para afirmar que existe tal norma ou tal regra. A minha opnião, então, era que a lógica deôntica consistia numa lógica de formulações normativas formalizadas interpretadas descritivamente".

Esta posição está ligada ao expressivismo: as normas expressam uma atitude do falante mas não descrevem facto algum. O mais que se pode fazer com elas é descrevê-las, afirmar proposições acerca delas.

A ideia geral consiste em poder explicar as relações existentes entre as normas, partindo de que só há relações lógicas entre as proposições normativas. Para tanto haveria que inventar um cálculo paralelo tal que tivesse todas as propriedades da verdade e da falsidade mas que reflectisse as relações que efectivamente existem entre normas. Tendo em conta que, negando as relações lógicas entre normas, não sabemos sequer o que se possa querer dizer com a afirmação de que o cálculo proposicional deve "reflectir todas

as propriedades e relações existentes entre as normas'', a tarefa parece sumamente árdua, mesmo intuitivamente.

4.3 Proposições Normativas e Normas Verdadeiras

Existe uma variante da versão anterior que, pelas consequências que comporta, deveria ser tratada fora deste capítulo e num outro ponto que já insinuámos, mas que vamos manter aqui por comodidade de exposição. É aquele ponto de vista que sustenta a necessidade de se distinguir normas e proposições normativas, mas que afirma que as normas também são susceptíveis de verdade ou falsidade porque descrevem factos normativos particulares como "ter um dever''.

É a posição dos cognitivistas éticos como Moore ou David Ross. Há uma longa tradição em ética, segundo a qual as normas expressam proposições verdadeiras ou falsas: praticamente todo o cognitivismo, seja naturalista ou intuicionista, Spencer, Kant, Locke.

Numa ética descritiva, os enunciados que expressam normas são descrições e os factos que os tornam falsos ou verdadeiros existem na realidade.

Que factos? O facto de que Fulano tenha tal ou tal dever, o que é similar àquilo que Moore pensava ao considerar que ''bom'' indicava uma qualidade das coisas. Este facto encontra-se no mundo independentemente de haver quem o pense: eis o que determina a verdade ou a falsidade da afirmação de que Fulano deve fazer isto ou aquilo.

Ora esta não é uma proposição normativa; é um enunciado normativo que não expressa a proposição acerca de normas: expressa uma norma que é verdadeira ou falsa e tal verdade ou falsidade dependem do que acontece no mundo, de que no mundo se produza o facto normativo correspondente. Kalinowski sustentou esta tese.

Neste ponto, a pergunta é: onde se verificam factos normativos que convertam as normas em proposições verdadeiras acerca do mundo? Há várias respostas na bibliografia da lógica deôntica.

Tal posição impede a formação do dilema mas traz consigo o problema dos factos normativos.

4.4 Mundos Possíveis

Sustentar que existe uma certa correspondência entre o que diz um enunciado normativo, entendido prescritivamente, e uma certa realidade que torna verdadeiro ou falso esse enunciado, parece demasiado forte pois a primeira intuição que se tem é que não há, na realidade, nada que corresponda à verdade ou à falsidade das normas. O que sucede, sustentam os que tomam esta posição a sério, é que a realidade não se esgota no mundo real; existem, além disso, outros mundos possíveis.

David Lewis [19] por exemplo, sustenta que os mundos possíveis são tão reais como o mundo actual e que tudo isto faz parte da nossa ontologia. Se sustento que a realidade é mais ampla e que é constituída não sómente pelo que de facto acontece como também pelo que pode acontecer, então posso predicar a verdade ou a falsidade das normas num mundo determinado, se além disso puder indicar qual desses mundos tem a característica de ser um mundo óptimo. Dizer que no mundo actual é verdade que devo fazer tal ou tal coisa, significa que nos mundos que são alternativas óptimas do mundo em que estou se cumpre semelhante obrigação.

Em metalinguagem pergunta-se o que quer dizer que 'Op' seja verdadeira. Qual o facto que tem de acontecer para que seja verdade. Se supusermos que na realidade há

distintos mundos e em alguns deles acontece 'p' e noutros não acontece 'p', compreendemos porque a verdade não se define em geral senão em relação a um mundo. Suponhamos o mundo 'n': dizer que Op é verdadeiro num mundo 'n' significa que o requerido pela norma (ou seja 'p') se dá em todos os mundos ideais de 'n'.

Dizer que um enunciado normativo que expressa uma norma possui valor de verdade tem neste caso um sentido claro porque indica que tem de dar-se na realidade para que tal enunciado seja verdade; e para que seja verdade um enunciado de dever no mundo actual, tem que acontecer que, nos mundos possíveis que são mundos ideais relativamente ao mundo actual, deverá dar-se o conteúdo da obrigação.

Tudo isto permite um apoio nas noções de consistência, derivação, implicação lógica definidas em função da verdade, mas tem o preço de ter de se aceitar essa ontologia tão rica e tão complexa. Para manter a correlação entre consequência e verdade há que estar disposto a pagar o preço desta ontologia.

Se não se está disposto a pagá-lo, talvez porque não se sabe sequer quais são os instrumentos necessários para identificar tais mundos e os factos dentro desses mundos, o melhor será voltar à velha primeira haste do dilema e continuar a defender que as normas não são verdadeiras ou falsas, antes respondem a critérios bivalentes, por exemplo.

4.5 Validade e Invalidade

Neste caso aceita-se que as normas não tenham valores de verdade mas faz-se notar que as normas estão sujeitas às leis lógicas, só que os valores com os quais se manejam as normas, se bem que bivalentes, não são verdade e falsidade.

Quais sejam estes dois valores depende dos autores, mas para a maior parte trata-se de 'validade' e 'invalidade' [20].

A primeira coisa que se deve dizer é que validade é um termo sumamente ambiguo. Mas há um sentido que é reiterado por todos os autores: uma norma de obrigação é válida quando se deve fazer o que a norma exige (força obrigatória), e é inválida quando não é caso que se deva fazer o que a norma exige.

Esta condição, que é univoca e clara, responde ao que se chama a condição de Tarski: 'a neve é branca' é verdadeira como enunciado se, e só se, a neve for branca. Nela se apreende a correspondência entre o facto e a predicação que numa proposição se faz do sujeito. "A neve é branca", como enunciado, é verdade se, e só se, o objecto mencionado pelo sujeito da proposição, que é uma realidade objectiva, tiver a qualidade indicada pelo predicado da proposição. E isto é análogo ao que se requer na "força obrigatória" da validade quando se diz que uma norma é válida se, e só se, o obrigado na norma for devido.

Ora bem, se o conceito de validade-invalidade satisfaz a condição de Tarski, então, ainda que se use a *palavra* validade, na realidade está a usar-se o *conceito* de verdade. Uma expressão meta linguística 'E' expressa o conceito de verdade se é verdade que 'E' é predicado de um enunciado quando ocorre o que é descrito pelo enunciado. Ou seja: o que expulsámos pela porta, volta a entrar pela janela (a noção de verdade).

As normas são válidas ou inválidas, mas as condições a que deve satisfazer a validade, pela condição de Tarski, constituem a noção de verdade e deste modo volta-se à alternativa de não permitir a criação do dilema, porque na realidade se está a predicar, das normas, verdade ou falsidade.

Se se usa outra noção para a palavra validade a solução do problema pode variar, mas não vamos abordar isso.

Ainda que todas as tentativas levem a saídas ou muito difíceis ou só aparentes, resta

sempre a possibilidade de um novo recurso da inteligência. Assim, quando tudo parece perdido perante a aparente irredutibilidade do dilema há ainda outra solução que pode ser proposta.

4.6 A Racionalidade do Legislador

Parece pouco plausível, pouco razoável, aceitar que as normas pelo facto de não serem verdadeiras ou falsas excluem a intromissão de problemas como os de consistência.

A solução consiste em se aceitar a condição a-lógica das normas, reconstruindo-se, no entanto, a sua racionalidade por via da do legislador.

Trata-se de expôr as leis lógicas que resultariam verdadeiras se satisfizessem "a racionalidade do legislador".

Resumindo: as leis lógicas não são a descrição, mas sim o reflexo de uma idealidade a que chamamos legislador racional. Um desenvolvimento desta concepção encontra-se em *Norma, Verdade e Lógica* de Von Wright [21]. Segundo este, duas normas são inconsistentes se não suportarem a prova da aceitação de ambas por um legislador racional. Deste modo, as noções centrais da lógica vão além do seu campo clássico de aplicação. Assim entendida, esta posição manifesta a opção pela segunda haste do dilema.

Esta solução, sumamente inteligente e refinada, enfrenta, no entanto, dois problemas de envergadura. O primeiro consiste no facto de que não sendo a noção de consistência predicada das normas por falta de um correlato intuitivo (verdadeiro-falso), a sua substituição pela ideia de legislador racional também não exibe correlatos intuitivos satisfatórios. O segundo é mais técnico e intrasistemático: no melhor dos casos a posição exposta poderia explicar as relações entre enunciados deônticos atómicos, quer dizer enunciados que dizem que algo é obrigatório, que algo é proibido, etc., mas continua a ser impossível obter enunciados deônticos moleculares pois o "e", o "ou" e sobretudo o "não" carecem de explicação, o que impede que possam ser usados com sentido.

Parece inegável que os conectivos lógicos mencionados se usam entre normas e que, além disso, têm um sentido: ora esse sentido há que reconstrui-lo, e isso não pode ser feito sem as condições semânticas que as normas não têm.

O curioso é que, se observarmos o comportamento de tais conectivos *entre as normas*, não se mostra diferente do seu comportamento *nas normas* e, mais em geral, do comportamento que têm em todos os contextos proposicionais. Mas, então o que os conectivos exigem como caracterização do seu sentido não é que as partes que ligam sejam enunciadas como verdade ou falsidade. Exigem, sim, uma explicação do seu comportamento que permita manejá-los em relação com a noção de consequência.

5. A Noção Abstracta, Sintáctica e Semântica de Consequência

Como já indicámos atrás, vamos passar o "Rubicão" da Verdade e aceitar a segunda haste do dilema: sustentaremos que as normas têm uma lógica e que para caracterizar a lógica em geral não é necessário apoiarmo-nos em noções de verdade ou falsidade, bastando partir de uma noção abstracta de consequência, desde que tomemos este termo como primitivo.

Estamos sinceramente convencidos de que se outros autores que trouxeram tantos e tão importantes elementos para esta posição a não enunciaram, isso se deve provavelmente a um velho preconceito filosófico [22].

5.1 A Noção Abstracta de Consequência

O sentido dos conectivos, tal como o dos operadores lógicos, caracteriza-se pelo seu uso em relação a uma noção abstracta de consequência. E esta especifica-se mediante um sistema de axiomas que se reporta à noção intuitiva.

Se há regras que estão ligadas ao sentido de um operador e podem ser dadas de um modo unívoco e preciso, não é necessário que afectem proposições descritivas: alterando as regras de uso alteramos o sentido do operador, mesmo dentro da lógica que se aplica a proposições descritivas, o que mostra que as regras que definem o uso dos conectivos são as que indicam o seu "sentido'', e não a relação com valores de verdade das proposições afectadas.

A tradição da lógica ligada às condições de verdade é tão pertinaz e tem desenvolvimentos tão ricos que resulta muito difícil prescindir dela. Mas não esgota a noção de lógica.

Aceitar a segunda parte do dilema de Jørgensen tem essa consequência. Alterando-se a caracterização de consequência, alteram-se também a da consistência e as definições dos operadores. A alteração importante consiste em mudar entre os primitivos, o primitivo. Partir das noções de verdade e falsidade põe limites insuperáveis; partir da noção de consequência lógica é mais fecundo desde que se facultem as propriedades desta noção, através de um conjunto de axiomas. Estas propriedades encontram-se já enunciadas na noção abstracta de consequência de Tarski, previamente à sua explicitação da condição semântica de consequência [23].

Na concepção tradicional procurou-se na noção semântica de consequência a justificação da noção sintáctica. Mas existe uma versão ainda mais geral que não é semântica nem sintáctica, a partir da qual começa o trabalho de reconstrução e fundamentação lógica.

A noção de consequência é expressa por uma operação que se faz sobre conjuntos de enunciados em qualquer linguagem e que dá como resultado um conjunto de enunciados da linguagem. Por outras palavras: uma função de conjunto de enunciados a conjunto de enunciados. É evidente que se trata de uma abstracção de operações tiradas da prática; o importante é o grau de abstracção e formalização obtido.

Quais são as propriedades destas operações? Em primeiro lugar, que o conjunto de enunciados contido nas premissas esteja contido no conjunto de enunciados das consequências, em segundo lugar, que as consequências das consequências sejam consequências, e em terceiro (simplificando), que embora aumentem as premissas, as consequências que se obtinham com o conjunto mais reduzido se devem manter.

Formalmente:

1. $A \subseteq Cn(A)$ Postulado de inclusão
2. $Cn(A) = Cn(Cn(A))$ Postulado de equipotência
3. Se $A \subseteq B$ então $Cn(A) \subseteq Cn(B)$. Postulado de monotonia [24].

É possível obter outras apresentações da mesma ideia através de uma relação de inferência '$\vdash$' que recolhe toda a rica tradição que se desenvolve em torno dela e preserva as propriedades que enunciámos.

Por exemplo:

1. Se $x \in A$ então $A \vdash x$. Reflexividade.
2. Se $A \vdash y$ e $A \cup \{y\} \vdash x$ então $A \vdash x$. Transitividade.
3. Se $A \vdash y$ então $A \cup \{x\} \vdash y$. Monotonia [25].

Esta noção abstracta de consequência permite ganhos de generalidade em relação às

noções sintáctica e semântica de consequência.

5.2 A Noção Sintáctica de Consequência

Uma definição desta noção de consequência é a definição sintáctica que se pode exprimir como segue: um enunciado RS consequência sintáctica de um certo conjunto quando há uma sequência de enunciados tal que cada enunciado da sequência é um elemento do conjunto em questão, ou é um dos axiomas, ou obtém-se dos anteriores com base em alguma regra de derivação e o último enunciado de sequência é o que se demonstrou. A noção sintáctica de consequência é uma noção de consequência porque satisfaz os postulados da noção abstracta de consequência.

Não se diz que regras vão a ser exactamente aquelas, pois a cada especificação da noção de consequência virá associada uma lógica. Por exemplo: poderá dizer-se que há duas condicionais: uma clássica e outra intuicionista, dependendo de como se definam sintacticamente uma e outra quais as propriedades que admitem (por exemplo a "lei de Pierce"). Tais condicionais são sintacticamente distintas.

A noção sintáctica de inferência pode-se expressar também dizendo que é uma operação definida para cada linguagem em função de regras de dedução e axiomas. Esta caracterização da lógica é totalmente sintáctica.

Para a mesma noção abstracta um operador Cn é simplesmente um operador que aplicado a A identifica um conjunto. Identificar um conjunto é saber quando um objecto pertence ou não a esse conjunto. O que faz a definição sintáctica é identificar quando um enunciado pertence ao conjunto das consequências.

Esta é a especificação sintáctica da noção abstracta.

5.3 A Noção Semântica de Consequência

Anos depois de ter dado a caracterização da noção abstracta de consequência, Tarski deu a definição semântica de consequência [26] ou seja, a definição de implicação lógica: que um enunciado X se obtém de um conjunto A de enunciados como implicado logicamente se com isso se quer dizer que não há nenhuma interpretação (ou seja, nenhuma maneira de adjudicar valores a esses enunciados) tal que todos os enunciados do Conjunto A sejam verdadeiros e X não o seja.

Esta noção de consequência semântica não é finitista, ao contrário da noção de consequência sintáctica apresentada anteriormente. Portanto, é importante mostrar a coextensividade de ambas as noções para cada sistema de lógica. Nisto consiste o problema da completude semântica de um cálculo sintáctico. Convém sublinhar que ambas as noções de consequência, a sintáctica e a semântica, são exemplos da noção abstracta de consequência, posto que ambas satisfazem os axiomas de Tarski, já descritos.

5.4 O Sentido dado por Regras de Uso num Contexto

Chegados a este ponto, podemos ver a noção de consequência como a primitiva e central da lógica que começa a dar sentido a todos os outros operadores e propriedades. Definir um operador lógico quer agora dizer: mostrar as funções que os signos lógicos cumprem na (ou com relação à) noção abstracta de consequência.

É este o modo como Gentzen caracteriza os operadores lógicos, dando as regras de introdução e de eliminação num contexto de dedução. Gentzen caracteriza a sua lógica com

dois tipos de regras: as estruturais, que têm de tipificar a própria noção de inferência, e as regras operativas para os conectivos que são operações lógicas.

Não sendo este o lugar para discorrer sobre as diferenças entre o Wittgenstein do *Tractatus* e o das *Investigações Filosóficas,* o que se pode sustentar é que o segundo Wittgenstein viu com clareza que as operações numa linguagem estão ligadas às funções que nela se realizam e que tais funções são muitas mais do que as de comunicar e informar sobre a realidade [27].

Não só as proposições desprezam uma parte importante da linguagem significativa, como se pode construir uma lógica tomando como primitiva a noção de consequência e definir os operadores do mesmo modo como se pode dar sentido aos enunciados, *dando simplesmente as regras para o seu uso.*

É importante determinar o contexto em que tal operação se realiza, pois a noção de contexto é prévia mesmo ao significado de conectivos e operadores. E é apenas num contexto de consequência (a noção abstracta de Tarski) que faz sentido perguntar pelas regras de uso de um conectivo.

Que significa dar as regras de uso dos signos lógicos num contexto de derivabilidade onde há premissas por um lado e conclusão por outro? Como disse Gentzen, consiste em dar as regras de como introduzir e eliminar os enunciados nos quais figurem esses signos lógicos.

É assim que se procede na lógica de sequências: introduzem-se os signos lógicos no presequente (o que está à esquerda de "$\vdash$") e no prosequente (o que está à direita de "$\vdash$"). Sabendo-se como manejar tais signos no contexto de derivabilidade, fica explicado o sentido dos conectivos de derivabilidade, pois explicar o sentido dos conectivos é dar as regras de uso.

5.5 Que Operadores para a Lógica?

Qualquer signo lógico? Em princípio sim. Qualquer signo. Só que, por razões de prudência, se deve começar por aqueles que têm larga tradição em matéria lógica e aos quais se pode predicar um sistema de propriedades lógicas já maduro. Mas, em princípio, qualquer um.

Por exemplo, pode caracterizar-se um operador deôntico como obrigatório, pois tudo o que se deve indicar é como manejá-lo num cálculo de sequências, sem ter de estar comprometido com os valores de verdade ou falsidade dos enunciados que afectam.

Tarski e Carnap assinalaram a arbitrariedade histórica que significa o limitarmo-nos, em geral, a poucos signos lógicos. A primeira resposta que surge é que tais poucos signos, os conectivos extensionais (digamos "ou", "e", "não", "se...então"), e os operadores extensionais (quantificadores) são os que reconstroem a lógica mais incontestável. O que é certo, mas também é certo que são os que guardam, com maior clareza, a tradição de se supôr como primitiva a noção semântica de consequência.

Nas lógicas modais os problemas que surgem devem-se à concepção semântica. Por esta razão os operadores modais foram contestados, com sortes diferentes. A concepção que caracteriza o seu sentido pelas regras de uso em contextos de derivabilidade elimina estes problemas e permite utilizá-los com os mesmos procedimentos que servem para introduzir os operadores extensionais.

A concepção semântica, que é a mais difundida, assegura o sentido dos compostos a partir do sentido dos componentes, procedendo de tal modo que as condições de verdade dos compostos sejam uma função de verdade dos componentes.

E é esta a ideia central de Prior no artigo citado [28]; a outra possibilidade de associar novos operadores a um cálculo lógico que descreva as regras de uso no contexto em que se aplicam, é, para ele, uma saída–genial–falsa pois é possível provar intuitivamente que um novo operador, por exemplo "tonk", viola as regras mais elementares do raciocinio. A posição semântica de Prior deu lugar, com referência às lógicas modais, às teorias semânticas dos mundos possíveis, das quais nos ocupámos anteriormente.

Prior pensa estar na linha de toda a tradição lógica. Na verdade está em linha só numa parte – se bem que importantíssima – que é aquela que explica e dá sentido ao raciocinio pelas propriedades das partes (verdade ou falsidade) que se mantêm em todo o raciocinio

5.6 Só num Contexto a Parte ganha Sentido

Na lógica e na filosofia existe uma outra vertente que se fundamenta na inteligibilidade da idéia global de raciocinio e explica o sentido das partes, inclusivé dos conectivos, pela função que cumprem num raciocinio. Tal é a resposta de Belnap a Prior [29].

Se à primeira linha se pode chamar 'analítica', a segunda pode chamar–se 'sintética', A exposição que Prior faz desta segunda linha está incompleta porque dá as regras de uso de "tonk" mas não caracteriza completamente o contexto na qual vai ser utilizado, e que é o contexto de consequência.

A conclusão da posição de Prior é que os conectivos não podem ser definidos apenas em termos de deducibilidade; há que ter uma noção prévia do significado do termo, independentemente do papel que desempenha como premissa e como conclusão.

Para poder aplicar a linha sintética representada por autores como Gentzen, Fitch, Popper e o segundo Wittgenstein, devemos interrogar–nos sobre como poderemos definir os conectivos num contexto de derivação sem que se produzam os danos que, por exemplo, "tonk" exibe.

Como disse Belnap, a solução consiste em ter presente que também na metodologia sintética é incorrecto definir os conectivos *ab initio*, pois é necessário dispôr antes de um contexto de deducibilidade a respeito do qual existam definições precisas. Isto resulta evidente no modo como Prior *utiliza* a noção de transitividade da dedução, já que, de contrário, não poderia obter o resultado pretendido [30].

Mas se existem já assunções no contexto de deducibilidade, então será necessário mantê–las presentes afim de evitar que as novas definições que se dêem provoquem inconsistências com as assunções prévias. Deste modo é fácil demonstrar que as regras de introdução do conectivo "e" mantêm consistência com as assunções do contexto de derivabilidade já enunciadas, ao passo que as regras de introdução do operador "tonk" as violam.

Do ponto de vista técnico é facilmente reconstruível, com faz Belnap, uma noção de deducibilidade como se fosse um pequeno sistema formal e que serve para definir o termo '$\vdash$'. A partir das regras estruturais de Gentzen:

Axioma $A \vdash A$

Regras:

Debilitamento: De $A1,\ldots An \vdash$ pode–se inferir $A1,\ldots, An, B \vdash C$

Permuta: De $A1,\ldots, Ai + 1, \ldots, An \vdash B$ pode–se inferir

$A1,\ldots, Ai + 1, Ai, \ldots An \vdash B$

Contracção: De $A1,\ldots An, An \vdash B$ pode–se inferir $A1,\ldots, An \vdash B$

Transitividade: De $A1,\ldots, Am \vdash B$ e $c1,\ldots, Cn,B\ 1 \vdash D$ pode–se inferir $A1,\ldots, Am.\ C1, \ldots Cn \vdash D$

Estas regras determinam completamente o contexto para as definições dos operadores. No momento em que os operadores são caracterizados produz-se uma extensão da noção de deducibilidade. Noção que se pode obter dos axiomas e regras indicadas, de tal modo que as novas regras do novo operador virão a agregar-se às existentes no contexto de derivabilidade. Chamar-se-á *conservativa* à propriedade que se exigirá aos novos axiomas e às novas regras que definirão os conectivos, entendendo por tal a propriedade de manter todos os enunciados já contidos na definição formal do contexto de derivabilidade e não admitindo novos enunciados que não impliquem o novo operador. Não conduzirá a nenhum novo enunciado de deducibilidade A1, ..., An $\vdash$ B que não contenha o novo operador, a menos que tal enunciado seja demonstrável independentemente das regras e dos axiomas do novo operador.

Estes requisitos mantêm, por um lado, a consistência com as assunções do contexto de derivação, por outro, realçam que os axiomas e regras do contexto de derivação são completos e, por último, evitam a reiteração de 'nomes' de operadores que afinal cumprem as mesmas funções porque se servem das mesmas regras de uso (unicidade).

A definição de "tonk" dada por Prior é inconsistente no que respeita às regras do contexto de derivabilidade no qual deve actuar. E dado que as condições de deducibilidade da definição abstracta de consequência dada por Tarski ou por Gentzen, ou enunciada por Belnap, são completas, podemos concluir que um operador como o proposto não pode ser admitido, enquanto é fácil demonstrar que as regras de introdução de um conectivo como "e" são *conservativas* no sentido de serem consistentes com as assunções do contexto de derivabilidade [31].

Deste modo pode demonstrar-se que é possível definir-se qualquer conectivo em termos de consequência, mas que isso comporta o ónus da prova da consistência do novo conectivo para com as assunções do contexto no qual se quer introduzi-lo. Esta é uma *conditio sine qua non;* a *conditio per quam* consiste em definir completamente as regras de introdução de tal operador no contexto, conforme as regras de uso. Como última condição, deve-se provar que não existe outro conectivo que já tenha as mesmas propriedades formais operativas.

6. A Lógica Deôntica

Tendo-se resolvido o dilema no sentido de apoiar a segunda parte, precisamente porque a lógica não se refere em exclusivo aos enunciados que são verdadeiros ou falsos, mas a todos aqueles aos quais se conseguem dar regras precisas sobre o uso dentro de um contexto, fica aberto o campo da lógica deôntica.

Neste ponto conviria facultar algum exemplo de sistemas de semelhante lógica.

Como é sabido os exemplos abundam. Mas, embora isto possa ser considerado filosoficamente escandaloso, não o é, em absoluto, nem para a lógica, nem muito menos para o direito ou para a informática, que devem servir-se de tais exemplos.

Em primeiro lugar, porque não é a única parte da lógica onde os sistemas se multiplicam. Não é demasiado importante que existam várias apresentações mas que tais apresentações sejam, em definitivo, expressões de um núcleo fundamental — algo a que von Wright chamou "a lógica deôntica Standard" — e que respondam aos critérios canónicos expressados anteriormente. Além disso, porque tal proliferação não deve paralisar os utilizadores, mas pelo contrário torná-los mais exigentes, de modo a saberem quais dentre as apresentações são as mais adequadas e em particular para que contexto normativo.

Provavelmente a *última ratio* na matéria será verificar nos contextos pragmáticos quais são as regras que melhor se ajustam às praxis existentes.

Os signos lógicos e os operadores deônticos devem ser incorporados com o mesmo critério que serve para dar as regras de os introduzir e eliminar num contexto de dedução.

Qual seja o seu sentido, pode ser mostrado no contexto de dedução mediante o uso e o funcionamento. A lógica deôntica aparece então como a prestadora desse tipo de regras. O que é relativamente simples pois toda a lógica deôntica pode ser introduzida numa apresentação sequencial.

A título de exemplo: Em que consistiria a introdução dos operadores deônticos do sistema Standard nos presequentes e prosequentes de uma lógica de sequências? Na admissão como regra operativa (à maneira de Gentzen) de:

$$\frac{A1,\ldots,An \vdash B}{OA1,\ldots,OAn \vdash OB}$$

onde A1,... An é um conjunto qualquer (pode ser vazio) e onde B é um enunciado ou uma sequência vazia, mas nunca pode ser uma sequência com mais de um enunciado. Além disso, dado que B é um enunciado, pode-se negá-lo e, em geral, realizar com ele todas as operações possíveis de realizar com enunciados.

O que a regra introduz poderia verbalizar-se dizendo que o que é implicado por um conjunto de deveres, é um dever.

Note-se que o operador deôntico 'O' tem, porque afecta um só enunciado, características formais muito parecidas com as da negação. Os outros operadores podem ser definidos com base em 'O', como por exemplo "Pp =df -O-p" e "Vp =df O-p".

Para obter resultados mais próximos da praxis jurídica, por exemplo, há que estabelecer aparelhos mais sofisticados, sendo necessário introduzir novos operadores, com observância dos cuidados aqui descritos.

7. Por que razão a solução do Dilema interessa a toda a Lógica

A solução do dilema não só permite recuperar para a lógica expressões que desde Aristóteles tinham sido afastadas, v.g. as prescrições, assim como, devido à mudança na ordem dos primitivos, consente que se parta da noção abstracta de consequência.

Passa a ficar mais claro o entendimento de que a negação intuicionista, por exemplo, e a negação clássica têm propriedades diferentes e, por isso, não são duas teorias sobre a negação, mas dois operadores diferentes com sentidos diferentes, não competitivos entre si e que, para completar o seu sentido, devemos conhecer exactamente todo o contexto de derivabilidade. O mesmo se pode dizer das condicionais.

Interpretar, como já indicámos, significará indicar as regras para introduzir e eliminar um operador num contexto de derivação. Isto é válido para todos os operadores clássicos, válido também para o novo operador 'O' e deve sê-lo ainda para outros operadores aos quais se podem dar as condições de introdução e de eliminação explicitadas anteriormente.

A lógica que interessa é a lógica da qual se podem dar, univocamente, as regras de introdução e eliminação dos operadores e que efectivamente aconteça.

Mas que efectivamente aconteça, é um facto empírico.

Entre os discursos que o preconceito filosófico deixou de lado estão, sem dúvida, as noções deônticas, mas o mesmo se pode ter passado com outras, como as noções de preferência.

O problema não consiste em determinar entre os diferentes sistemas qual "a verdadeira

lógica", mas em examinar – como disse Carnap –: "as propriedades formais e as possibilidades de interpretação e aplicação na ciência" [32].

8. Consequências para Informáticos

Do que atrás foi dito o mais importante para os informáticos foi o ter-se mostrado uma forma de apresentação da lógica deôntica e que esta se pode construir, com os mesmos critérios de quaisquer outras lógicas de sequências. Além disso, a lógica deôntica apresentada é decidível.

Quanto mais abstractas e sintácticas forem as regras lógicas mais fácil será fazê-las executar por um computador.

O preconceito filósofo da prioridade semântica verifica-se também no mundo informático. Mesmo aqui, onde se trabalha simplesmente com enunciados dentro de um contexto de derivação, muitos continuam a pensar que na realidade se "usa" verdade e falsidade [33].

Cremos que a reformulação abstracta das noções centrais da lógica em lugar da concepção semântica de Tarski não pode senão favorecer os desenvolvimentos informáticos já que não se podem transmitir às máquinas noções semânticas mas apenas o seu correlato sintáctico. Se a compatibilidade, a coerência e outras relações lógicas forem introduzidas, assim como os operadores, de modo totalmente sintáctico, o cálculo automático fica beneficiado, ampliando-se o seu horizonte.

Certamente, se o que foi dito está correcto, produz-se como consequência o abrir de novos campos ao cálculo lógico, a possibilidade de estabelecer critérios de calculabilidade e decisão mecânica por exemplo na deôntica, nas preferências e nos valores. Isto merece ser observado com atenção pelos informáticos aos quais se amplia o campo de aplicação das suas realizações, dado que tudo o que seja calculável e mecanicamente decidível, deve – em princípio – poder ser realizado por um computador.

Um problema diferente é o grau de complexidade que a execução concreta de tais regras significa para um computador. Além da noção de "decidibilidade", que consideramos crucial, pode agregar-se uma condição de eficiência, ligada, em princípio, aos progressos da computação e em particular da programação lógica. Mas este é tema a tratar noutra ocasião.

(Trad. do espanhol por Mª Helena Vilalva; revisão de J. Lopes Alves e J. Sousa e Brito).

Notas

(1) J. Jørgensen, *Imperatives and Logic*, "Erkenntnis", 7, 1937/38. A. Ross, *Direction and Norms*, Routlegde & Kegan Paul, London, 1967, VI cap. 30. Nesta obra Alf Ross atribui-se a paternidade da expressão "dilema de Jørgensen" no seu artigo *Imperatives and Logic*, "Theoria", p. 53 e seg..

(2) E. Mally, *Grundgesetze des Sollens. Elemente der Logic des Willens*, Leuschner & Lubensky, Graz, 1926.

(3) G. Kalinowski, *Introducción a la lógica jurídica*, Eudeba, Buenos Aires, 1973, p. 71.

(4) G.H. von Wright, em *Norm and Action*, Tecnos, Madrid; A. G. Conte, *Studio per una teoria della validità*, 1980, p. 382 e seg..

(5) H. Kelsen, *Teoria pura del derecho*, Eudeba, Buenos Aires.

(6) H. Neri Castaneda, *Thinking and Doing: The Philosophical Foundations of Institutions*, Reidel, Dordrecht, 1975, sustenta que as normas são verdadeiras ou falsas mas que os factos

que falseiam ou tornam verdadeiras as normas são a legitimidade ou ilegitimidade das prescrições nelas contidas.

(7) Hao Wang, *From Mathematics to Philosophy*, Routledge Kegan Paul, 1974, p. 8.

(8) Aristóteles, *Organon, Da expressão ou interpretação*, cap. 4, 17a: "Ora bem, conquanto toda a sentença ou juizo tenha significado, embora não como um instrumento da natureza mas, como temos observado, por convenção, nem todos podem chamar-se proposições. Chamamos somente proposições às que têm em si verdade ou falsidade. Uma suplica é, por exemplo, uma sentença ou expressão, mas não tem nem verdade nem falsidade. Passemos adiante, pois o seu estudo caberá melhor ao campo da retórica ou da poética. Teremos somente como tema da nossa investigação actual as proposições" (da Edição Aguilar, Madrid, 1964, p. 259/60).

(9) A. Tarski, *Logic, Semantics and Metamathematics*, Oxford, 1956, VIII, The Concept of Truth in Formalized Language, vê com toda a clareza que somente se está a ocupar de enunciados declarativos. Ver também *The Semantic Conception of Truth*, "Philosophy and Phenomenological Research", 4, 1944, p. 341 e seg..

(10) L. Wittgenstein, *Tractatus logico-philosophicus*, Routledge & Kegan Paul, Ltd. London, 1922, tradução espanhola de J. Munoz e I. Reguera, Alianza Editorial, Madrid, 1987; "2.2. A figura tem em comum com o figurado a forma lógica da configuração. 2.201. A figura figura a realidade na medida em que representa uma possibilidade de se dar ou não o estado das coisas. 2.21. A figura concorda com a realidade; é correcta ou incorrecta, verdadeira ou falsa."

(11) Ibidem, 4.31.

(12) R. Carnap, *Foundations of Logic and Mathematics*, The University of Chicago Press, Chicago, 1939, cap. 2º., 12 "O resultado da nossa discussão é o seguinte: a lógica, ou seja as regras da dedução (na nossa terminologia, as regras sintácticas de transformação), podem-se escolher arbitrariamente, e portanto são convencionais sempre que sejam assumidas como base para a construção do sistema linguístico e se a interpretação do sistema for dada num segundo tempo. Se, ao contrário, for dada precedentemente uma interpretação dos signos lógicos o sistema de lógica não é susceptível de escolha, mas ajustado ou equivocado. Porém também neste caso as convenções têm uma importância fundamental, porque a base sobre a qual está edificada a lógica, quer dizer, a interpretação dos signos lógicos (mediante, por exemplo, uma determinação das condições de verdade) pode ser livremente escolhida".

(13) A.N. Prior, *The Runabout Inference-Ticket*, "Analysis" vol. 21, (Blackwell, 1960), pp. 38-9.

(14) G. Gentzen, *Untersuchungen über das logische Schliessen*, "Mathematische Zeitschrift", 1934, t. 39 p. 176 e seg..

(15) A. J. Ayer, *Language, Truth and Logic*, London, 1936.

(16) H. Kelsen, *Allgemeine Theorie der Normen*, Manzsche Verlags und Universitätsbuchhandlung, Wien 1979, capítulo LX, onde acrescenta, de modo mais ou menos claro, que às normas são aplicáveis outros príncipios lógicos como, por exemplo, a subsunção do particular no geral.

(17) M. Hare, *The Language of Morals*, OUP, London, 1952.

(18) G.H. von Wright, *Norms, Truth, and Logic. Practical Reason*, Basil Blackwell, London 1983, p. 131.

(19) D. Lewis, *Confractuals*, cap. Possible worlds.

(20) Assim, por exemplo, em certos escritos de Kelsen.

(21) G.H. von Wright, *Norms, Truth and Logic*, cit..

(22) G.H. von Wright, afirmou mais de uma vez que a lógica vai além da verdade e da falsidade; F. Miró Quesada afirmou que o problema da lógica não é de verdade ou falsidade mas de hereditariedades de qualquer valor transmitidas através da noção de consequência. Tarski, Jaskowski, Gentzen, Wittgenstein e Belnap, de modo que procuraremos aqui ilustrar de forma breve, se não o expressaram nos nossos termos, criaram todas as condições e critérios para o explicarem.

(23) Costuma dizer-se que a noção semântica de consequência enunciada por Tarski era já de Gödel e que este, por sua vez, a tinha tomado de Skolem. Por isso alguns colocam toda a lógica moderna com início em Skolem.

(24) Em Tarski apresenta-se como 1. A $\subseteq$ Cn (A), 2. Cn (A) = Cn (Cn (A)), 3. Cn (A) = U {Cn (B): B$\subseteq$ A & B é finito}. Esta última condição permite obter: 3.1. Se A$\subseteq$B então Cn (A) $\subseteq$ Cn(B) (monotonia) e 3.2. se x $\in$ Cn(A) então há um B$\subseteq$A tal que B é finito e x $\in$ Cn(B) (compacticidade). A. Tarski, *Logic, Semantics and Metamathematics*, citado, cap. III. Foundamental Concepts of Metamathematics, p. 30 e cap. V, Foundamental Concepts of the Methodology of the Deductive Sciences, p. 60.

(25) Esta é uma reconstrução da mesma noção abstracta de 'consequência', só que em Tarski consistia numa *operação* e agora é apresentada como uma *relação*, na concepção de Gentzen, *Untersuchugen...* cit..

(26) A. Tarski, *Logic, Semantics and Metamathematics*, citado, XVI, On the Concept of the Logic Consequence, p. 409.

(27) L. Wittgenstein, *Philosophische Untersuchungen*, Basil Blackwell, Oxford 1953, "23. Wieviele Arten der Sätze gibt es aber? Etwa Behauptung, Frage und Befehl? – Es gibt unzählige solcher Arten: unzählige verschiedene Arten der Verwendung alles dessen, was wir "Zeichen", "Worte", "Sätze", nennen. Und diese Mannigfaltigkeit ist nichts Festes, ein für allemal Gegebenes; sondern neue Typen der Sprache, neue Sprachspiele, wie wir sagen können, entstehen und andere veralten und werden vergessen. (Ein ungegähres Bild davon können uns die Wandlungen der Mathematik geben). Das Wort "Sprachspiel" soll hier hervorheben, dass das Sprechen der Sprache ein Teil ist einer Tätigkeit, oder einer Lebensform".

"Mas quantos géneros de orações há? Asserções, perguntas, ordens? – Há inúmeros géneros: inumeros géneros diferentes de emprego de tudo a que chamamos "signos", "palavras", "orações". E esta multiplicidade não é fixa, algo dado uma vez por todas: pois novos tipos de linguagem, novos jogos de linguagem, por assim dizer, nascem e outros envelhecem e são esquecidos... A expressão "Jogo da linguagem" deve sublinhar aqui que falar uma linguagem faz parte de uma actividade ou de uma forma de vida". E no final da proposição 22: "se ouço alguém dizer "chove", mas não sei se ouvi o início e o final do período, essa oração não é para mim um meio de compreensão", (da tradução castelhana de A. Garcia Suárez e U. Moulines, Editorial Crítica, Barcelona, 1988, p. 39).

(28) V. nota 12.

(29) N.D. Belnap, *Tonk, Plonk and Plink*, "Analysis", vol. 22 (Blackwell, 1962) pag. 130–4 e reproduzido, como o original de Prior citado, em P. Strawson (organizador), *Philosophical Logic*, Oxford University Press, 1965.

(30) Ibidem.

(31) É a mesma razão, como bem exemplifica Belnap, que levou Peano a sustentar que a introdução do operador '?' era incorrecta para os números racionais, dado que das regras de uso podia derivar-se que 2/3=3/5.

(32) R. Carnap, *Foundations of Logic and Mathematics*, 1939, citado, Capítulo segundo *in fine*, dizendo imediatamente antes: "É importante atender às partes convencionais na construção de um sistema linguístico. Tal concepção leva a um estudo despreconceituado das várias formas de novos sistemas lógicos que diferem mais ou menos da forma canónica (p. ex. a lógica intuicionista construída por Brouwer e Heyting, os sistemas de lógica modal construídos por Lewis e outros, os sistemas de lógica plurivalente construídos por Lukasiewicz e Tarski, etc.) e dá alento à construção de outras formas novas".

(33) Repare-se como, por exemplo, o algoritmo de Wang é apresentado em termos de Verdade e Falsidade, quando na realidade trabalha sequências. P. ex. M.L. Schagrin, W.J. Rapaport e R.R. Dipert, *Logic: a Computer Approach*, 1985, McGraw-Hill, cap. 6.

Ruy J.G.B. de Queiroz

Dept. Computing, Imperial College
Univ. of London, London (UK)

Meaning, Use, Consequences - Interconnected Concepts

Abstract

Verificationist theories of meaning based on assertability conditions, *while escaping from the denotational view of language, must remain essentially extensional. Wittgenstein's 'theory of meaning', which can be said to be verificationist but not* assertability conditions-*based, is neither denotational nor extensional given the strong intentionality element inherent to the connections between the meaning/use of a sentence and the consequences one is allowed to draw from it.*

Resumo

As teorias verificacionistas do sentido baseadas em condições de asserção, *embora se furtem ao ponto de vista denotacional da linguagem, permanecem essencialmente extensionais. A "teoria do sentido" de Wittgenstein, que pode ser qualificada como verificacionista mas não baseada em* condições de asserção, *não é denotacional nem extensional dado o forte elemento intencional inerente às conexões entre sentido/uso das proposições e as consequências que daí advêm.*

1. Motivation

When explaining the meaning of a sentence, one can hardly avoid referring to the context under which the sentence is being considered. In his use-based semantic teachings, L. Wittgenstein consistently insists on the crucial rôle played by the context of utterances, and that is very clear from the very conceptual basis provided by the notion of 'language-games'. Whenever it is relevant to ask what a certain sentence really means, Wittgenstein reminds the reader of the crucial part played by the context under which one is asserting the sentence:

> We ask "What does 'I am frightened' really mean, what am I referring to when I say it?" And of course we find no answer, or one that is inadequate.
> The question is: "in what sort of context does it occur?"
>
> (*Investigations*, Part II, section ix, p. 188[e].)

Nevertheless, he does not seem to neglect the intentional ingredient which is inherent to a conception of meaning based on its identification with use. By attempting to demonstrate that the consequences one can draw from a proposition constitute one of the key ingredients of Wittgenstein's definition of meaning as use, we endeavour to show that it is not unreasonable to say (with Wittgenstein) that the meaning of a sentence depends on but is not *ultimately* given by the circumstances under which it can be correctly asserted. Taking the explanation of the what consequences one is allowed to draw from a proposition as its ultimate intentional aspect, as an explanation of how the proposition functions in the

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 151/166
1992, APDC & SPF, Lisboa

circumstances, one remains congenial with Wittgenstein's concerns as to the semantical rôle of the consequences, and yet taking into account the important part played by the circumstances. A clear evidence of Wittgenstein's concern with the functionality aspect of the meaning of signs is present in Prof. P. Geach's notes of Wittgenstein's *Lectures on Philosophical Psychology 1946–47*:

> *Malcolm*: You say "a voluntary action" gets its sense from circumstances. And you contrast 'voluntary action' with 'feeling' as different concepts. Therefore one would expect that "a pain" should *not* get its sense from circumstances; whereas *you* say that "a pain" *also* gets its sense from circumstances. So where is the contrast between, e.g. pain and voluntary action?
>
> *Wittgenstein*: ***All signals get their sense from circumstances, it is true; but the way they function in the circumstances is different***. Thus, impressions have a place and emotions have not; and to say an action is voluntary is not like uttering an impression *or* an emotion.
>
> (Geach (ed.) 1988, p. 76.)

(The ***emphasis*** is ours.)

Despite constituting a necessary ingredient in a semantic analysis, a description of the context of utterance must remain in the extensional territory, whereas the clarification of the consequences that can be drawn from the sentence appears to play the ultimate (intentional) rôle in the meaning explanation. It seems equally reasonable to say that, e.g., no matter how accurate the description of a context under which one can correctly assert a sentence may be, one will not make its meaning clear until the aspect of intentionality, which can hardly be said to be part of the context, is uncovered.

> "I know that that's a tree." Why does it strike me as if I did not understand the sentence? though it is after all an extremely simple sentence of the most ordinary kind? It is as if I could not focus my mind on any meaning. Simply because I don't look for the focus where the meaning is. As soon as I think of an everyday use of the sentence instead of a philosophical one, its meaning becomes clear and ordinary.
>
> Just as the words "I am here" have a meaning only in certain contexts, and not when I say them to someone who is sitting in front of me and sees me clearly,–and not because they are superfluous, but because ***their meaning is not determined by the situation, yet stands in need of such determination.***
>
> (*On Certainty*, 347–8, p. 44[e].)

(The ***emphasis*** is ours.)

Attempts have been made to develop semantic theories based on states of affairs, situations, etc., where the basic ingredients include more complex structures than the ones used in model–theoretic truth–values based semantic theories. Such ingredients in most cases find their ontological status justified in the framework of possible world's–semantics perhaps first introduced by Prof. S. Kripke's semantical studies of modalities. The important lesson that we seem to be offered by Wittgenstein's investigations into the connections meaning–use and the separation of the intentional (functional) aspect from the aspect concerning the context (state of affair, circumstance, etc.) is that more than simply devices to describe contexts must be incorporated in any workable semantic theory. And this seems to be justified by Wittgenstein in the form of a warning against our thinking that it must be possible to find different states of affairs corresponding to meaningful sentences which apparently have the same truth–conditions but have different meanings.

> (...) In any case the true state of affairs can only be seen when we look into the detail of the usage of our expressions. And it is clear that *if in our present case we should decide to say that the two expressions have different meanings, we shall certainly not*

> *be able to say that the difference is that the fact which makes the second sentence true is a different one from the fact which makes the first sentence true.*
> We are justified in saying that the sentence "He can continue..." has a different meaning from this: "He knows the formula". *But we mustn't imagine that we can find a particular state of affairs 'which the first sentence refers to'*, as it were on a plane above that on which the special occurrences (like knowing the formula, imagining certain further terms, etc.) take place
>
> (*Brown Book*, Part I. 64, p. 115)

(The *emphasis* is ours)

Based on a distinction between sense and force, perhaps first made by Frege, and on the connections between meaning and proofs, some verificationist theories of meaning often take the *assertability conditions* as the central aspect which determines the meaning of a proposition. It seems to have been Wittgenstein who first insisted on the rôle of *verification* in the settlement of sound foundations of objective knowledge. Incidentally, he also referred to the *sense* of a proposition as being determined by the way it can be verified. However, Wittgenstein's point of view on the rôle of *verification* for the particular case of logical propositions seems to have been much different from the general 'dogma' of the so-called 'verificationist theories of meaning', simply because it appears to prescribe that the process of *verification* should take primarily into account the *conclusions* one is allowed to draw from the proposition, or simply, the *intention*, the *utility* of making the assertion. And this seems to suggest the fundamental rôle of the *consequences* one can draw from a particular assertion as a key aspect of its meaning, in addition to the equally as important *assertability conditions* taken to be the sole aspect of the meaning of an assertion by most of the 'verificationist theories of meaning'. Prof. Prawitz's semantical investigations of a conceptual basis for a theory of meaning for the language of mathematics based on the connections between meaning and proofs clearly acknowledges that the two aspects of a proposition, namely sense and force (following Frege and Dummett), are two independent aspects, but concludes that the central aspect of its meaning must reside in the sense aspect (conditions for assertion):

> 4. *Does the condition for asserting A constitute the meaning of A?* The following two facts are relevant for this question:
> (i) From the rules governing when a sentence is correctly asserted, one can derive the condition which a situation *S* has to satisfy, in order that a given inference is to be acceptable in *S*, viz., that a procedure is known in *S* which applied to canonical proofs of the premisses yield a canonical proof of the conclusion. This is the harmony required by Dummett to hold between the conditions for asserting a sentence and the commitments made by asserting it.
> (ii) From an assumption that a person knows the condition for asserting a given sentence, it does not follow that he knows the acceptability of any particular inference involving those inferences that correspond to the conditions for asserting the sentences (introduction inferences); he knows the criterion for accepting an inference, but not necessarily whether the criterion is satisfied – in particular, ***he does not need to know what is usually considered to be the immediate commitments made by asserting the sentences in question.***
> ***(i) and (ii) show to what extent the use of language can be derived in a meaning theory which identifies the meaning of a sentence with the condition for asserting it.***
>
> (Prawitz 1977, pp. 38–9.)

(The ***emphasis*** is ours.)

A theory of meaning constructed on the basis of *assertability conditions* as the central

aspect of the meaning of a sentence can hardly be said to incorporate any intentional ingredient. As it is clearly acknowledged in the passage above, by knowing the conditions for asserting a particular sentence one can hardly know the commitments made by asserting it, and it is this very aspect of 'commitments' which plays a central semantic rôle in what Prof. Dummett calls a 'falsificationist' theory of meaning (Dummett 1976, p. 126). And 'reflection should make us admit' that a falsificationist theory of meaning would be better than a verificationist *assertability conditions*-based theory of meaning, according to Prof. Dummett (Ibid.,p. 137). Now, as we have pointed out in de Queiroz 1989a, it is reasonable to suppose that the commitments one makes by asserting a sentence are inseparable from the consequences he/she allows to be taken from it.

On the other hand, the element of intentionality is clearly present in the connections between meaning and utility which Wittgenstein makes more explicit than anywhere else in the remark 'Meaning, function, purpose, usefulness – interconnected concepts', 291, p. 41[e] of the *Last Writings on the Philosophy of Psychology,* vol. I. In this respect, it should perhaps be sufficient to recall that the conception of 'words as tools' discussed in the *Investigations* (Part I, 11), is already present in the *Blue Book*:

> Think of words as instruments characterized by their use, and then think of the use of a hammer, the use of a chisel, the use of a square, of a glue pot, and of the glue.
>
> (*Blue Book*, p. 67.)

It appears to be in the same *Blue Book* that the crucial rôle played by the context of utterances in a semantical theory is mostly emphasised over and over again. In some passages, however, the intentionality aspect is also brought into the focus of attention, although the general picture still reflects a clear emphasis on the importance of looking at the circumstances under which the utterances are made.

Wittgenstein's view on the insufficiency of context descriptions in making meaning explicit takes a great variety of forms. In the *Investigations,* where the use of the device of saying 'the question is: under what circumstances does one assert the sentence?' is often, if emphatic, he is already trying to warn us against thinking that there can be a totality of conditions (state of affairs) which gives an utterance its meaning:

> But did "Now I can go on" in case (151) mean the same as "Now the formula has occurred to me" or something different? We may say that, in those circumstances, the two sentences have the same sense, achieve the same thing. But also that *in general* these two sentences do not have the same sense. We so say: "Now I can go on, I mean I know the formula", as we say "I can walk, I mean I have time"; but also "I can walk, I mean I am already strong enough"; or: "I can walk, as far as the state of my legs is concerned", that is, when we are contrasting *this* condition for walking with others. ***But here we must be on our guard against thinking that there is some totality of conditions corresponding to the nature of each case (e.g. for a person's walking) so that, as it were, he could not but walk if they were all fulfilled.***
>
> (*Investigations*, Part I, 183, p.74[e].)

(The ***emphasis*** is ours.)

The intentionality aspect is not ultimately uncovered by an increase on the amount of information about the context. And this is what characterises the intention as the final and ultimate interpretation. In a given social pratice, neither need one to define for each sentence what can be deduced from it, not need the hearer to guess, the right conclusions are readily taken due to the ability that both have of playing the language-game. Wittgenstein comes to this specific point still at Part I of the *Investigations*:

> "I was going to say. .."–***You remember various details. But not even all of them together shew your intention***. It is as if a snapshot of a scene had been taken, but only a few

> scattered details of it were to be seen: here a hand, there a bit of a face, or a hat–the rest is dark. And now it is as if we knew quite certainly what the whole picture represented. As if I could read the darkness.
> These 'details' are not irrelevant in the sense in which other circumstances which I can remember equally well are irrelevant. But if I tell someone "For a moment I was going to say....." ***he does not learn those details from this, nor need he guess them. He need not know, for instance, that I had already opened my mouth to speak. But he can 'fill out the picture' in this way. (And this capacity is part of understanding what I tell him.)***
> (*Investigations*, Part I, 635–6. pp. 164^{e}–5^{e}.)

(The ***emphasis*** is ours.)

In a language–game and social pratice, intentions and expectations participate in a 'give–and–take': an utterance and what will expectedly be drawn from it are linked by the very social practices. When hearing an utterance as part of a certain language–game one does not need to be given a description of details of the circumstance under which it is asserted, nor need he/she guess it. That he/she can fill out the picture in such and such away is part of understanding what is being said.

2. Intentionality and the Consequences

For a statement to have a meaning at all, it needs to allow certain conclusions to be made from it, and the intencionality aspect would be determining which is to be taken as the ultimate and intended one.

> "I have already remembered three times today that I must write to him" Of what importance is it what went on in me then?–On the other hand what is the importance, what the interest, of the statement itself?–*It permits certain conclusions*.
> (*Investigations*, Part II, section xi, pp. 217–8^{e}.)

(The *emphasis* is ours.)

In search of a logical account of the 'meaning of signs', and as a substitute for the usual semantical device of 'interpretation' Wittgenstein finds in the (apparently psychological) notion of 'intention' a strong candidate to do the job. Having remarked that 'any interpretation still hangs in the air along with what it interprets and cannot give it any support', and 'interpretations by themselves do not determine meaning'(*Investigations*, Part I, 198), he could find in the concept of 'intention' the final point in the endless chain of interpretations. The apparent disappointment was that 'intention' would not be a logical endpoint but rather a psychological one:

> By "intention" I mean here what uses a sign in a thought. The intention seems to interpret, to give the final interpretation; which is not a further sign or picture, but something else–the thing that cannot be further interpreted. But what we have reached is a psychological, not a logical terminus.
> (*Zettel*, 231, p. 41.)

In many other opportunities, however, Wittgenstein does consider relevant to demonstrate that the notion of intention is perhaps not as 'psychological' as one is usually led to think. For example, in the *Brown Book*, 'meaning' and 'intending' refer to 'certain acts':

> We use the words "meaning", "intending" in such a way that they refer to ***certain acts*** states of mind given certain circumstances; as by the expression "checkmating somebody" we refer to the ***act*** of taking his king.(...)
> (*Brown Book*, Part II, p. 147.)

(The ***emphasis*** is ours.)

And in the *Remarks on the Philosophy of Psychology*, *intention* is neither a feeling nor an experience.

> Intent, intention, is neither an emotion, a mood, nor yet a sensation or image. It is not a state of consciousness. It does not have genuine duration. Intention can be called a mental disposition. This term is misleading inasmuch as one does not perceive such a disposition within himself as a matter of experience. The *inclination* toward jealousy, on the other hand, is a disposition in the true sense. Experience teaches me that I have it.
>
> (*RPP II*, 178, p.34e.)

Now, if we look at how the words meaning and intention are actually related in ordinary language, we can see that the act of explaining the consequences one is allowed to draw from a sentence does play the rôle of a logical counterpart to that 'psychological' terminus. Cf. 'what do you mean?' with 'what do you want to say?', 'what message do you intend to get across by saying that?', etc.. As it happens, the connections between the meaning of a proposition and the consequences one can draw from it seem to be in Wittgenstein's own conception of meaning throughout most of his writings. We have recently attemtped to demonstrate that by drawing a triangle-like semantical diagram with vertices in meaning, use and consequences, we can still fill in the sides corresponding to meaning-consequences and use-consequences which are made much clearer and much more incisive in his very late writings, be it from his own hand or in the form of lecture notes taken by someone else, namely the *Remarks on the Philosophy of Psychology (I & II)*, the *Last Writings on the Philosophy of Psychology I, On Certainty* and the *Lectures on Philosophical Psychology 1946-47* (de Queiroz 1989b). A great number of examples of passages conveying that particular aspect of later Wittgenstein can be given as, e.g.:

> "As I heard that word he occurred to me. "What is the practical significance of this point of time?-. For I want to say; "*It seemed to me* that he occurred to me at that word" -and the subjective element of this statement makes no difference. The question still remains the same: ***"what consequences does such a report have?"***
>
> (*Last Writings I*, 100, p. 16e.)

(The *emphasis* is ours.)

> What does someone who says "Now I see it as ..." convey to us? That is, *what follows from this report, what sort of use does it have? It could have many different kinds of consequences.*
>
> (*Last Writings I*, 622. p. 79e.)

(The *emphasis* is ours.)

In order to understand what is meant by a certain utterance one considers the situation in which the utterance is made. It seems unwarranted, however, that in the general case a description of the situation (without mentioning the intention, the purpose of the utterance) will provide enough information to clarify the meaning of the utterance. Most certainly, any attempt to understand the meaning of sentences without due attention to the circumstances under which they can be used, is likely to be incomplete, if not nonsensical. The meaning of a sentence being as it is inseparable from the situation in which it is placed, a description of the latter would not ultimately determine the former. It does not appear to be too convincing to say that by enriching the description of the situation one will uncover the meaning, the intention, the purpose of a certain word or sentence.

3. Logical Calculi and the Rôle of Consequences in the Early Wittgenstein

The intuition regarding the connections between meaning and use (as utility), which

appears explicitly in very late remarks such as the one shown in the opening lines as well as in the title of this paper, and underlies the 'words are tools' conceptual link made mainly in Part I of the *Investigations*, finds its roots in very early remarks documented in early writings such as the *Notebooks* and the *Tractatus*.

> The requirement of determinateness could also be formulated in the following way: if a proposition is to have sense, the syntactical employment of each of its parts must have been established in advance. For example, it cannot occur to one only subsequently that a certain proposition follows from it. *Before a proposition can have a sense, it must be completely settled what propositions follow from it.*
>
> (*Prototractatus*, 3.20103, p. 65.)

(The *emphasis* is ours.)

And this is almost verbatim another remark which can be found in the pre-*Tractatus Notebooks 1914–16*:

> We might demand definiteness in this way too: if a proposition is to make sense then the syntactical employment of each of its parts must be settled in advance. It is, e.g., not possible *only subsequently to come upon* the fact that a proposition follows from it. But, e.g., ***what propositions follow from a proposition must be completely settled before that proposition can have a sense!***
>
> (*Notebooks*, dated 18.6.15, p. 64[e].)

(The ***emphasis*** is ours.)

One can find remarks in the *Tractatus* which reflect rather clearly how early the connections between meaning and use were present in Wittgenstein's understanding of semantical issues.

> The way in which language signifies is mirrored in its use.
>
> (*Notebooks*, dated 11.9.16, p. 82[e].)
>
> In order to reconize a symbol by its sign we must observe how it is used with a sense.
>
> (*Tractatus*, 3.326, p. 16.)
>
> (In philosophy the question, 'what do we actually use this word or this proposition for?' repeatedly leads to valuable insights.)
>
> (*Tractatus*, 6.211, p. 65.)

4. Logic and Semantics

From the *Tractatus* to the *Investigations* Wittgenstein replaces his account of propositions as 'a proposition is a truth-function of elementary propositions' (*Tractatus*, 5), by 'and what a proposition is is in one sense determined by the rules of sentence formation, and in another sense by the use of the sign in the language-game' (*Investigations*, 136). Now, important questions arise from such a change in the 'definition' of what a proposition really is. What rules of logical inference can best represent the two aspects of Wittgenstein's later account of propositions? How does this particular change fit in with his shift of the semantical emphasis from truth-values to *use*? Furthermore, where is the emphasis of traditional logical systems? In Prof. Geach's *Wittgenstein's Lectures on Philosophical Psychology 1946–47* we can perhaps find an answer to the latter:

> Linguistic devices are like the handles of a machine. They are all made to fit the hand, but what they do and what you have to do with them are different. One has to be pulled hard, one can be adjusted continuously, one will take only two positions.
>
> *The tradicional logic is concerned with the handles, not with what they do.* E.g. the copula doesn't matter. In German we say "es blaut", "it's blue", there might be a language where all colour words are verbs.

(Geach (ed.) 1988, p. 25.)

(The *emphasis* is ours.)

From his (early) view of language as a calculus and calculus as a game, which is shown clearly in Waismann's *Ludwig Wittgenstein and the Vienna Circle,* we learn that by looking at rules of deduction as rules of a game establishing a frame of reference, we can see why the rules of *elimination*, so to speak, convey the intentional, functional aspect of a logical sign. The 'rules of the game' delimit the intentionality aspect, while the very notion of 'game' delimit the context – nothing else can be done with the particular logical sign than whatever is indicated by the rules of the game:

> What I mean by that is the following: when we are talking about negation, for instance, the point is to give the rule '¯¯p = p'. I do not assert anything. I only say that the structure of the grammar of'¯' is such that '*p*' may be substitued for '¯¯p'. Were you not also using the word 'not' in that way? If that is admitted, then everything is settled. And this is how it is with grammar in general. The only thing we do is to tabulate rules. If by questioning I have found out concerning a word that the other person at one time recognizes these rules and, at another time, those rules, I will tell him, ***In that case you will have to distinguish exactly how you use it; and there is nothing else I wanted to say.***
>
> (LW & VC, p. 184.)

(The ***emphasis*** is ours.)

> If someone says the rules of negation are not arbitrary because negation could not be such that ¯¯p = p, all that could be meant is that the latter rule would not correspond to the English word "negation". The objection that the rules are not arbitrary comes from the feeling that they are responsible to the meaning. But how is the meaning of "negation" defined, if not by the rules? ¯¯p = p does not follow from the meaning of "not" but constitutes it. Similarly, $p.p \supset q. \supset .q$ does not depend on the meanings of "and" and "implies"; it constitutes their meaning.
>
> (Ambrose (ed.) 1979, Part I (Philosophy. Lectures 1932–33), 2, p. 4.)
>
> I have said that "¯¯p = p" gives a rule for the use of the word "not"; it gives the meaning by stating how it functions.
>
> (Ibid., Part III (Lectures 1934–35), p. 153.)

If, on the one hand, the emphasis on 'giving the meaning by stating how it functions' is actually present not only in the early middle period as we have seen above, but pervades also his later phase, on the other hand, the distinction between rules of sentence formation and rules of use of the sign in the language-game, which is very much a mark of later Wittgenstein's, is already present in his early middle period as it becomes clear from his view of language as a calculus and calculus as a game in the early thirties. When reading Waismann's *LW & VC* one often comes across passages where the distinction is made clear between rules of grammar and rules of application:

> If I explain the calculus to you by specifying the rule[s] of grammar and application, then you will also understand the statement '*a* cannot be substituted for *a*'.
>
> (*LW & VC*, p. 179.)

while the key semantical aspect is clearly given in the latter:

> I understand a proposition by *applying* it. Understanding is thus not a particular process; it is operating with a proposition. *The point of a proposition is that we should operate with it.*
>
> (Ibid., p. 167.)

We found appropriate to add the 's' to the word 'rule' in the first line, given that in the original it says 'Wenn ich Ihnen den Kalkül erkläre, indem ich Ihnen <u>die Regeln</u> der Grammatik und der Anwendung angebe, dan verstehen Sie auch die Aussage »a ist nicht

durch a ersetzbar«.' Observe that the plural 'die Regeln' is translated into the singular 'the rule' which seems to prevent the interpretation of 'the rules of grammar and application' as 'the rule of grammar *and* the rule of application', which is what is said with different words in Wittgenstein's later account of propositions as in 'and what a proposition is is in one sense determined by the rules of sentence formation (in English, for example), and in another sense by the use of the sign in the language-game.' *Investigations*, Part I, 136, p. 53[e]. As we can readily see, such a later account of propositions distinguishes *two* aspects, namely the 'grammar' aspect (the rules of sentence formation) and the 'application' aspect (the use of the sign in the language-game).

In what concerns the term 'grammar', something needs to be said about the various uses Wittgenstein makes of the word. In connection with 'the use of the sign in the language-game', rules of grammar and rules of use are sometimes collapsed into a single class of 'rules of grammar'. There is, however, a great many evidences that although the concept of 'grammar' as a certain body of rules did play an important part in Wittgenstein's teachings on the position sustained by the meaning-is-use semantical dictum, it was not considered by Wittgenstein himself the ultimate semantical instrument:

> Grammar does not tell us how language must be constructed in order to fulfil its purpose, in order to have such-and-such an effect on human beings. It only describes and in no way explains the use of signs.
>
> (*Investigations*, Part I, 496, p. 138[e].)

Nevertheless, it would seem fair to say that by using the semantical dictum 'meaning-is-use' in favour of *assertability conditions*-based semantic theories, in order to challenge theories of meaning based on the traditional logicist truth-values semantics strongly criticised by Brouwer and the intuitionists, as reminded by Prof. Prawitz in :

> As pointed out by Dummett, this whole way of arguing with its stress on communication and the role of language of mathematics is inspired by ideas of Wittgenstein and is very different from Brouwer's rather solipsistic view of mathematics as a languageless activity. Nevertheless, as it seems, it constitutes the best possible argument for some of Brouwer's conclusions.
>
> (...)
>
> I have furthermore argued that the rejection of the platonistic theory of meaning depends, in order to be conclusive, on the development of an adequate theory of meaning along the lines suggested in the above discussion of the principles concerning meaning and use. Even if such a wittgensteinian theory did not lead to the rejection of classical logic, it would be of great interest in itself.
>
> (Prawitz 1977, p. 18.)

is to insist that the 'rules of sentence formation' and not 'the use of the sign in the language-game' are to be taken as the central aspect of the meaning of a proposition. As we have seen previously, the rules of assertability conditions can hardly said to convey any intentional ingredient.

As a matter of fact, one can distinguish two views of logic according to principles as to how one should look at the links between language and reality. While one view is essentially 'extensional', the other could be classified as 'intentional'. The usual classification, however, is made in terms of whether on the one hand one uses interpretations (denotations) are taken to be the ultimate meaning-giving devices, or on the other hand operations are given the key rôle in settling the quest for meaning.

The denotational view finds its most notable representative in G. Frege, having R. Montague as one the more recent exponents. What characterises this view is that sentences of a given language are endowed with meaning through interpreting the language in some

valuation system. A numerical value from a 'universal domain' is the final symbol which ends the interpretation. The non-denotational view is usually subdivided into the 'sense' view and the 'use' view. The sense view is said to be characterised by the semantics based on *assertability conditions* (sense, as opposed to force, according to Dummett and Prawitz) finding in Gentzen, Heyting and others its early origins. In this view language is said to be understood by syntactic relations between sentences and set of sentences (consequence relations, turnstiles – '$\vdash$') and the central aspect of the meaning of a propositions is given in its conditions for assertion. The use view is attributed to Wittgenstein, and, given his emphasis on the rôle played by the contexts of utterance, this view is said to be inadequate as a basis for any 'workable' theory of meaning because, if what gives meaning to sentences is the *context of utterance*, then 'language is really unanalysable'. We feel that the latter is by no means a positive attitude towards encouraging the use of Wittgenstein's (recognisably appealing to the intuitive understanding of language) use-based semantical theory to analyse formalised languages such as logics, programming languages, etc., because it (apparently) says that one cannot neglect the 'context' of use of expressions sentences, propositions, and therefore language is unanalysable. And this view unfortunately appears to leave one with the impression that the 'sense' (Gentzen's) view should be considered more appropriate because it does not resort to denotations and it only involves 'syntactic relations', no consideration of the 'contexts of utterance' being required.

5. Normalisation and its Semantical Rôle

After D. Hilbert's pioneering attempts, G. Gentzen was certainly one of the first to put into effective terms the analysis of issues concerning the foundations of mathematics in a precise and purely syntactic way. Procedures like proving the consistency of elementary number theory using only syntactical arguments were straightforward results of this 'theory of proofs'. The key idea was that of using the notion of logical steps, *introduction* and *elimination* steps, for a suitable formal representation of mathematical proofs as objects of study in themselves. Most importantly, he also introduced the notion of *normalisation* representing the procedure of showing how 'redundant' steps could be removed from a logical proof without any loss. By performing an *introduction* step for a certain logical connective, and immediately afterwards applying an *elimination* step, one would be making 'redundant' steps which could be removed from the formal representation of the proof without any loss. Applying these ideas to extensions of his *calculus of natural deduction*, he arrived at a precise formulation of what is called the *cut-elimination* procedure. Being a procedure which operates on the objects of the theory – proofs in a certain logic –, in order to allow proving properties of the particular theory – the presentation of a particular logic – such as normalisation, strong normalisation, consistency, etc., it is easily understood as a 'meta-level' procedure.

Starting from Gentzen's ideas, Prof. Prawitz attempts to further explore the *calculi of natural deduction* as workable tools to define/present logics. In his thorough and seminal studies on Natural Deduction (Prawitz 1965), he develops *reduction* rules to perform *normalisation* on proofs constructed from natural deduction inference rules. Recalling that the relationship between *introduction* and *elimination* rules of natural deduction systems is expressed by what is called 'the logical principle of inversion', Prof. Prawitz defines procedures which operate on proof-trees, and which he initially calls reduction *steps*. Only in further developments of those normalisation procedures for natural deduction presenta-

tions of logics are those 'meta-level' procedures referred to as *rules*. However, due to the very nature of those particular presentation systems, where only *introduction* and *elimination* rules take part, the *reduction* rules remained at the 'meta-level', until a certain formalisation of the basic principles of constructive mathematics was presented via natural deduction rules from a new perspective emerging from the identification of propositions with types, and the *reduction* rules were incorporated as 'part of the system' so to speak[1]. This was exactly Prof. P. Martin-Lof's *Intuitionistic Type Theory* (Martin-Lof 1984) which has since had so much attention from many different perspectives on the rôle and the nature of language on the foundations of logic and mathematics, as well as on the logical basis of the 'science of computing'.

Now, what good can such an 'internalisation' of *normalisation* rules do for a better understanding of the meaning of logical signs? In what ways can it be explored and how does it relate to Wittgenstein's 'the use of the sign in the language-game'? With a brief explanation of *Intuitionistic Type Theory*, departing from its original formulation which has essentially an *assertability conditions*-based semantical explanation, but from our own perspective on the relation rules-meaning of such a particular presentation of constructive mathematics, we have attempted elsewhere to make clear some of these important issues (de Queiroz 198?). More importantly, we have endeavoured to explain why our own *'Meaning-As-USE' Type Theory*, which emerged from a semantical reformulation of Prof. Martin-Lof's *Type Theory* as an attempt at recovering the place of *normalisation* as a semantical device, can be seen as a more convincing theory of meaning for the language of mathematics (de Queiroz & Maibaum 1989). Roughly speaking, it would seem fair to say that it convinces better simply because it offers itself as a reasonable attempt to formalise the notion of *utility*, and put it into its proper semantical rôle.

At this point, an experienced proof-theorist who is used to prove consistency of logics presented via proof rules by making use of the canonical argument that 'the formal system is normalisable', might be left wondering whether the 'internalisation' (from 'meta'-level to 'object'-level) of normalisation rules would affect such a universal procedure. The answer is very simple, and the witness very well known: in Church's λ -Calculus one has the 'conversion' ('reduction', 'normalisation', etc.) rule in the object-level (Church 1941, Chapter II, 6. CONVERSION, p. 12.), and yet proofs of consistency, existence of normal forms, etc., making use of the ß-rule of normalisation are commonplace in mathematical studies of Church's calculus. To acknowledge the semantical rôle of *normalisation* rules is not by any means to invalidate or to ignore the part they have to play in proofs of consistency, confluence, etc.. By extending the work of A. Church of formalising mathematical concepts through explicit rules of operation, from the single notion of computable function to a much broader context of logics, types and theories, Prof. Martin-Lof has also 'brought in', so to speak, the rules of *normalisation* from the 'meta'-level to the 'object'-level. Again, similary to Church, he also makes use of such semantical of *normalisation* in proving essential meta-mathematical properties of the formal system such as 'Normalization theorem', 'uniqueness of normal form', 'Church-Rosser property' and 'decidability of the convertibility relation' (**theorem** 3.3, and PROPOSITIONS 3.6, 7.7 and 3.8, respectively, p. 101 and p. 114 of Martin-Lof 1975).

It seems relevant to emphasis that when we say that the *normalisation* rules have a fundamental semantic rôle to play in the meaning theory supporting what we call *'Meaning-As-USE' Type Theory*, we refer to the rôle they play in terms of explaining the meaning of the actual logical connective. The fact is that the rules of *normalisation* are in some sense formalising the explanation of the *consequences* one is allowed to draw from

the corresponding proposition, and for that reason they serve as a semantical device which we take to be a formal counterpart to L. Wittgenstein's informal teachings on the concepts of meaning and proposition, namely 'the meaning of a word is its use in the language' and 'what a proposition is is in one sense determined by the rules of sentence formation, and in another sense *by the use of the sign in the language-game*' (de Queiroz 1989ab). This particular view is reflected directly in the title of another paper of ours, 'To Show the Consequences of a Proposition is To Give its Meaning' (de Queiroz 1989a), which attempts to demonstrate that the connections between the meaning of a proposition and the consequences one is allowed to draw from it, seems to appear in differents forms throughout Wittgenstein's early as well as later writings.

6. Computing Science and the Notion of Types

The usual mathematical notion of set as a subset of the universe of values, which has proved adequate for most developments in the mathematics of the second half of this century, proved inadequate to support the notion of data type as understood by computer scientists. Instead of defining a type by what it contained, it was felt that a data type definition had to convey the information demonstrating how the elements of the type were to be used, computed, etc.. That was the view already put forward in 1973 by J. Morris in a paper entitled 'Types are not Sets'. He begins by saying:

> The title is not a statement of fact, of course, but an opinion about how language designers should think about types. There has been a natural tendency to look to mathematics for a consistent, precise notion of what types are. The point of view there is extensional: a type is a subset of the universe of values. While this approach may have served its purpose quite adequately in mathematics, defining programming language types in this way ignores some vital ideas.
>
> (Morris 1973, p. 120.)

A whole body of ideas developed mainly in the seventies aiming at gaining a better understanding of the concept of data types began to show that data type definitions were worth studying as a theory in itself. With the help of algebraic techniques, the notion of data type as defined by what can be done with its elements proved as a convenient alternative to the mathematical notion of set as defined by what the set contains. In the words of the ADJ Group, which were responsible for many of the seminal ideas on the mathematical characterisation of the notion of data types:

> At least since Morris, it seems to be generally accepted that types are not just sets, but sets equipped with operations, and that is *exactly* what many-sorted algebras are.
>
> (Thatcher, Wagner & Wright 1982, p. 711.)

And, yet, there is a clear mismatch between the notion of 'types' as understood by the theories of abstract data types, and the concept of type as understood from the semantical explanations of *Intuitionistic Type Theory*. For example, the type of 'lists of items' is presented in *Intuitionistic Type Theory* in a rather different form than the usual computer science, abstract data types manner (cf. the usual LISP-list presentation through 'cons' *and* 'CAR-CDR'). Due to the semantics of 'canonical values', the type of lists is not presented through its usual operations 'CAR-CDR' or 'HEAD-TAIL', but only the rule demonstrating how to form the canonical values of the type is exhibited ('cons'-rule), following the semantical principle that the meaning of a type is given by its canonical elements, that the meaning of a proposition is given by its canonical proof[2]. As a matter of fact, such a semantical standpoint does not sound too convincing for a computer scientist. One has to

make a sharp distinction between the type of 'queues of items', which is 'destroyed', 'computed' via the discipline 'first-in-first-out', and the type of 'lists of items' which is computed through a rather distinct discipline 'last-in-first-out', despite both having 'isomorphic' canonical elements: both contain sequences of items, but are only recognised to be either 'lists' or 'queues' as soon as the *normalisation* of non-canonical elements is shown. The canonical element by itself does not make one see the difference, apart from a judicious choice of 'names' ('cons' for lists, as opposed to 'addq' for queues.) Looking at (algebraic) specifications of the two types one realises that, although the rules of syntax which determine how to form canonical elements (in both cases 'sequences of items') are identical up to names, the way a non-canonical element of the type 'list of items' is *normalised* via its destructors 'CAR' and 'CDR' is very different from the way a non-canonical element of the type 'queue of items' is *normalised* by means of its destructors FRONT' and 'REMOVE' (see algebraic Specification of *queue* in Ehrig & Mahr 1985, p. 55). Cf. also J. Guttag's explanation of the axiomatic approach to data type definitions using the same example of the data type 'queue of items', which comes after the justification for the axiomatic approach to the specification of data types given the insufficiency of the 'syntactic' information given by procedures to form the 'canonical' values of types. The axiomatic is supposed to demonstrate the distinguishing characteristic of a type, which is hardly given by the form of its elements, but rather by what can be done with them:

> The distinguishing characteristic of a queue is that it is a first in-first out storage device. A good axiomatic definition of the above operations must therefore assert that and only that characteristic.
>
> (Guttag 1977, p. 398.)

Despite the 'breakthrough' made by such an alternative and rather convincing account of types as characterised by the operations that could be done with the elements, the classical extensional view proved stronger, and the notion of data types found a mathematical counterpart in the notion of many-sorted algebras. An 'operational' or 'intentional' ingredient was introduced when the view on 'sets are characterised by their content' was challenged by 'types are characterised by their operational behaviour'. The strong traditional view on numerical values and 'a sign denotes some object from a valuation domain' prevailed, given that many-sorted algebras were just another part of the classical extensional mathematics. Therefore, the chief semantical concern, even in the 'computer science' alternative for a formal characterisation of the concept of data type, was still directed towards the classical view that 'the quest for meaning is settled by denotations', 'meaning is a (numerical) value', and many difficulties were encountered to explain the meaning of oddly formed 'terms' such as the top of an empty stack, the front of an empty queue, etc..

As a matter of fact, computing science provides interesting examples of how one can have an intuitively appealing and yet 'pratical' semantics by relating meaning with use/utility. Data structures defined to help organising information in certain structured ways (lists, stacks, queues, trees, etc.) are generally understood by looking at their utility as storage devices: stacks as last-in-first-out, queues as first-in-first-out devices, etc. In one way or another the distinguishing characteristic of each of them is made 'precise/formal' through rules resembling rules demonstrating their *use/utility/consequence*, namely the *normalisation* rules: when demonstrating how to recover data 'introduced' in the memory device, one describes how destructors operate on constructors. It would seem fair to say that it was based on some form of the logical principle of *inversion* (see, e.g., Prawitz 1965,

chapter II) that a theory of data type definitions began to be developed in the early seventies, where data types were said to be defined by rules of 'grammar' indicating when to form the 'canonical' data type expressions, and rules of application demonstrating the effect of *eliminatory* operators (the 'destructors') on the terms resulting from the application of *introductory* operators (the 'constructors').

7. Concluding Remarks

When attempting to fill in the two less widely recognised sides of (our view of) Wittgenstein's semantic triangle-like diagram with vertices in *meaning*, *use* and *consequences* we have tried to demonstrate how one can bring his celebrated 'meaning-is-use' semantical paradigm more 'down to earth', so to speak, by demonstrating that if 'the use of a sign' is understood as 'the utility of the sign' then 'use' can be made more precise and more 'concrete' for the case of formalised languages by means of some kind of *elimination/consequences-explanation* rules. In the context of formal logic, for example, there are logical inference rules which can be taken as being the rules playing the part of explaining the utility of logical signs by showing what can be deduced from a proposition formed with that sign as its major one. To realise such a demonstration of the conclusions one has the usual logical rules of *elimination*. It would seem fair to say that the best candidate for a *consequences-explanations* logical rule would be the rules of *normalisation* which constitute more informative devices as 'formal explanations of the consequences' than the relatively poor and restrictive natural deduction-like *elimination* rules.

Moreover, by emphasising the semantical rôle of the proof-theoretic notion of *reduction* rule, we believe we have provided a useful perspective on the connections between proof theory and the logical foundations of computer programming. The idea that a *reduction* (*normalisation*) rule can be looked at as a semantic instrument should prove a useful conceptual view that could allow the unification of techniques from theories of abstract data type specification with techniques from proof theory, constructive mathematics and λ -Calculus. For this reason we are also led to believe that these connections stand as an important conceptual basis for a new foundational view on the form of frameworks for formal software development, which could prove fruitful for a more methodological view of computer programming.

(**Draft**, September 12, 1989)

Acknowledgements

I am very much indebted to Dr T. S. E. Maibaum for his invaluable academic support. Financial support from 'CNPq' (Brazilian national research council) through grant 20.2724/84-CC, and from GENESIS, ESPRIT Project 1222(1041), CEC, is here gratefully acknowledged.

Notes

(1) 'In the interpretation of sets as propositions, the formation rules are used to form propositions, introduction and elimination are like those of Gentzen, and the *equality rules correspond to the reduction rules of Prawitz*', Martin-Lof 1984, p. 24. (the *emphasis* is ours)

(2) '[The Meanings of the Logical Constants] The meaning of each constant is to be given by specifying, for any sentence in which that constant is the main operator, what is to count as

a proof of that sentence, it being assumed that we already know what is to count as a proof of any of the constituents.', Dummett, 1977, p. 12.

References

Ambrose, A. (ed.): 1979, *Wittgenstein's Lectures: Cambridge. 1932–35*, Basil Blackwell, Oxford, xii+225pp.

Church, A.: 1941, *The Calculi of Lambda-Conversion*, series Annals of Mathematical Studies, N. 6, Princeton University Press, Princeton, 77pp.

Dummett, M.: 1976, *'What is a Theory of Meaning? (II)'*, in *Truth and Meaning: Essays in Semantics*, Gareth Evans & John Mcdowell (eds.), Clarendon Press, Oxford, xviii+419pp, pp. 67–137.

– : 1977, *Elements of Intuitionism.* (With the assistance of R. Minio), series Oxford Logic Guides, General Editor: D. Scott, Clarendon Press, Oxford, x+467pp.

Ehrig, H. & B. Mahr: 1985, *Fundamentals of Algebraic Specifications 1. Equations and Initial Semantics*, series EATCS Monographs on Theoretical Computer Science, vol. 6, W. Brauer, G. Rozenberg & A. Salomaa (eds.), Springer– Verlag, Berlin, XI+321pp.

Geach, P.T. (ed.): 1988, *Wittgenstein's Lectures on Philosophical Psychology 1946–47*, (Notes by P.T. Geach, K.J. Shah & A.C. Jackson), Harvester–Wheatsheaf, Hemel Hempstead, xv+348pp.

Guttag, J.: 1977, *'Abstract Data Types and the Development of Data Structures'*, in *Communications of the ACM*, June 1977, Volume 20, Number 6, pp. 396–404.

Martin Lof, P.: 1975, *'An intuitionistic theory of types: predicative part'*, in *Logic Colloquium' 73*, H.E. Rose & J.C. Shepherdson (eds.), series Studies in Logic and The Foundations of Mathematics, North–Holland, Amsterdam, pp. 73–118.

– : 1984, *Intuitionistic Type Theory*, (Notes by G. Sambin of a series of lectures given in Padova, June 1980), series Studies in Proof Theory, Bibliopolis, Naples, iv+91pp.

Morris, J.H., Jr.: 1973, *'Types are not sets'*, in *Conference Record of ACM Symposium on Principles of Programming Languages*, held in Boston, Massachussetts, October 1–3 1973, pp. 120–124.

Prawitz, D.: 1965, *Natural Deduction. A Proof-Theoretical Study*, series Acta Universitatis Stockholmiensis. Stockholm Studies in Philosophy 3, Almqvist & Wiksell, Uppsala, 113pp.

– : 1977, *'Meaning and proofs: on the conflict between classical and intuitionistic logic'*, in *Theoria* (Sweden), vol. XLIII, pp. 2–40.

de Queiroz, R.J.G.B.: 1988, *'A Proof-Theoretic Account of Programming and the Rôle of Reduction Rules'*, in *Dialectica*, vol. 42, n. 4,pp. 265–282.

– : 1989a, *'The Mathematical Language and its Semantics: to Show the Consequences of a Proposition is to Give its Meaning'*, Abstract published in the *Abstracts of the Thirteenth International Wittgenstein Symposium*, organised by the Österreichische Ludwig Wittgenstein Gesellschaft, Institut für Wittgensteinforschung, Analytische Philosophie und Wissenschaftstheorie, 93pp, p. 17, Kirchberg/Wechsel, Austria, August 14–21 1988. Full paper in the *Reports of the Thirteenth International Wittgenstein Symposium 1988*, (Edited by P. Weingartner & G. Schurz), series Schriftenreihe der Wittgenstein–Gesellschaft, Vol. 18, (Edited by E. Leinfellner, R. Haller, A. Hübner, W. Leinfellner & P. Weingartner), Holder–Pichler–Tempsky, Vienna, 304pp, pp. 256–266.

– : 1989b, *'Meaning, Function, Purpose, Usefulness, Consequences – Interconnected Concepts'*, (presented at the *Fourteenth International Wittgenstein Symposium (Centenary Celebration)*, Kirchberg/Wechsel, August 13–20 1989. Abstract in p. 20 of the *Abstracts* of the Symposium, and full paper to appear in the *Reports* of the Symposium.)

– : 198?, *'Normalisation and Language-Games'*,(to appear).

de Queiroz, R.J.G.B. & T.S.E. Maibaum: 1989, *'Proof Theory and Computer Programming'*, (to appear in the *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*).

Thatcher, J.W., E.G. Wagner & J.B. Wright: 1982, *'Data Type Specification: Parameterization and the Power of Specification Techniques'*, in *ACM Transactions on Programming Languages and Systems* (TOPLAS), vol. 4, N. 4, October 1982, pp. 711–732.

Waismann, F.: 1979, *Ludwig Wittgenstein and the Vienna Circle*, (Reprinted 1983), (edited by B. McGuinness), (German edition 1967), (translated by J. Schulte and B. Mcguinness), Basil Blackwell, Oxford, 266pp. (referred to simply as *'LW & VC' passim.*)

Wittgenstein, L.: 1922, *Tractatus Logico–Philosophicus*, (transl. by D.F. Pears & B.F. McGuinness, with an Introduction by B. Russell), ed. 1961, Routledge & Kegan Paul, London, xxii+89pp. (referred to simply as *'Tractatus' passim.*)

– : 1953, *Philosophical Investigations*, (transl. by G.E.M. Anscombe), (2nd edition, reprinted 1984), Basil Blackwell, Oxford, viii+250ᵉpp. (referred to simply as *'Investigations' passim.*)

– : 1958, *The Blue and Brown Books. (Preliminary Studies for the 'Philosophical Investigations')*, (2nd edition 1969, reprinted 1989), Basil Blackwell, Oxford, xiv+192pp. (referred to separately as *'Blue Book'* or *'Brown Book' passim.*)

– : 1961, *Notebooks 1914–1916*, G.H. von Wright & G.E.M. Anscombe (eds.), (with an English translation by G.E.M. Anscombe), Basil Blackwell, Oxford, 140+91epp. (referred to simply as *'Notebooks' passim.*)

– : 1967, *Zettel*, G.E.M. Anscombe & G.H. von Wright (eds.), (transl. by G.E.M. Anscombe), (2nd revised edition 1981), Basil Blackwell, Oxford, vi+138pp.

– : 1969, *On Certainty/Über Gewissheit*, G.E.M. Anscombe & G.H. von Wright (eds.), (transl. by D. Paul & G.E.M. Anscombe), Harper & Row Publishers, New York, vii+90+90ᵉpp.

– : 1971, *Prototractatus*, (An early version of *Tractatus Logico–Philosophicus*), B.F. McGuinness, T. Nyberg & G.H. von Wright (eds.), (With a transl. by D.F. Pears & B.F. McGuinness), (an historical introd. by G.H. von Wright), (and a facsimile of the author's manuscript), Routledge & Kegan Paul, London, 256pp.

– : 1980, *Remarks on the Philosophy of Psychology*, Vol. II, G.H. von Wright (ed.), (transl. by C.G. Luckhardt & M.A.E. Aue), Basil Blackwell, Oxford, 143+121ᵉpp. (referred to simply as *'RPP II' passim.*)

– : 1982, *Last Writings on the Philosophy of Psychology*, Vol. I, *(Preliminary Studies for Part II of 'Philosophical Investigations')*, G.H. von Wright & H. Nyman (eds.), (transl. by C.G. Luckhardt & M.A.E. Aue), Basil Blackwell, Oxford, 148+127ᵉpp. (referred to simply as *'Last Writings I'passim.*)

Eugene J. Meehan
Univ. of Missouri, St. Louis (USA)

Logical Models of Human Judgments: Limits and Possibilities

Abstract

The central thesis of the paper is: if judgments are necessary for creating the kind of knowledge needed to direct or manage human affairs on defensible grounds, and judgments cannot be formalized, then those concerned with creating logical models that are to be used for such purposes need to be aware of those limits as well as potential benefits. Although two basic limits can be identified, the formal limit implied by Gödel's theorem is relatively minor, and the primary concern is the limit imposed by the need to incorporate a set of normative assumptions in every decision and to provide an adequate justification for doing so.

Resumo

A tese central deste estudo é a seguinte: se são necessários juízos para se criar o modo de conhecimento requerido pela direcção ou gestão dos assuntos humanos consoante fundamentos defensáveis e se os juízos não podem ser formalizados, segue-se que os que visem a criar modelos lógicos destinados a intervir naqueles propósitos devem ter em atenção as decorrentes limitações, tanto como consideram os potenciais benefícios. Embora possam ser identificados dois limites básicos, o limite formal implicado pelo teorema de Gödel é relativamente menor, sendo primacial como preocupação o limite imposto pela necessidade de integrar um elenco de assunções normativas em cada decisão e de justificar adequadamente que se decida de dada maneira.

Construed in broad terms, my purpose in this paper is to analyse and thus to clarify the potential uses and limitations of formal logical systems in the development, justification, and application of knowledge. The discussion is meant to apply to all formal structures, whether or not they are referred to as "models", or incorporated into an electronic or mechanical apparatus. To facilitate the analysis, and obtain a much sharper delineation of the role of formal logic in the development and use of knowledge, the focus of the paper has been narrowed to an examination of the function of logic in complex intellectual activities that involve processes to be labeled "judgments", identified (tentatively) as ways of reaching conclusions that cannot be formalized. Since judgments can be shown to play a crucial part in creating and using every form of knowledge, some means of circumventing that limitation is essential if logics or calculi are to be used to make decisions. The reasons why judgments cannot be formalized, and the manner in which that limit can be evaded, will be shown by a systematic examination of the use of logical systems to direct human actions in ways that can be justified as fully as human capacity allows — of the use of logic to make defensible or justifiable decisions or choices leading to actions. And since actions serve to link knowledge to the conditions of life of the human

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 167/175
1992, APDC & SPF, Lisboa

population affected by its use, an evasion is a matter of considerable practical importance.

To summarize the central thesis of the paper: if judgments are necessary for creating the kind of knowledge needed to direct or manage human affairs on defensible grounds, and judgments cannot be formalized (the process cannot be captured by a formal system), then those concerned with creating logical models that can be used for such purposes need to be aware of these limits as well as any potential evasion routes. The thesis is stated and argued in wholly general terms, without specifying either the kinds of decisions to be made or the kinds of formal models employed.

The context required for achieving the purposes set for the paper, the perspective from which the role of logic can fruitfully be examined, is an overall conceptualization of human knowledge in which the requirements for directing actions on defensible grounds can be established. In effect, an analytic framework must be created able to show the kind of knowledge that is (a) adequate for directing human actions and (b) corrigible out of experience within the limits of human capacity. The discussion must, therefore, first identify the purposes which knowledge must serve (for improvement is always measured relative to some purpose) and the human capacities available for fulfilling them. That allows specification of the kind of knowledge that is required to fulfill those purposes and of the manner in which claims to have such knowledge can be justified or assessed. A framework can then be sketched that allows a fairly precise statement of the knowledge requirements for directing human actions on defensible grounds. That in turn allows the user to show the functions that judgments perform (and logic must equal) in the overall process of developing and applying knowledge. From this base, a systematic discussion of the uses and limits of formal logic in intellectual performances that involve judgments can be carried out. It need hardly be said that space constraints enforce a horrendous compression and simplification of the argument.[1]

Definitions and assumptions: the analytic framework

A logic or calculus is simply a set of rules for manipulating a set of nominally-defined symbols. Logics or calculi are purely formal: they have no relation, *qua* logic, to the world of observation. Applications of logic to the world of experience require a theoretical linkage, which must in turn be justified. Further, logics are limited in capacity; at most, they rise to the level of a "Turing machine" that can reproduce its own content fully. Logics cannot "create" in the human sense of that term. Logic has enormous power in human affairs, for it provides a way of calculating implications that can be validated in absolute terms, but that power is contingent on creating an acceptable theoretical link between the symbols and processes incorporated into the logical system and the substance of human experience, adequately conceptualized.

The nature of "judgment" is most clearly visible in a context supplied by the techniques available to humans for reaching conclusions. Three fundamental procedures may be used for that purpose: First, decisions can be based upon direct observation or measurement, but only if they refer to propositions in the form "X was (or was not) the case". Second, conclusions can be generated inferentially, by logical calculation from an established (or assumed) set of propositions, some of which must be generalized in form. In all other cases, here labeled "judgments", conclusions are based on the functioning of the neural apparatus but the specific content of the constituent elements of the decision process remain unknown. In judgment, the neural system functions like a "black box" whose products can be observed but whose processes remain hidden. A judgment, then,

can be regarded as a conclusion reached by the functioning of the total intellectual system: the elements of the system that are involved in any particular judgment cannot be isolated. Of course, all conclusions depend ultimately on such judgments; both observation and applied logic, further analyzed, require prior judgments. Judgments may well be produced formally within the neural apparatus, but the assumptions from which the inferences are drawn are not known. For that reason, the judgmental process defies formalization; no incomplete system can be formalized. In these terms, acquiring the capacity to deal effectively with human judgments is a major problem in every knowledge system.

Knowledge is here construed as "organized" human experience, meaning that a logical pattern has been imposed on an incoming flow of raw perceptions. The organizing process is subject to all of the limitations entailed by the need to rely on the sensory apparatus for raw data and to express the results in a language that is itself a human artifact. In these terms, logic appears as the basic apparatus for organizing experience, and knowledge is usefully regarded as the result of applying logic to sensory experience. It must be emphasized, of course, that the application of logic to experience is not a problem *within* logic; the use of logic is always a matter for judgment and argument and not calculation.

Criticism and improvement of such knowledge, like the criticism of any other human instrument, depends on the purposes which knowledge is meant to serve. The kind of knowledge that concerns us here must be able to direct human actions on defensible grounds and must be corrigible over time out of past experience. An analytic framework that is adequate for dealing with such knowledge must therefore: (1) identify, in suitably generalized terms, the purposes that must be achieved if human actions are to be placed on defensible grounds knowledge in suitably generalized terms, (2) summarize the relevant human capacities available for fulfilling those purposes, and (3) provide a metalanguage or theoretical structure (precisely equivalent to the kind of "theory of play" available to coaches in well-developed sports) that will show how those purposes can be achieved within the limits of available capacity plus any other relevant constraints.

Development of the set of purposes that must be fulfilled through the use of knowledge *if knowledge is to provide a justifiable base for action*, is absolutely contingent on acceptance of an overall purpose for the human enterprise. In practice, that rather formidable-sounding task turns out to be relatively simple. The prime purpose that knowledge must fulfill is always and everywhere the maintenance and/or improvement of the conditions of life of *some* human population. The genetic inheritance is inadequate even for species maintenance; it must be augmented by creating knowledge and learning how to improve it. That much agreed, the principal corollaries to the overall purpose set for the human intellectual enterprise are readily determined. At a minimum, knowledge must provide the capacity to: (a) predict future events from presently available observations; (b) control future events (produce or prevent them) by human actions; and (c) make choices. The list need not be exhaustive, though at present there seems no good reason to extend it, but even if extension proves necessary, it already includes so much of what is significant in human affairs that an analytic apparatus that can deal effectively with those three purposes is well worth prusuing.

Four types of human capacity, all widely distributed within the species, are essential for creating knowledge adequate for fulfilling these three primary purposes — and thus achieving the overall goal of maintaining and/or improving the conditions of life of some human population. First, the neural apparatus must be able to discriminate, and thus impose a basic ordering on the incoming flow of perceptions; second, the system must be capable

of generalizing particular conclusions or judgments, a capacity that is much strengthened (but made riskier) by the availability of language; third, it must be able to calculate, to draw logical inferences or explore the content of generalized propositions or combinations of propositions; and fourth, the individual must react affectively, must *care* sufficiently to act about differences in conditions of life — members of the species cannot be wholly indifferent to difference.

Finally, the task of creating knowledge is much facilitated by two characteristics of the human situation, each a commonplace. First, every individual lives in an ongoing knowledge system and at least some of that knowledge is in principle justifiable or defensible; there is no need to begin with a *tabula rasa*. Second, all that is needed to succeed is the capacity to deal with recurring or repetitive events. Both conditions are essential, for if it were necessary to begin the search for knowledge anew with each individual, very little would be achieved and species survival would be unlikely, and the human intellectual apparatus is incapable in principle of dealing with any event that is *wholly* unique.

In addition to a supply of logics that include a variety of relational or logical terms, and a language, (both of which must be taken as given here) a number of different types of patterns is needed to fulfill the purposes set for the knowledge-creating enterprise. Five such patterns are fundamental: (1) *concepts* or classifications, which enumerate the shared or "distributed" attributes of class members; (2) *forecasts*, which link two or more events through formal rules so that one event can be used to predict the other under stipulated limiting conditions; (3) *theories*, which also link two or more events by rule, but which include an assumed causal relation among the events and thus provide one of the essential requirements for directing actions; (4) *priorities* or preference-orderings, which when applied select the preferred outcome (the outcome whose selection can be justified or defended) from within the set of outcomes available for choice; and (5) *policies*, or action programs (comprised of theories but serving to reify priorities) that provide a way to achieve the preferred outcome under stipulated limiting conditions.

The need for different instruments to serve different purposes, and the consequent need for an appropriate mode of justification for each instrument, underscores the principal weakness of the concept "modeling". In effect, the term "model" refers to any type of pattern included in the knowledge system — it lumps together all of the basic intellectual instruments into a single category. But no justification can be provided for accepting and applying knowledge until the purpose the knowledge is expected to serve (which can be generalized very broadly so long as criteria can be determined for fulfilling it) has been stipulated. The ambiguity introduced by the use of "modeling" largely accounts for the tendency for economists, among others, to insist that predictability is an adequate test of theory, or to claim, following Milton Friedman's famous essay on the methodology of economics,[2] that *any* model may contain false assumptions — an assertion that is tenable with respect to forecasts but grossly mistaken if a basis for action is being created. The distinction is particularly important if the goal in modeling is to create an instrument able to produce decisions or direct actions in real cases.

Classifications are created by generalizing the shared attributes of a set of "things" that can presumably be treated in the same manner; they are tested and refined by further observations of other things that share the same attributes. Forecasts, which link two or more events by rule so that observing one event allows prediction of another, are generalized from past experience, contain no causal assumption, and can be "tested" only against past experience. Experimentation is ruled out. Theories, which have the same

structure and function, include an assumed causal connection between events and therefore allow experimentation. That makes possible the deliberate production of stronger supporting evidence for the assumed relation (testing), and provides a basis for corrigible action. Policies are comprised of theories, and are developed and tested in the same manner as theories. Priorities, which are the most difficult part of the knowledge system to both produce and justify, are based on past experience but cannot be tested in the usual sense of the term. That is, a proposition in the form "Do X to produce Y under conditions C" is easily tested given its everyday meaning, for if X is done, then Y should be produced. But propositions in the form "Prefer Outcome A to Outcome B" do not have a meaning that allows for testing in action in the same sense — the results of acting cannot, given the meaning of terms, produce evidence suggesting that one outcome *should* be preferred to another. Propositions about preferences must be argued, creating a special class of problems to be discussed more fully below.

For directing actions on defensible grounds, a number of such patterns is combined with a range of other assumptions that form part of the overall epistemology. Some of those assumptions have far-reaching implications. For example, the assumption that all species members are wholly dependent upon the sensory apparatus for information forces inductive development of knowledge, generalization from specific and particular observations. And choices or actions (which are analytically identical) require an initial normative commitment — implied in the assumption that the overall purpose for which knowledge is needed is to maintain and/or improve the conditions of life of some human population — to the principle *homo mensura*. Without the assumption that the conditions of human life in the various outcomes available for choice (a function of the capacity of the actor doing the choosing) are the central focus of justification, the knowledge system remains incorrigible.

Action, which provides the central focus of the analysis, is usefully defined in analytic terms as the capacity to produce change in the environment. If that capacity can be exercised voluntarily, action or choice is forced — even if the capacity is not exercised, the world will be different than it could have been were capacity exercised differently and that is the indicator for action. That approach to action has a number of advantages for systematic inquiry. First, the content of any choice is determined by the capacity of the actor, who may be either an individual or a collectivity, and is analytically distinct from the actor's awareness or intentions — the focus of criticism is the action and not the actor. If choice is to be defensible, the content of the outcomes available for choice must be known and compared; that requires a form of knowledge able to link actions to their consequences (a theory, which can link two or more events causally by rule) but whether or not this is part of the outcomes must be expressed in terms of the conditions of life of the population affected by the action, an appropriate set of concepts must already be available, and it must be possible to translate the concepts used in the theories that project the content of the available outcomes into the concepts used in the normative apparatus. The outcomes are compared, seeking reasons in experience (which may be purely affective reactions, in the simplest case possible, but are more likely to involve complex intellectual considerations in most cases) for preferring one to the others. In principle, a priority must be created that will actually *force* a specific action or choice in a situation to which it applies, otherwise the normative apparatus cannot be applied or improved.

Implications for applied logic.

Briefly summarized, the knowledge-creating enterprise emerges within the analytic

framework as an ongoing effort to organize past experience into formal patterns that can be used to fulfill the basic purposes required for maintaining and improving the human condition. In that context, the development of a field of knowledge becomes a matter of articulating and elaborating organizing assumptions that will order past experience in humanly useful ways — usefulness being determined by the kinds of purposes that can be fulfilled using the patterns created. To create knowledge, formal patterns or calculi are used to organize past experience: the kind of pattern that is created, and the kind of evidence that is needed to justify accepting and using it, depend on the purpose for which the pattern will be employed. The normative purposes accepted for the theory of knowledge provides the measure of achievement for the overall enterprise. The dynamics of the knowledge system are provided by the calculating capacity of formal logic. If the elements and processes contained in a calculus can be linked firmly to the world of observation, if an adequate set of rules of correspondence can be devised, a useful and corrigible piece of knowledge can be created.

The role of formal logic in the creative phase of knowledge-production is extremely limited. Some formal inferences may be employed in the overall process but they are interwoven heavily with judgments, often of a very complex nature. All conclusions relating to the adequacy of supporting data for a general proposition, or the adequacy of the conceptual apparatus used to structure a descriptive account, the relevance of data to any specific proposition, or the defensibility of a preference or priority, are matters of judgment. Moreover, the concept of "rational" judgment, to the extent that the meaning of "rational" is equated with formal validity, is either inconsistent or meaningless, and therefore cannot provide an escape route from the predicament. Further, there are good reasons to suppose that the intellectual apparatus operates in a number of different modes, and little is presently known about the ways in which these modes interact.(3) In sum, the process by which knowledge is created is as opaque as the process used to make judgments, and requires the same kind of evasion effort. Formal reasoning may assist in the creative process, as J.S. Mill's "methods" simplify the search for causal relations by eliminating from consideration some of the factors in a list of potential causes selected on theoretical or other non-formal grounds. And there are excellent reasons for seeking to extend the influence of formal reasoning as far as is legitimately possible. Accordingly, the remainder of the paper will be concerned with the factors that limits or conditions the use of formal systems in knowledge development, and how their influence can be reduced or avoided.

To establish the precise nature of the problem, and the character of the solution proposed, the knowledge-production system is construed as a "Black Box" function involving the whole intellectual apparatus. In principle, the content of the black box is opaque to human observers; in that sense, knowledge-creation as a process is analogous to the operation of the natural universe, for its functioning principles, if indeed there are any, are also "unknowable" — even if they were correctly identified, there would be no way of demonstrating the correctness of the identification. The process of judgment, then, lies beyond human capacity to replicate and incorporate into a formal logical structure. Indeed, there is no way to actually "test" a judgment, for what is tested in the ordinary meaning of the term is the validity of the set of assumptions that led to a particular conclusion; obviously, if those assumptions cannot be stated they cannot be tested. The only feasible test of a human judgment is another human judgment, a comparison of the reasoning and evidence that can be marshalled to supported different judgments when they appear, carried out by competent human judges. Whether in physical science or in human

affairs, including normative inquiry, the court of last resort is always the consensus of the community of informed and competent persons. The members of that community may be very difficult to identify and in many cases, they will not agree among themselves. In the absence of consensus, the case in question can only be regarded as moot.

Some way of evading the limit imposed by the need to rely on judgments must be created, otherwise formal logical systems are of little value for the development and use of knowledge, and knowledge is in fact impossible to achieve. The "solution", which is a commonplace in the philosophy of inquiry, is to ignore the processes by which judgments are reached and concentrate on replicating the conclusions. Since that is the only procedure available for dealing with "Black boxes", at least to the extent that they are truly opaque, it must be accepted as a point of departure. The task of producing knowledge, from that perspective, appears as an effort to create a logical apparatus that reproduces the external characteristics of "black box" performance. There is no justification for assuming that the axioms incorporated into the logical system replicate the internal logic of the "black box", nor is there any need to do so. So long as logical structures are created with a specific purpose in mind, and linked firmly to the conduct of human affairs, successive applications and the elimination of discrepancies as they appear can lead to intellectual tools of immense power and accuracy.

The way in which the need for a trained nurse in a hospital recovery room could be eliminated over time by formalizing performance functions will suggest both the procedure that is required and its limits. Initial formulations of a formal apparatus to perform the nurse's functions could be expected to be grossly inadequate: the inadequacies would emerge in application, presumably carried out under suitable safeguards. As the logical apparatus was modified and improved, it would tend to stabilize: the number of aberrant cases would decrease, the margins of error would be reduced, the scope and precision of the apparatus would improve. At some point, the marginal cost of maintaining a nurse and a formal system would become too high, or the cost of further improvements in the logic structure would far outweigh the cost of maintaining a nurse, and there development could cease.

Of course, if the situations to which the logical structure is applied show wide ranges of variance, formalization may be severely restricted until the level of differences can be reduced. For exemple, a computer cannot, or better should not, be programmed to perform an appendectomy so long as appendixes display the degree of variance in size, location, configuration, and so on that is presently observed. Future improvements in conceptualization may reduce, or even eliminate such variations, but if that does not occur, the programming requirements for developing an adequate feedback system to guide the computerized operation would be prohibitive. Training a physician to perform appendectomies is a much simpler task.

The great virtue of computerizing knowledge is their capacity to build very large complexes, combining large numbers of patterns, while maintaining the logic coherence of the overall structure. In a sense, that potential is the major contribution to formalization, or potentially to improvement of the available knowledge supply, that has appeared thus far in human history — the construction of "knowledge" employed here can be taken to equate with valid formalization. However, knowledge remains of a different order from strictly formal propositions, if only because knowledge claims are always in some degree uncertain — a function of the artificial closure required to allow calculation or formal inference. Knowledge is likely to remain dependent upon human judgment into the indefinite future, unless and until the need for further knowledge vanishes.

There are, and presumably will remain, immutable limits to the use of logic in the development and application of knowledge. There are two reasons for asserting that position, two grounds for the judgment: First, if Gödel's theorem is correct, there will always be an outer limit to size of systems of interacting variables that can be created while maintaining the logical integrity of the overall structure. Second, every knowledge claims rests in principle on a set of normative judgments, and their peculiar character serves as a different kind of limit on the formalization of human experience.

The need to base knowledge systems on normative decisions is far and away the more important of the foreseeable limiting factors. Again, there are two major reasons for asserting that position. First, in a definitive formalization of knowledge, future human reactions to changes in their conditions of life would have to be anticipated and included in the formal apparatus. But there is no way to anticipate what those conditions will be since they are at least partially contingent upon interim human creativity, and future human reactions to changes in potential cannot be projected because they are presently trans-experiential — they will be made under conditions of life that have not yet been experienced, and refer to options that are also beyond present experience. If that limitation is ignored, the resulting knowledge system becomes an intellectual prison rather than an evolving human product, and potentially at least a self-fulfilling ordinance rather than a product of human judgment and capacity. Second, the justification for a normative conclusion, the kind of argument that *should be* persuasive to a competent critic, will vary with knowledge and circumstances. The human life that is the basis for justifying preferences does not comprise a definable continuum, no standard unit for measuring that continuum is possible, and one life is not wholly interchangeable with another, even with respect to the present. The features of human life that matter in normative judgments vary like the features of a painting that are taken into account when the *value* (not the price) of a painting is being judged. Procedures can be developed for making the comparisons needed for preferences, but they are valid only with respect to particular times and places and for specified classes of persons — they cannot be culture free or absolute. The procedures that are used to evade judgment with respect to "empirical" conclusions, or matters of "fact", cannot be employed with respect to normative conclusions. Ultimate reliance on formal models is therefore effectively ruled out.

To conclude, a sustained effort to extend and integrate the body of patterns created out of past experience, to formalize it to the fullest extent possible, should certainly be continued, for there is much to be gained from it. And as the overall apparatus grows larger and more complex, its power increases significantly, together with its potential expandability; higher levels of accuracy and reliability are to be expected. That is all to the good. The corollary, however, bears watching, for as such knowledge structures increase in power, they become very much harder to modify; the exceptions that do appear in experience are more readily treated as aberrations rather than indicators of a need for modification in the basic structure. The problem is not new, of course, but that does not make it any more tractable. The antidote or remedy is hard to apply systematically. Though it requires no more than adherence to two fundamental rules. First, so long as logic remains firmly linked to experience and action (avoiding what might be called in our own era the "theoretical physics" syndrome) a fairly simple rule will suffice to maintain system integrity: when logic and experience come into conflict, it is the logic that must yield. Second, the primary concern is always the condition of life of the human population and the basic criterion (and justification) to be applied by those who make decisions is the relative desirability of the human conditions that appear in the projected alternative futures available for choice.

Notes

(1) For a fuller account of the analytic apparatus, see Eugene J. Meehan, *Reasoned Argument in Social Science: Linking Research to Policy*, (Greenwood Press, 1981), or *The Normative Dimension of Policymaking*, (Greenwood Press, forthcoming).

(2) Milton Friedman, *Essays in Positive Economics*, (University of Chicago Press, 1953).

(3) See the very interesting summary by J. Engelkamp, "Language and Modes of Thought", unpublished mss, 1988.

Ruth Manor

Dep. Philosophy, San Jose State Univ.
San Jose (USA)

Rational Question Raising

Abstract

The author is concerned with the issue of how a question may arise in a person's mind so that he/she acts to seek its answer. The theory here sketched is written from a philosophical (rather than an empirical) perspective, as it seeks to determine the assumptions which can be used to solve certain puzzles associated with question generation.

Resumo

O objecto deste estudo é como pode uma questão emergir na mente de uma pessoa de forma que esta procure dar-lhe resposta. A teoria aqui esboçada é-o numa perspectiva mais filosófica do que empírica, visando a determinar as assunções que podem ser usadas para resolver certos puzzles *associados com a geração de questões.*

Introduction

In this paper I am concerned with the issue of how a question may arise in a person's mind so that he/she acts to seek its answer. The theory I will sketch is written from a philosophical (rather than an empirical) perspective, as it seeks to determine the assumptions which can be used to solve certain puzzles associated with question generation. For instance, one of the puzzles can be described as follows:

We often feel that some information given to us, whether it is new information or just presented in a new way, raises a certain question. Thus, in writing papers, we often use the expression "this raises the question...". Assuming that the information is expressed by an (ordered) set of assertions, how can it *raise* a question? To answer this we need to explain what a question is, in what sense some information is given to us, and what does this form of question raising consist of. Though the quoted expression suggests that this type of question raising is a semantic relationship holding between assertions and a question, I believe that it is a pragmatic relationship, because whether some information we have raises a question in our minds depends also on our informational goals. If question raising depended only on the information itself, then the same information should always raise the same question in anybody's mind. This, I believe, is false.

So we need to explain what are these "informational goals", and how they are expressed. "Informational goals" consists of information which I seek to have. Presumably, the information I have is expressible by a set (or sets) of assertions which I believe or accept (temporarily) as true, and my informational goals are expressible as a set of questions, whose answers I am determined to have (as beliefs or assertions which I accept).

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 176/188
1992, APDC & SPF, Lisboa

Hence, the claim above, that question raising depends on a person's informational goals, amounts to the claim that questions do not (cannot) arise in someone's mind merely from a bunch of assertions, but only if they in addition have other questions the answers to which they are already seeking[1].

This view is in accordance with Collingwood (1939, 1940), who claimed the primacy of questions over assertions. Assertions are viewed as answers to questions, and hence the logic of assertions could be reduced to the logic of questions. Questions, however, could not be similarly reduced to assertions. In the present context this means that certain kinds of questions occur in our minds automatically whenever we process information. One obtains and understands information only in answer to questions, so that even when they are not currently involved in seeking an answer to any particular question, the encounter with and processing of new information involve the automatic occurring of certain questions.

These questions concern the issues of what in our world view is the new information relevant to, whether and in what ways the information is new, how it is comprehended or interpreted, and how true or reliable it is. Unless we have some answers to these questions, we cannot be said to possess the information. In other words, I assume that these questions occur automatically, when we obtain new information. These four questions correspond to Grice's four Maxims of Conversational Cooperation (1975). As we shall see below, the explanation of how these questions arise rests on a view of how we organize and manipulate information. The role of the above questions in information organization can be viewed as explaining why the Gricean maxims govern cooperation.

The view of how we organize and manipulate information, as well as our theory of question generation, should be consistent with our view of scientific information manipulation, and scientific question generation. In this framework another puzzle associated with question generation is brought to light, concerning the relationships between questioning and creativity. Creativity is often associated with the ability to come up with new questions, not just new answers (Dillon, 1982). It has been argued, for instance, that computers will not be capable of intelligent thinking because even if they could be great answering machines, they cannot come up with the right questions, as we sometimes do (Leibowicz, 1978). Again, while scientific progress is often attributed to cases where we have found answers to current questions, scientific revolutions and conceptual leaps were often attributed to the ability to formulate new, good, and interesting questions. The intuition underlying such claims, I believe, is that since questions indicate what could count as their answer, what is a good answer is restricted and determined by the question, while questions are not similary restricted. Given certain information, we can probably construe it as answering a wide range of types of questions, and it is not at all clear how the criteria for evaluating questions are determined.

In a recent paper (Manor, 1986) I have suggested a model of the process by which a person, a scientific community or a machine may come up with an original idea or a new theory, given that they have multiple sources of information, a memory and some deductive abilities. The suggested model is close to Minsky (1975) frame theory and to Kuhn (1962). More specifically, the paper supports the Popperian (Popper, 1959) view that science neither employs nor needs to employ inductive reasoning in order to come up with new generalizations. I will not repeat the arguments presented there, which are not necessary for the understanding of the present issues. However, there are two points of relevance. First, that paper deals with the creative aspect of coming up with a new hypothesis – i.e., an assertion – and does not deal with questioning. Yet it does suggest that we search for new

hypotheses to remove contradictions from our information. The present paper supplies the missing discussion of question ocurrence, including an explanation of how questions arise when one discovers inconsistencies in their beliefs.

The second point of relevance concerns the adequacy conditions of the theory. I want to deal with question generation and creativity within the same restricted assumptions, i.e., that we have a constant flow of information from a variety of sources, that we have a memory and some deductive capabilities. In this context, it is often useful to talk in terms of a metaphorical machine (as in the earlier paper), of which the restricted assumptions hold, to ensure that we do not assume additional abilities.

Any attemp to provide a theory of creativity faces a problem in that as a theory, it should have predictive force, but as a theory of creativity, what it attemps to predict is unpredictable. Indeed, it seems that most often we will be able to predict that certain questions will be asked. However, we may also predict that certain unexpected and unpredictable questions will be asked, without being able to predict what these questions are. It is this type of a theory that I suggest.

Concerning predictability, note that there may be physiological causes for certain questions to occur to us. For instance, a feeling of hunger may cause a question like "what should I eat now?" to occur in our minds. Alternatively, a question may be caused to occur because of informational causes. For instance, the information in an article may typically cause certain questions to occur to the readers even when the article does not mention or imply it. Thus, to some extent we can predict that certain questions will occur to people in certain situations. In the following I will be concerned only with these "informational causes" and not other causes for questions occurance(2).

To account for both the predictable and creative aspects, I distinguish between the *occurrence of a question* in my mind, and *my raising* of that question. The first is not an action of mine. It includes the processes by which the question is brought to my attention. Once it occurs to me, I consider it and choose whether to raise it or not. My raising of the question is the action involved in the decision to look into the question and to exert further efforts to find its answer.

Question raising is like the action of asking-oneself a question. Often when we are interested in certain information and we seek an answer to a question, it may be rational of us to ask someone else the question. However, this asking as a communicative action may promote or demote our goals independently of whether we seek the answer to the question. Thus, when I do seek an answer to a question it may be wiser, sometimes, not to ask it directly. In addition, we may ask a question even when we are not interested in this answer, but we believe that the asking would further some other goal. Since I wish to concentrate here on questioning in the sense of the mental activity of answer seeking rather then the social interaction of asking a respondent certain questions, I refraine from talking about asking but only of question occurrence and question raising(3).

One motivation for this distinction lies in my hope that sometime we may be able to come up with theories that will help us raise better questions. Though we may not be able to *cause* directly the occurrence of better questions, we may be able to *guide* the decision to raise better questions. The present paper relates to this wish only indirectly, in raising the Rational Question Raising (RQR) problem, and in a very preliminary attempt to characterize criteria for the RQR, namely,

RQR: Given a certain situation, which questions is it rational to raise?

First, a question is "semantically good" to raise by a person if he is interested in its answer, in the sense that either it will get him closer to a answering another (semantically

or pragmatically) good question, or it will help reorganize his information (or world view) better. Second, a question is "pragmatically good" to raise in situations in which it is both semantically good and we can expect to obtain a satisfactory answer to it with a reasonable amount of effort.

So far I have described the general motivation and background of the theory I will outline below. The theory is not intended as a general theory of questioning, in the sense of Dillon or van der Meij, but rather it is an attempt to address only some of the issues associated with questioning, which I believe are important. As the discussion above is rather intricate, let me summarize the main objectives of the theory from the more abstract to more specific goals, using the distinctions introduced above:

0. One of the tasks a general theory of questioning should aim at is to direct us how to raise "good" or "better" questions. The present theory attempts only to provide preliminary criteria for the Rational Question Raising problem, by characterizing what a "good" question is.
1. The theory accounts for both the predictive and creative aspects of questioning, by explaining how a certain question is caused to occur in the mind (predictive aspect) in certain situations, and how we may choose whether to raise it or to ignore it (creative aspect).
2. The theory accounts for question occurring and raising in terms of information organization and processing on both the individual and the scientific level (without assuming that we have an inductive inference mechanism).
3. The account given for question occurring and information processing supports both Collingwood's (1940) view of the primacy and the irreductubility of questions, and Grice's (1975) view of the logic of conversations.
4. The theory relies on an acceptable view of the logics of questions, answers, and dialogues.

This last point has not been mentioned and will be discussed in the next section.

Concepts from the logic of questions, answers and dialogues[4]

It will be useful to distinguish an *interrogative sentence* of a given language, which is a syntactic entity, from a *question*, which is a semantic entity, and is whatever interrogative sentences express when they are normally used. *Asking, raising, mentioning*, and the like, are pragmatic concepts as they refer to the various uses of questions. Thus, for instance I may ask my son whether he is hungry, by uttering the interrogative: "Would you like to eat anything?"

In general we assume that (indicative, interrogative, etc.) sentences have a syntactic structure and that propositions, questions, etc. have a semantic (or logical) form, so that the meaning of a composite sentence is determined by the semantic meanings of its components as well as the context of its utterance. The connection between the syntatic structure of an interrogative sentence and the logical form of the question it expresses is that the syntactic structure of the sentence together with the context of its utterance determine the logical form of the question. In the following, I am concerned mainly with questions and their logical forms, rather than with interrogatives and syntactic structure (although of course, I will need to use English interrogatives to express questions).

Atomic (i.e., non-composite) questions are divided into two types according to their logical form, alternative questions and Why-questions. Alternative questions offer the respondent a set of alternatives to choose from. Any one of the alternatives is a direct

answer to the question. "Yes–No" questions are alternative questions requiring a choice among propositions, while "which–questions" like

1 *Which suspect killed Mrs Jones?*

require that the respondent choose among the suspects those who did the killing. Why–questions, like

2 *Why did the butler want to kill her?*

require explanations as direct answers. Without going into details, let us say that an explanation provides (explicitly or implicitly) the information necessary for the respondent to derive the explanandum (Hemple & Oppenheim, 1948; Bromberger, 1970). Thus, a direct answer to (2) would consist of a set of claims which would allow the respondent to derive that the butler indeed wanted to kill her. Atomic questions (i.e., why– or alternative–questions) can be combined into complex questions using connectives and quantifiers. For instance, certain questionnaires are (ordered) conjunctions of questions and there are various kinds of conditional questions.

Questions can be informative, and this idea is incorporated in the concept of presupposition. A question semantically presupposes that it has a true answer. For instance, the which question (1) semantically presupposes that some suspect did indeed kill Mrs. Jones; and the why–question (2) semantically presupposes that the butler wanted to kill her (and that that is derivable from certain assumptions). The act of asking a question pragmatically presupposes that the respondent is (or, in a sense, should be) in a position to provide a true direct answer to the question. Thus, asking the which–question (1) pragmatically presupposes that the respondent knows, or is in a position to point out, which suspects killed Mrs.Jones; and asking the why–question (2) pragmatically presupposes that he knows, or is in a position to know, how to derive that the butler wanted to kill her. In general, by asking a question, the speaker expresses his commitment to the truth of its presuppositions. Conditional asking of a question on condition that a presupposition holds, however, serves to "dis–presuppose" the question's presupposition. For instance, my asking you the unconditional question

3 *What is your wife's age?*

semantically presupposes that you have a wife, and pragmatically presupposes that you know her age; the conditional question,

4 *If you are married, what is your wife's age?*

does not presuppose that you are married.

The concepts of 'direct answer' and 'presuppositions', may be sufficient to describe ideal communication, when language is used literally. Strictly speaking, if I ask you a question and you give me less than a true direct answer to it, then you are uncooperative. If I ask you a question whose presupposition is false, then I acted uncooperatively. Of course, if I believed that the presuposition is true and you gave me information which is as close to a true direct answer as you could, then our responses will be viewed as involving innocent mistakes rather than being intentionally uncooperative. Hence, in actual conversations, participants are normally happy if they get what they can. Except in some contexts (e.g., legal cross examination or tests) presupposing falsehoods is not such a great sin, and it is correctible by a "corrective answer" like,

5 *Nobody killed her, she's alive and kicking in Memphis.*

In the following I will include corrective answers in "direct answers".

This leads to the distinction between responses which satisfy a request and those which satisfy a person. When we ask a question, the request is satisfiable only with a true direct answer. If, however, the respondent is unable to provide a direct answer, the questioner

should be satisfied with any answer which may help him find the answer. For instance, answering the why-question (2) by a suggestion of a hypothesis which, if true may serve as an explanation, such as

6a *Maybe the butler was her heir.*

may help the questioner find the right explanation or eliminate a possible explanation. Similarly, an answer which helps narrow down the number of alternatives, should satisfy the questioner. For instance, the responses (6b–6d) express eliminative answers to the "Whodunnit" question (1)[5].

6b *Whoever had a motive.*

6c *Mr. Jones didn't do it.*

6d *One of the children did it.*

We desire, of course, true information and true answers (rather than merely direct answers) to our questions, because only true information can reliably guide our choice of action. However, we are aware that we often obtain false information, and that there is no way to check whether the totality of our beliefs are true. Hence, we assume that new information is true unless we have reason to doubt it, that is, unless we have information to the contrary. Thus, although what we desire is the Truth (with a capital 'T'), what we seek is maximal consistency of our beliefs, and we accept as tentatively true that information which is coherent with whatever we have already somehow accepted.

Now, consider what happens when we discover a contradiction in our belifs. This situation is intolerable and requires some decision and action on our part, even if all we decide to do is to ignore the contradiction and pretend that we did not discovered it. We discover a contradiction when we realize that some part of our beliefs contradicts another. To restore consistency in our beliefs, we may replace the conjunction of the conflicting parts by their disjunction, representing that the two parts are alternatives. Thus, the occurrence of the alternative question ("Which of the conflicting views should be retained?") is caused by the discovery of the inconsistency.

For example, suppose that we accepted that the murderer must have had a motive, that only Mr. Jones had a motive, and that Mr. Jones is not the murderer. This contradiction is intolerable, so we need at least to rearrange our information as a disjunction, e.g., that the murderer had a motive, only Mr. Jones had a motive, *or* Mr. Jones is not the murderer. The need to formulate a set of alternatives is the occurrence of the alternative-question, i.e., the question which of these alternatives are to be accepted and which should be rejected. Whether we choose to raise this question and actually seek its answer by finding grounds for choosing among the alternatives, is open and depends on our own will.

We now sketch the main points of the theory of question occurring and question raising.

Information Organization

Let us assume that we have a constant flow of incoming information from multiple sources, a memory, and some deductive capabilities.

Since information is necessary to guide actions, and to interact with an ever-changing environment, information in the memory is systematically organized and new information is systematically added to it, to make retrieval maximally effective. Our informational goals, that is, the need for information about the environment causes the question

7 *What can I do now to promote my goals?*

to be constantly on our minds; and in attempting to answer it, questions like

8a *What is going on?*
8b *What is happening?*
8c *What is the significance of all this?*
or simply
8d *What?*
occurs constantly. This means that we constantly record, process, and store new information as answers to these questions.

Given these assumptions, the issue we face is when will new questions be caused to occur. There are two such situations. First, during the process of storing certain questions occur "automatically", and one may choose to raise them. Second, given that a question is raised, attemps to retrieve its answer may lead to raising further questions. But before we consider these processes, let us describe how the information is organized. There are, in general, a static and a dynamic aspect of the organization of the information. The static aspect is that at every moment, the contents of memory is organized to make retrieval maximally efficient. The dynamic aspect is that the constant flow of incoming information may affect the memory contents as well as its organization. Namely, both the informational contents and informational organization may change.

Let us first consider the static aspect. Systematic storage is achieved essentially, in the way that we store scientific information, i.e., in terms of theories. If a statement A entails another, B (and not conversely), then A is more informative then B in the sense that I do not have to keep B in the forefront of my mind in order to retrieve it quickly. Rather, I can retrieve it from A by deduction. Hence, a scientific theory consists of "basic principles" which are used to derive less basic principles. Similarly, I assume that we have a primary memory containing "basic assertions" as well as derivation principles. In addition, we have a "focusing" mechanism, which contains the information we focus and concentrate on, so that when we suddenly remember something and bring it "from the back of our mind to the forefront" that means (in terms of the present metaphor) that it is brought into focus. Finally, we have a "secondary memory" containing details. Both storing and retrieving is done in terms of focusing first on the primary memory, and using it to bring into focus fragments of the secondary memory. Ideally, information processing would be most efficient if the primary memory would be so reduced that we could have it all in focus at once. However, our information is so vast and fragmental, that it is divided into sections or "relevance areas". Thus, processing involves determining which relevance area to bring into focus; then, when the relevant section in the primary memory is in focus, it points to the relevant sections of the secondary memory, which can also be brought into focus.

Relevance depends on a deductive and on an "associative" relation. A piece of information is relevantly related to another if one can be used to deduce the other. For example, basic generalizations in the primary memory are relevantly related to those items in the secondary memory which they entail. Alternatively, bits of information may be relevantly related even when they are not deductively related, if they are about the same thing. Thus, though John's being tired and his being a redhead may not be deductively relevant, they are associatively relevant in the sense that both are about John (Manor, 1982, 1987; Harrah, 1985).

Understanding information is related to relevance, in the sense that ideally I understand it if I know what information in my memory it entails, what entails it, and what it is about. But there may be two situations in which some information is incomprehensible to us. One is when we cannot determine its aboutness or its deductive relation to the rest of our world view. The other is when it is relevantly related to some one section in our world view but

contradicts another. Since we aim at possessing true information to guide our decisions, and since an inconsistent theory can neither be true nor an effective guide in decision making, we are faced with the problem of figuring out what went wrong. It is this case that reflects the dynamicity of informational contents and organization of our memory.

We constantly obtain and derive new information which is immediately processed, stored and used in parallel to any other processes we are involved in. Thus the memory contents changes dynamically. Accepting a new basic principle or finding a new deductive relation may cause us to move items from the primary to the secondary memory and vice versa, corresponding to the case where we find a new formulation of a theory. Thus, the contents of memory components may change dynamically. Finally, when as a result of new information or deduction we discover inconsistencies in our beliefs, we need to reevaluate our information to find out "what went wrong" or how to make our total world view maximally consistent. This process may involve major reorganization of our beliefs, including the adoption of new relevance classifications (Manor, 1986).

To summarize, the principles governing information organization are as follows:
Informativeness. The goal is to store maximum information in order to make retrieval as effective and efficient as possible.
Relevance. Information is classified into relevance areas which facilitate focusing.
Perspicuity. Comprehending information consists in knowing its deductive relation to the rest of our beliefs. Knowing these deductive relations is essential to determining informativeness, relevance, and consistency of the information.
Truth. As suggested earlier, the truth (coherence) of some information is evaluated by its consistency with our world view.

These principles are similar to and can be used to explain Grice Maxims of conversational cooperation. Grice (1975) attempts to explain the phenomena of 'conversational implicatures' by postulating that conversations are governed by a Principle of Conversational Cooperation. The principle is conventional, and normally we follow it and assume that others do the same, but sometimes we elect not to follow it. The Principle includes four maxims, calling for a participant's contribution to a conversation to be relevant, informative, perspicuous, and truthful. Grice explains successful communication of an implicature as follows. If in a conversation a speaker appears to act uncooperatively but we have no reason to assume that he intends to be uncooperative, then we will look for a hypothesis which explains why his apparently uncooperative speech-act is actually cooperative. If the hypothesis matches the speaker's intentions, then we have grasped the implicature.

Suppose, for example, that we are involved in a juicy conversation, happily gossipping about Jim, when you suddenly respond by a comment on the wonderful weather. That your response is utterly irrelevant suggests that you are acting uncooperatively. This forces me to look for an explanation for your unexpected behaviour. Noticing that Jim has just joined us, I conclude that your comment, which was irrelevant to the topic, was relevant to the situation, because we do not want to share this conversation with Jim. If this is the conclusion you intended me to reach, then I have grasped your implicature.

The principles of information organization and processing are related to Grice's four maxims in that during the storing and retrieving of information, the questions concerning the informativeness, relevance, comprehensibility, and truth of the information are caused to arise. This view explains the Gricean maxims. Informativeness, relevance, perspicuity, and truth mark conversational cooperation because they also guide the way information is organized in our memories. It should be stressed that what I call here "storing and

retrieving'' of information does not mean merely that we type the information or read it. In order to store or retrieve some information, I have to understand it first. For example, in our gossip session I first understood only what you literally said about the weather, without understanding the implicature, and the question of whether your comment was relevant to the discussion, arose. If I had not expected you to behave cooperatively, I would conclude that your comment was irrelevant, and I would have stored it as such. Since I did expect you to be cooperative, I assumed your comment to be relevant to the conversation and I raised the question of what in the conversation it is relevant to. Whatever answer I give to the initial relevance question, it had to be answered before I could feel that I understood your comment and before I could store it. Namely, the questions concerning relevance, informativeness, perspicuity, and truth should not be viewed as side issues which are somehow associated with the information we possess. Rather, only when we can answer these questions can we say that we really grasp that bit of information. This mechanism can explain how we may choose to raise questions during storing and retrieving, as well as what a good question is.

Question raising in storage

Whenever we encounter an information bit to be stored, we face the following questions which arise as a consequence of the principles of information organization. First, ''what is it relevant to?''. The answer suggests relevance areas to which it may possibly belong. Second, ''To what extent is the information bit informative?'' i.e., whether it enables us to deduce anything that was not retrievable before. Third, ''How can we understand it within the current theory?'' i.e., which of our beliefs imply or are implied by it. Fourth, ''How true or reliable is it?'' i.e., which of our beliefs, if any, contradict it or are in conflict with its assertion.

Normally, given that we want to store an information bit, we are able to answer these questions immediately to our satisfaction, and we store the new information in the right place. In some cases, however, we may encounter problems in answering these questions immediately. Either there is some problem in understanding it, finding its aboutness and deductive relations; or we understand it but it conflicts with our previous beliefs. We are then faced with a choice: we can ignore the problem or raise new questions. For instance, if the new bit is incomprehensible, we can store it in an area reserved for incomprehensibles. We can also raise the question of how to understand it and say ''I don't understand'', ''How is it relevant?'', or ''What are you talking about?''. Alternatively, we can consider some of our beliefs about the same topic and raise questions concerning their deductive relationships to the new bit. If the new bit is found to contradict our old beliefs, we can either store it as unbelievable or we can decide to raise the alternative question, which of the old beliefs or the new bit should be accepted[6].

As suggested in a previous section, when we discover a contradiction in our beliefs, then since this is intolerable, we will rearrange our beliefs so that the contradiction is removed, at least by replacing the conjunction of contradictory beliefs by their disjunction, causing thereby the corresponding alternative- and why-questions to occur, i.e., ''Which of my previous (relevant) beliefs is false, and why?''.

For instance, suppose that at some point we have accepted that the butler had an alibi for the time of the murder and hence the husband and the maid are the only real suspects of Mrs. Jones' murder. Now, you discover that the fingerprints on the gun were the butler's, from which (togheter with the rest of your expert detective knowledge) you

conclude that the butler did it. But this contradicts your belief that one of the "real" suspects did it. This causes the question,

9a *Which bit of (relevant) information is unreliable, and why?*

to occur in your mind, which if you decide to raise may cause you to entertain several hypotheses. You can then choose to raise some of them, for instance,

9b *Has the butler's alibi been faked, or the murder time miscalculated?*

Raising and pursuing this question may lead you to raise further questions, e.g., the conditional Why–question

9c *If the alibi supplied by the maid was faked, why did she lie?*

Note that another detective may consider other hypotheses and hence raise different questions. Which hypotheses come to our minds and which questions we raise depend on our beliefs, deductive abilities, and our relevance classification systems. The point here is that the more difficulties we encounter in storing the new information, as long as we are still determined to overcome these difficulties, the more irregular or uncommon the questions we raise will be. It is at this stage that new ideas (hypotheses) occur to us, and our creativity depends on our tendency to raise or not raise and pursue certain questions. Once a stubborn contradiction has been discovered, there is no limit as to which beliefs could be questioned. A conceptual or scientific revolution is involved when after discovering a contradiction, we hypothesize (and raise it as a question), new principles of information classification into relevance areas[7].

To raise a question is to decide to make an effort to find its answer. This means that we entertain hypotheses almost randomly and try to formulate more and more specific questions, until we find a satisfactory answer or we decide not to exert any more effort in pursuing the answer. Now, when we seek an answer to a question, new questions arise which again we may choose to raise. This is the case of question raising in retrieval.

Question raising in retrieval

Suppose that I decide to raise the question

10 *Who killed Mrs. Jones?*

Initially, I will consider and focus on certain relevance areas; everything I know about Mrs. Jones, about the circumstances of her death, and the generalizations I believe about murders. Since I cannot retrieve its true direct answer from memory, I have to look for clues. Suppose for instance, that I believe that every murder has a motive. So the questions

11 *Why would anyone to kill (this dear old lady)?*

and

12 *Who had a motive to kill Mrs. Jones?*

should rationally arise. The answer to these questions may not provide a direct answer to the original whodunnit question, but it is an eliminative answer which may serve to eliminate some of the alternatives. More generally, we can say that given a question Q1, which is not directly answerable, together with the assumption that its true direct answer implies a true direct answer to another question Q2, then Q2 is a semantically good question.

That the question is a semantically good question does not mean that it is pragmatically good, namely, that it is a rational to raise it in a given situation. For a question to be pragmatically good, it must be semantically good and its pragmatic presuppositions need to be justified. That is, we must believe that we can answer it with reasonable efforts. For, suppose that I don't know who killed Mrs. Jones but I assume that the fingerprints on the

gun were the killer's.

Then, the question

13 *Whose fingerprints are on the gun?*

is semantically good. But it is pragmatically good only if I believe that I have a better change of finding its answer than the answer to the original whodunnit question.

The situation described in this fingerprints example is the converse of the previous example about motive. Given a question Q1 whose true answer is not directly retrievable, and the assumption that the true direct answer to another question Q2 implies a true direct answer to Q1, then Q2 is a semantically good question.

Q2 is a semantically good question, given that question Q1 is raised and that Q1 entails Q2 (or that Q2 entails Q1). Q2 is a pragmatically good question to raise (it is rational for us to raise it), if

1 Q2 is semantically good.
2 There is an interest in exerting effort to answer Q1.
3 An answer to Q2 is expected to be more easily retrievable than an answer to the original question Q1.

Let us go back and reconsider information storage. As mentioned, if there is a problem in answering the Gricean questions, the non-specific question "What went wrong?" arises. Attempts to answer it consist of formulating hypotheses and considering the resulting Which-questions. Since a true direct answer to the Which-question entails a true answer to the "What-went-wrong" question, the Which-question is semantically good. If in addition, the more restricted question is expected to be answerable, then it is pragmatically good. This means that in the face of a puzzle, a contradiction, or a paradox, any relevant question which we raise is semantically good. This explains the feeling that the non-specific questions are open, unbounded, and unrestricted. The key to solving these puzzles lies in raising the right questions. In facing such puzzles, we should rationally raise answerable questions, rather than those we somehow wish we could answer, but don't know how to.

Is the rational-question-raising question a good one?

The main idea here is that there are constantly lots of questions which are caused to occur and are brought into focus. Of these we choose to raise only a few, and our choices are rational if the questions raised are good. This distinction between question occurrance and the raising of a question suggests why I have refrained from using the 'question generation' term. 'Question generation' suggests to me that the whole process resulting in our pursuing a question is involuntary and that, given any question, there is an effective procedure for generating that particular question in any mind or system. On the present approach, questions are caused to arise but, given a particular question, we may not be able to predict whether it will or not arise in some mind. Since we differ in our deductive abilities and classifications, and since both change dynamically, different minds or the same mind at different times will in similar situations cause different questions to arise. In addition, I assume that, normally, many questions arise at once, rather than one after the other. We do not evaluate each question raised on an absolute basis but we compare them, relative to our goals and expectations. This explains another aspect of creativity, namely, that a question may arise in the minds of two persons but only one of them will raise it.

Hence, on the present account our ability to ask interesting questions about a topic depends on the occurrance of a multiplicity of relevant questions, and possibly, an

optimism as to which of these questions is answerable with reasonable effort. The choice element, I believe, is extremely important, for no question is absolutely good, but good relative to other questions and our ability to answer them. Thus, though the non-specific question "What is a good question to pursue now?" arose in the minds of scientists and philosophers for centuries, its general form, the rational-question-raising <RQR> is only now becoming a good question to raise. First, because we are beginning to have theories concerning questions, answers and their rational uses. Second, because the current technological development makes it worthwhile to pursue it. Within this framework, *the RQR question raises the question of when a machine, satisfying certain conditions, should raise certain questions.* This question can be made more or less specific depending on the conditions that the machine is supposed to satisfy. Hence it is answerable, i.e., a pragmatically good question.

Notes

(1) This is possible in contrast with other views, see e.g., Dillon.

(2) The importance of a theory of questioning that has predictive power has been stressed by Flammer. The present paper is not in conflict with his views. It does imply, however, that since questioning is associated with creativity and originality, the predictability of questioning is limited.

(3) As suggested above, I believe that the predictability of questions is limited. However, note that if one wishes to claim that the question we raise are in principle completely predictable, one can use the above fact (i.e., that we do not ask all question we raise and that we ask questions which we do not raise) to explain the high degree of failure to predict questions asked.

(4) The present discussion includes points which appear in earlier publications, listed at the end of the paper.

(5) It was suggested to me by van der Meij that there are empirical results showing that actual questioners often presuppose a falsehood in a cooperative manner, and by Dillon that empirical studies show that questioners are happy with any tolerable response. The points I am making here is not empirical but evaluative: that these behaviors are in certain cases rational and justifiable from the point of view of the person's informational objectives.

(6) I am not suggesting, of course, that we discover all the contradictions in our beliefs. The above problem exists in so far as we are aware of the contradiction.

(7) See Manor, 1986, for more details.

References

Bromberger, S. 1970 *"Why-questions"*, B. Brody (ed) *Readings in the Philosophy of Sciences*, Prentice Hall, 66–87.

Collingwood, R.G. 1939 *An Autobiography*, Oxford University.
1940. *An Essay on Metaphysics*, Oxford University.

Dillon, J.T. 1982 *"Problem finding and solving", Journal of Creative Behavior* 16, 97–111.

Grice, P. 1975 *"Logic and conversation"* P. Cole, J.L. Morgan (eds) *Syntax and Semantics* 3, 41–58, Academic Press.

Harrah, D. 1985 *"A logic of message and reply", Synthese* 63, 275–294.

Hemple, C.G., and Oppenheim, P. 1948 *"Studies in the logic of explanation", Philosophy of Science*, XV, 135–75.

Kuhn, T.S. 1962, *The Structure of Scientific Revolution*, University of Chicago Press.

Leibowicz, 1975 *"Human and artificial intelligence"*, presented at the Tel Aviv University Collooquium.

Manor, R. 1981 *"Dialogues and the logics of questions and of answers", Linguistische Berichte* 73, 1–28.

1982 *"Answers and Other Reactions", Theoretical Linguistics* 9, 69–94.

1984, *"Dialogue representation", Topoi* 3

1986, *"Simulating Imagination", Logos* 7, 69–81

1987, *"Harrah's theory of questions", Questioning Exchange*, forthcoming.

Minsky, M. 1975 *"A framework for representing knowledge"*, P.H. Winston (ed), *The Psychology of Computer Vision*, McGraw–Hill.

Popper, K. 1959, *The Logic of Scientific Discovery*, Basic Books.

Rescher, N., Manor, R. 1970 *"On Inferences from inconsistent premisses", Theory and Decision* 1, 179–219.

3

INTELIGÊNCIA NATURAL & INTELIGÊNCIA ARTIFICIAL

NATURAL & ARTIFICIAL INTELLIGENCE

Roland Fischer

Former Georgetown and G. Washington Universities,
Washington D. C., and Lecturer with
Johns Hopkins Univ., Baltimore (USA)

Neuroepistemological Aspects of Natural and Artificial Intelligence

Abstract

Realities are models of (living) matter's self-experience. Through an ongoing process that took about 3 billion years, matter evolved to its own awareness through the self-experience of the brain that we have in mind. The perceiving-observing mind - mostly a wave aspect of matter - transubstantiates the a-temporal and a-spatial sub-stance of an indivisible universe to substance with form - that is, mostly a particle aspect of matter. Hence, perceiving-observing species generate a variety of models of the real whenever they make a (literally) substantial categorical observation. The variety of species represent matter's intelligence for solving problems of self-experience: environmental events are to be specified as internal states that are (in mammalian brains) the neuronal system's own symbols, as these states stand in regular relation to events in the world and signify potentials for action. The conscious mind is a domain of internalized motion or self-reflexive knowing to make decisions about "the next move"; it is insight into oneself as a moving experience. Intelligence may be re-cognized as a co-evolutionary process, creating perceptual-conceptual worlds, and enabling matter to experience itself. Individuals, then, are spatio-temporal localizations of intelligence distributed over a hierarchy of units (extending from genetic elements to multispecies communities).

Resumo

As realidades são modelos de auto-experiência (viva) da matéria. Mediante um processo progressivo que levou cerca de 3 biliões de anos, a matéria evoluiu para a sua própria consciência por via da auto-experiência do cérebro que trazemos na mente. A mente perceptiva/observante -- preponderantemente um aspecto ondulatório da matéria -- transsubstancia a sub-instância a-temporal e a-espacial de um universo indivisível numa substância com forma - i.e. preponderantemente num aspecto de partículas da matéria. Logo, as espécies perceptivas/observantes geram uma variedade de modelos do real à medida que praticam uma observação substancial e categórica (isto, literalmente). A variedade das espécies representa inteligência da matéria para resolver problemas de auto-experiência: os eventos ambientais são especificados como estados internos que se tornam (nos cérebros dos mamíferos) os símbolos próprios dos sistemas neuronais, por isso que aqueles estados permanecem em relação regular com eventos no mundo e significam potenciais de acção. A mente consciente é um domínio do movimento de interiorização ou conhecimento auto-reflexivo na tomada de decisões sobre "o lance seguinte"; é visão interior dentro de si mesma como experiência movente. A inteligência pode ser re-conhecida como um processo co-evolucionário que cria mundos perceptuais/conceptuais e habilita a matéria a experimentar-se a si própria. Os indivíduos são, portanto, localizações espacio-temporais de inteligência distribuídas segundo uma hierarquia de unidades (extensiva desde os elementos genéticos até às comunidades de multiespécies).

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 191/200
1992, APDC & SPF, Lisboa

Is there anything new under the sun in the cognitive domain? Previously, information was passed along units in form of messages. Now, in general terms, units represent hypotheses, and connections capture inferential dependencies among hypotheses. The interpretation of the processing is not in terms of messages being sent but in terms of self organization: the systems are tuning themselves – adjusting their weights – to become scalars not symbols. Thus the change in emphasis is from symbolic processing to states that reflect the quality of the match between the data to be described and the configurations of the underlying processing structures. In this new, parallel distributed processing (PDP), approach learning takes place through changes in the system itself. The information is reflected in the very shape, form and operation of the processing structures (Norman, 1986, p. 545). Clearly, the information does not come from the outside to be manipulated but it is created[1] through the very shapes, forms and operations of the self-organizing structures: "the medium is the message." Or, to give it a twist with a slightly different slant, the internal states are the neuronal system's own symbolic interpretations (Rothschild, 1958), as these states stand in regular relation to events in the world and signify potentials for action (Werner, 1987).

The units of the PDP system, it should be pointed out, represent hypotheses as symbol structures. Thus if one unit has an excitatory connection to another, this indicates that support for the first hypothesis provides some degree of positive evidence (in terms of strenght measure) for the second. Summation of activation at each unit serves to integrate multiple sources of converging or contradictory evidence regarding a hypothesis. The various learning algorithms allow revision of the inferential relationship among hypotheses. When symbolic information is thus cognized as the system's own creation, any reference to representation becomes superfluous. The idea of representation carries with it the notion of registration, storage and retrieval of "engrams" – that Penfield (1955) had in mind – engrams of "memories that could be reexperienced by electrical stimulation of the temporal lobe" in epileptics but, in fact, such re-experiences were in the eye of the experimenter (be)holding the stimulating electrode, and could not be elicited in non-epileptics by anyone else during the past quarter of the century (Fischer, 1989).

It is apparent by now that the PDP approach does not fundamentally contradict the functionalist approach to cognition (Holyoak, 1987). Any cognitive model, including the connectionist one, can be described at a level more abstract than any particular implementation. The connectionist model (without "representations") and Skarda & Freeman's (1987) neural model, based on EEG research in olfaction and vision (Freeman & van Dijk (1987), accordingly converge in many ways. Both rely on parallel distributed processes among highly interconnected units in interacting networks to produce behavior; both emphasize a self-organized or bottom up, rather than a rule-driven or top-down, explanatory approach; and both rely heavily on organized feedback among components of the system. There is no central processor; learning and memory are functions distributed throughout the neural network. In Werner's (1987) words: history is not represented as a stored image of the past, nor is the present a mirror of the environment. Instead, environmental stimulous configurations are specified by states of neural activity that are the result of the neuronal system's internal organization and dynamics.

Anatomical studies by Goldman-Rakic (1988) confirm the presence of highly integrated but parallel distributed networks in the primate association cortex. Several basic parallel functional systems bridge all major subdivisions of the cerebrum. Synaptogenesis proceeds at the same rate and reaches peak values at the same age in areas of sensory, motor, limbic and association cortex indicating an unexpected degree of integration in

maturational sequence (Rakic, et al., 1986). We are dealing here with developmentally linked and functionally unified PDP networks.

Consider now, that a typical neuron has, perhaps 10,000 synaptic inputs, and the nervous system contains around 10^{12} neurons. Hence brain function may involve the operation of something approaching 10^{16} reinforcing feedback loops with each of approximately 10^{12} neurons exploring, in parallel, the consequences of activating something like 10^{4} reinforcing feedback loops. The pattern of connections formed in the *developing* mammalian brain does not follow a determined scheme but suggests the involvement of competitive growth: neurons invade unoccupied synaptic space in much the same way that growing grasses tend to invade an initially empty field (Schwartz, 1988). During the formation of the brain in the embryo selection among competing neuronal cells and their processes determines the anatomical form and the pattern of synaptic connectivity of the nervous system. During this first stage, which spans the nine months preceding birth, dendrites sprout from the surfaces of neurons, branch out in all directions, and overlap other dendrites, creating a unique network of arbors, that Edelman calls the primary repertoire. Between the 10th to 18th week of gestation the formation of neurons accelerates, and a fierce competition for neurons to *connect* with other neurons and with sensory cells can be observed. Those who could not link up, die off. These interconnections are formed during development, prior to experience, and there is evidence for extensive variance of connectivity, even in genetically identical animals (Edelman, 1988).

A secondary repertoire is in the making during *postnatal* experience: selection among diverse preexisting groups of cells is accomplished by differential modification of synaptic strenghts or efficacies *without change in connectivity pattern*. Responses of neuronal populations are significantly influenced by similarities between present and past constellations of sensory signals. Perception – memory – learning is the consequence of selective amplification which leeds to increased speed or strenght of selected responses when similar patterns of stimulation are repeated. Edelman's theory of neuronal group selection holds that the transformation of the primary repertoire into the secondary repertoire operates according to some of the same principles that Darwin discovered govern the evolution of species.

Thus it becomes plausible that such neuronal substrate could develop onto a vast compartmentalized (modular) *knowledge base* (Klopf, 1982). It takes, then, many years of environmental exposure until a third, a more global, problem solving repertoire is acquired from the brain's self-experience that we have in *mind*. This knowing system can generate alternatives that subserve goal seeking, alternatives that are selected by the environment for fitness.

Theme with variation number one

More than two centuries ago Berkeley and Hume already argued for *associations* – instead of rules – as a basis for knowledge. Today's connectionist networks or PDP networks are rule-less systems. The behavioral stages through which the PDP network passes during learning turns out to be determined by the nature of the system's experience. In this respect its working is analogous to that of a mammalian brain "the only organ that develops (and learns to interpret its interactions as real) by experiencing itself" (Ey, 1963).

Once a PDP network has learned a presentation, the interpretation of what was learned is not a representation but the state of all its neuron-like connections. As mentioned above, the information is the very shape, form and operation of the processing structures, "the

medium is the message.'' A new conceptualization of cognition – and of the body–mind problem as well – follows by changing from symbolic processing to states reflecting the match between the data to be described and the possible configuration of the underlying processing structures (Norman *passim*). Edelman's recognition automata on the other hand, particularly Darwin–II, have the following characteristics which make their intelligence more biological than artificial: their connectivity is set but experience alters preset connection strenght; there are networks which detect local features; and re–entrant interactions between the two networks make associative memory possible. No forced learning, explicit semantic rules, or *a priori* instructions are used (Edelman & Reeke, 1982). Clearly, these authors have taken into account that intelligent biological systems evolved and were not programmed, either as species nor as individuals. The form of Darwinian selection operating among collections of neurons in the brain of a single organism during its life time provides for Edelman the only tenable basis for a theory of categorization and intelligence. Evolution, by extending selection on the way to the individual brain, starting, as we have seen, with the brain of the embryo, removes the need for a program. Indeed, remarks Deary (1988), an Edelman automaton would make as good an immune system as it would a visual recognition system. The epistemology of selectionist theorists, like Edelman, in both immunity and in intelligence/cognition is evolutionary. Perhaps we have arrived here at the biological roots of the hermeneutic circle: the central nervous system–brain (CNS) legitimates reality, that is, the interpretation of its interactions, while the CNS is legitimated by the interpretation. The brain, and a PDP network, or an Edelman automaton, for example, are systems that learn to interpret their interactions as real by experiencing them. This same self–referential circularity is the essence of the hermeneutic circle (Fischer, 1987).

All is well with both types of artificial intelligence systems PDP networks and Edelman automata, but the intention (desire) and the meaning have to be supplied by the designer of these neuron–like nets. Although explicit symbolic descriptions may not exist for many important thought processes, trainable neural nets can ''describe'' such syntactic processes. However, the trainer needs to know a lot about the structure of these processes (Cowan & Sharp, 1988).

Sokolowski (1988) compares the gradual diffusion of writing into human affairs as a historical analogue for the seepage of AI into human exchanges. A major impact of writing was to make new kinds of activities possible. It enlarged and differentiated economic, legal, political, and aesthetic activities, and it made oral history possible. And just as the people who experienced the beginning of writing could hardly imagine what Borges and Bohr could do, or what a national libray or a medical research center or an insurance contract could be, so we – if AI is indeed a renovation of writing – will find it hard, concludes Sokolowski, to conceive what form the flowering of machine thinking may take.

Theme with variation number two: Meaning

What do Cowan & Sharp (1988) mean by saying that meaning has to be supplied by the designer of neuron–like networks? Meaning, in effect, does not have to be supplied if the pattern to be recognized by a system is the outcome of the functioning of the system itself. As it happens, Atlan (1985; 1987) published a computer model of a machine, a pattern recognizer, and the pattern to be recognized is the outcome of the function of the machine itself. Atlan's pattern recognizer is a model of self–creation of meaning. We are confronted here with functional self–organization in the form of a machine that recognizes

strings on the basis of self-organizing criteria. Atlan's model is a machine that generates meaning in terms of its own functioning.[2] What the system perceives is determined by the way in which it perceives, and, closing the hermeneutic circle, the way the system perceives is determined by what the system is capable of perceiving. Do we account for self-created meaning in terms of our own functioning? Do we exist as selves because of self-registration? Very much so! *Sub*-stances become (real objects or) substances by being perceived by a subject. Prior to being perceived, electrons, to give an example, are still insubstantial entities, and the principle of form does not arise prior to the atomic realm. Hence, a five billion light years old galactic explosion has no existence except as observed-percieved, i.e., transubstantiated in the present (Wheeler, 1974).

The creative relevance, significance and meaning (note the increasing arousal-induced tension in these "three little words") of perception is related to both the Anthropic Principle (Barrow & Tippler, 1986), which postulates that the physical world (the "what") must be such as to permit the observer's existence as intelligent observer (who knows "how" to be self-creative), and the quantum principle.

That meaning is self-created through perception-cognition, was already postulated by Basil, the brother of the churchfather Gregory of Nyssa (335-394 A.C.): Try to take away black, white, weight, density, the qualities which concern taste, in one word all that which we see in it, and the substance vanishes. There is nothing for Basil 'behind' appearances (Milbank, 1988). It is these kind of considerations which the Irish bishop Berkeley (1685-1753) was able to develop into a new metaphysics: if "things" are composed of sensory qualitites, then knowledge occurs in the single medium of signs that can be no further explicated than as the reading of signs according to conventions. The new metaphysics is accompanied by a new semantics. If substance drops out, then so also does subject-predicate logic. To attribute something to a thing is now simply to further explicate its definition, which is no longer 'once and for all', but infinitely revisable: encyclopaedic explication replaces dictionary denotation. From a historical vantage point, submits Milbank (1988), the new semantics echoes the world view of the stoics.

Let me now reformulate - step by step - Basil's and bishop Berkeley's concept in the contemporary terminology of neural networks. If substancial things are composed of sensory qualities (self-creation of meaning!), then knowledge (the message) occurs in the single medium of signs ('the medium is the message') that can be no further explicated than as the reading of (neurophysiological) signs according to PDP conventions. Relate this to von Neuman's visionary postulate that "the visual object may constitute the simplest description of itself. It is possible that the connective pattern of the visual brain itself is the simplest logical expression or definition of this principle." Or, in Sejnowsky's words (1986, p. 377), "...the analysis of a single image requires the processing power of the entire visual system; in a sense, the interpretation of the image is the state of all the neurons in the visual system."

Assuming that the universe of matter generates meaning in terms of its own evolutionary functioning (that we now have in mind), we, neocortically endowed humans, are the outcome of this self-created meaning process, and realities (perceived by the variety of living species) are models of living matter's self-experience. Through an ongoing process that took about 3 billion years, matter is evolving to its own awareness through the self-experience of the brain that we have in mind. The perceiving-observing mind, mostly a wave aspect of matter, transubstantiates the a-temporal and a-spatial sub-stance of an indivisible universe to a substance with form, that is, mostly a particle aspect of matter. Hence, perceiving-observing species generate a variety of models of the real (real is that

wich has properties) whenever they make a (literally) substantial categorical observation. The variety of species represent matter's intelligence for solving problems of self-experience. The properties of objects – the primary and secondary qualities – are categories of biological intelligence that are embodied in the observer. Or, in v. Foerster's (1981) words: objects and events are representations of relations determined by modes of internal computations; hence, "objects" appear to reside exclusively in the subject's own experience of his sensory/motor coordination. The externality of communal space is established within the social context of other subjects also recursively experiencing "objects" as tokens of sensory/motor closures. Thus tokens for stable behavior are "tokens for Eigenfunctions" (v. Foerster, 1981, pp. 274, 280; and 1974, p. 830). Recursion, it should be added, enters the consideration whenever the changes in the creature's sensations are accounted for by its movements, and – within a hermeneutic circle – its movements by its sensations. The two accounts taken together form "recursive expressions" of the system (the creature) in terms of these very states. Hence, environmental events are specified as internal states that are – in mammalian brains – the neuronal system's own symbols, as these states stand in regular relation to events in the world and signify potentials for action. The conscious mind is a domain of internalized motion or self-reflexive knowing to make decisions about "the next move"; it is insight into oneself as a moving experience. Intelligence may be re-cognized as a co-evolutionary process: creating perceptual-conceptual worlds, and – at the same time – enabling matter to experience itself. Individuals, then, are spatio-temporally perceived localizations of intelligence, that is distributed over a hierarchy of units, ranging from conceptual and genetic elements to gene pools, organisms, sub-populations, and multispecies communities. And Schull (1988) argues that an evolving species is a better simulation of "real" intelligence than even the best computer program likely to be produced by cognitive scientist for many years. In the evolutionary modeling of intelligence semantic representations are the products of cognitive activity and not the basis for it. Evolutionary models are embedded in both the framework of the organism's *actions* in the world and the ecological environment of the organism, that is undergoing long term very slow changes far from equilibrium. Such a system displays a large variety of intelligences, and human intelligence, in strong association with language, is one of these intelligences.

Intelligence

Intelligence (from the Latin *inter legere* or reading in between) refers to problem solving. Intelligence is an adaptive trait, it is inherited, and its level is correlated with brain size. Moreover, there is a very high correlation between total brain size and the surface of the neocortex. The correlation between intelligence and brain size, if corrected for body size, is called encephalization, that is, the residual capacity that represents additional information that an animal handles beyond that used in routine control of body functions. Encephalization is the phenotypic expression of intelligence (Jerison, 1988). The nature of the evolution of intelligence is an aspect of the evolution of encephalization.

The size of the brain and particularly the cerebral cortex is almost completely determined by its neural elements; hence, by correlating the cortical surface parameters to the number of functional units or modules and the cortical volume parameters to the number of neuronal elements and their *interconnections*, we may have a measure of encephalization (Hofman, 1982).

The principle of PDP implies that activities of ordered sets of nerve cells can be

considered to be mathematical vectors. An important aspect of this vector approach is that it focuses on the explanation of brain function in terms of neuronal networks, i.e., mass action (Freeman, 1980) and that it is, therefore, compatible with the modular organization of the brain (Hofman, 1988, p. 440). The complexity of the neural circuitry of the cerebral cortex can be looked upon as a neural correlate of intelligence. The overall speed by which the circuitry functions may be, then, a *measure* of intelligence. Variation in intelligence as measured by standard I.Q. tests is to some degree a simple consequence of the information processing as measured by reaction time (RT) scores. 300 msec. as a mean RT is the minimum mean time of the various RT tests. The input time (from eye to visual cortex) is about 85 msec., while the output time from motor cortex to finger tip is about 20 msec. (conduction time from motor cortex to wrist is about 17 msec., and wrist to index finger tip about 3 msec.). This leaves 200 msec. for brain processing time (BPT) in the various regions of the brain involved in RT response. It is BPT that is heritable, and may be divided into axonal conduction time (ACT) and dendritic conduction time (DCT). Since RT is negatively correlated with intelligence (I.Q. score), BPT and its components ACT and DCT should also be negatively correlated, and indeed, they are.

It is remarkable that differences in RT so small that subjects themselves cannot discriminate between their own faster and slower trials (Jensen, 1980, p. 691), when consistently observed between subjects, are apt to reflect differences in fine neurophysiological detail. Due to the high neural connectivity, structural differences can lead to physiological differences at the level of mass action. Minute structural differences, according to Jensen, may correlate with global I.Q. because of the dependence of the latter on the presence of the former. Jensen's hierarchical model of this functional dependence accounts for the fact that correlations between I.Q. and reaction time, although often substantial, are less than perfect. Evidently, this has to be so if mental abilities that normally are positively correlated with each other are independent abilities. Gordon (1980) gives a few examples by considering extraordinary mnemonists, idiot savants with spectacular single talents, Turner syndrome patients, whose (non-verbal) performance I.Q. averages 17 points *below* their verbal I.Q., and the unique cognitive profiles peculiar to practically all ethnic groups, which in the case of the profile for Jews, for example, yelds a mean verbal I.Q. 15 points *above* the average performance I.Q. Clearly, not everything that usually goes together need do so all of the time. It is plausible that higher levels of g (general intelligence) assist in creating and then implementing the many specific abilities that are usually found in subjects of higher g and higher I.Q. Apparently, intelligence means a variety of different things, so long as one has enough of it.

But the measurement of intelligence through chronometry does not have to be restricted to studies of RT. One may also use, I am proposing, an evolutionary time measure of intelligence on an interspecies level, namely *lifespan*, that is, the capability of a mammalian species to endure temporarily by beating the 2nd Law of thermodynamics.

Enduring in time involves the use of available energy at a particular rate, and this rate – the metabolic rate – is proportional to body mass $Mb^{0.75}$, and, hence, endurance time, i.e., lifespan is proportional to the ratio of body mass to metabolic rate. The correlation of body weight with basal metabolic rate is about +0.99, that is, life span increases with the size of the animal. The highly significant relation between logarithm of *life span* and logarithm of *metabolic body size*/in $kg^{3/4}$ / (Kleiber, 1961) – for 63 species of mammals, from the shrew to the whale (Fischer, 1968) – hints at a possible relation between intelligence and life span. Indeed, when metabolic brain size/in $kg^{3/4}$ / is plotted against life span, an approximately twofold superiority of metabolic brain size over metabolic body size is obtained as a

predictor of lifespan; a tighter clustering to the regression lines becomes apparent and, specifically, man's closer fit within the relationship. Since man lives three times longer than expected in terms of his body size (Sacher, 1965, p. 276), the closer fit indicates that the human brain, or better, human intelligence, may have something to do with the prolongation of life span. We may now quantitatively express the index of encephalization (w) as a measure of relative brain size that is orthogonal to (i.e., independent of) body size (Schimdt-Nielsen, 1984, p. 149):

$$x=0.636w + 0.198y =0.471$$

where x is log life span, y is log body weight and

$$w= \log/(\text{brain weight})/(\text{body weight})^{2/3}/.$$

Some of the above data may be embedded into a more global, in effect, a cosmic context. The mass of a human is the geometric mean of a planetary and an atomic mass, while the mass of a planet is the geometric mean of an atomic mass and the mass of the observable universe (Barrow & Tipler, 1986, p. 12).

Intelligence too may be redefined within a more global context as matter's self-reflection (or self-perception) in its own image, that is, in the image of an immense variety of (still) evolving species. The variety of individual species may be considered matter's self-reflective function that emerges from its dynamics according to self-generated spatio-temporal criteria. Self-reflection, then, translated in Gibson's terms of 'direct perception' (Michaels & Carello, 1981), refers to the perception of a given pattern through a *resonance* between a given structure "out there" and matter's sensorium, the internal perceptual-cognitive network of its species. The adaptive *resonance* theory of Grossberg (1987) updates Gibson's by describing in differential equations a neural network architecture that can self-organize, self-stabilize, self-scale, and carry out self-adjusting memory searches in an arbitrary environment. Moreover it can build up a world view of (Kantian) categories from individual experiences that grows as it continuously self-stabilizes (Johnson & Brown, 1988). (3)

The identification of intelligence with self-reflection is not a new concept; it may be traced to Leibniz's concept of aperception and Heidegger's "(self-)reflectiveness" of Dasein (Dascal, 1989).

Epilogue

Karl Rahner (quoted by Morin, 1983) declares man to be "the question to which there is no answer." For even if the self explained itself as a series of semi-autonomous computerized systems, it would remain ineluctably the agent of its own judgement. If there were "an identity of spirit and matter," Rahner writes, it would only be "an identity in the thought of the subject; the thought of unity and unity in the thought would still be two different things."

Perhaps it is fundamentally impossible to totally idealize and rationalize the universe. *It just might be more "intelligent" than the mind at its present evolutionary development.*

Notes

(1) That information from "out there" is not processed but created may be impressively illustrated in the perceptual realm of hearing. The inner ear is an activ non-linear filter, and, hence, hearing is based on non-linear dynamics. As a result of the inner ear's anharmonicity, new self-organizing information is being created, that is a new combination of sounds; but the inner ear is also capable of destroying information. Musical sensibility and non-linearity are

tightly coupled phenomena, the latter enabling man to dynamically adapt to sound intensities spanning 12 orders of magnitude. Non-linear phenomena also allow us to extract pattern from noise, that is, perceive pattern in frequency space. Harmony, it should be emphasized, emerges not as a result of ideal, linear harmonic, i.e., physical oscillatory phenomena but through a subtle co-operation of *anharmonic*, non-linear self-organizing processes (Euler, 1984).

(2) Of course, Atlan's model cannot display consciousness. A conscious system would have to know more, namely, that *it* is expressing self-referential propositions.

(3) In Grossberg's adaptive resonance theory (1987) learning and memory are basically identical processes. Raw sense data from "out there" combine with the associations from the past and resonate until a coherent image is formed. The resonance is comparable to what happens in a stringed instrument when one string vibrates in response to the sound of a certain pitch. What resonates in Grossberg's model is a cloud of neurons – an interconnected group of cells surrounded by inter-neurons – that together form an electronic feedback system which resonates in response to a particular class of sensations. Patterns of sensations are captured internally in a resonating neuron cloud whose feedback responds to patterns as they are experienced. Thus the elements of one's ongoing experience of the world are a set of short term memories whose physical counterparts are resonating neuron clouds. The functional unit of short term memory is a resonating pattern (of synaptic weights) that fades in about 15 seconds if not committed to long term memory through association with "remembrances of things past". The conscious experience of daily routine then, is a pattern-matching game between the novelties perceived and the known patterns (Johnson & Brown, 1988).

References

Atlan, H. (1985). Natural complexity and self-creation of meaning. In *The Science and Praxis of Complexity*. Tokyo: U.N. University.

Atlan, H. (1987). *Self-creation of meaning*. Physica Scripta 36, 563–576.

Barrow, J.D. & Tipler, F.J. (1986). *The Anthropic Cosmological Principle*. Oxford: Clarendon Press.

Brues, A.M. & Sacher(1965). *Ageing and Levels of Biological Organization*. Chicago: The University of Chicago Press.

Cowan, J.D. & Sharp, D.H. (1988). *Neural nets and artificial intelligence*. In *The Artificial Intelligence Debate*. Graubard, S.R. (Ed.), Cambridge, Mass.: The MIT Press, pp. 85–121.

Dascal, M. (1989). Personal communication.

Deary, I.J. (1988). *Apllying evolutionary epistemology: from immunity to intelligence. J. Soc. & Biol. Struct. 11*, 399–408.

Euler, M. (1984). *Zur Biophysik des Horens*. Deutscher Volkshochschulverband, Materialien zum mnt Bereich nr. 88. Frankfurt.

Edelman, G.M. & Reeke, N.Jr. (1982). *Selective networks capable of representative transformations, limited generalizations, and associative memory. Proc. Natl. Acad. Sci. U.S.A. 79*, 2091–2095.

Ey, H. (1963). *La Conscience*. Paris: Presses Universitaires de France.

Fischer, R. & Rockey, M.A. (1968). *Psychophysics of excitation and tranquilization from a steady state perspective*. In *Neuro-Science Research 1*, New York: Academic Press, pp. 263–314.

Fischer, R. (1987). *Emergence of mind from brain; the biological roots of the hermeneutic circle. Diogenes No. 138*, 1–25.

Fischer, R. (1990). *Why the mind is not in the head; on single-minded units of society's connectionist network. Diogenes* (in press).

Foerster, v.H. (1974). In *L'Unité de L'Homme*. Morin, E. & Piatelli-Palmerini, M. (Eds.), Paris: Seuil; p. 830.

Foerster, v.H. (1981). *Observing Systems*. Searside, California 93955: Intersystems Publ. pp. 274 & 280.

Freeman, W.J. & van Dij (1987). *Spatial patterns of visual cortical fast EEG during conditioned reflex in a rhesus monkey. Brain Research* 422, 267–276.

Freeman, W.J. (1980). *Mass Action in the Nervous System*. New York: Academic Press.

Goldman–Rakic, P. (1988). *A topography of cognition. PDP networks in primate association cortex. Ann. Rev. Neuroscience 11*, 137–156.

Gordon, R.A. (1980). *Comment on Jensen's Chronometric Analysis of Intelligence. J.Soc. & Biol. Struct.* 3, 123–124.

Grossberg, S. (1987). *The Adaptive Brain I: Cognition, Learning, Reinforcement, and Rhythm*. Amsterdam: Elsevier.

Holyoak, K.J. (1987). *A connectionist view of Cognition. Science 236*, 992–996.

Hofman, M.A. (1982). *Encephalization in mammals in relation to the size of the cerebral cortex. Brain, Behavior & Evolution 20*, 84–96.

Hofman, M.A. (1988). *Brain, mind and reality*. In Jerison, H.J. & Jerison, I. (Eds.) *Intelligence and Evolutionary Biology* NATO ASI Series. Berlin, New York: Springer; pp. 437–446.

Jensen, A.R. (1980). *Bias in Mental Testing*. New York: Free Press.

Jerison, H.J. & Jerison, I. (Eds.) *Intelligence and Evolutionary Biology*. NATO ASI Series, Berlin, New York: Springer.

Johnson, R.C. & Brown, C. (1988). *Cognizers, Neural Networks and Machines that Think*. New York: Wiley.

Kleiber, M. (1961), *The Fire of Life*. New York: Wiley.

Klopf, H. (1982). *The Hedonistic Neuron. A Theory of Memory, Learning and Intelligence*. Washington: Hemisphere.

Michaelis, C.F. & Caretto, C. (1981). *Direct Perception*. Englewood Cliffs, N.J.: Prentice Hall, Century Psychol. Series.

Milbank, J. (1988). *Theology without substance*, Part II. *J. Litterature and Theology* 2, 131–152.

Morin, E. (1983). *Beyond determinism. Sub-Stance* No. 40, 22–35.

Norman, D.A. (1986). In *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*. Vol. 2, Psychological and Biological Models. McLelland, J.L., Rumelhart, D.E. & the PDP Research Group: Cambridge, Mass: The MIT Press.

Penfield, W.C. (1955). *The role of the temporal cortex in certain psychical phenomena*. 25th Maudsley Lecture. J. Mental Sci, 101, 451–465.

Rakic, P. et al. (1986). *Concurrent overproduction in the synapses in diverse regions in the primate cerebral cortex*. Science 232, 232–235.

Rothschild, F.S. (1958). *Das Zentralnervensystem als Symbol des Erlebens*, Basel: Karger.

Sacher, G. (1965). In *Ageing and Levels of Biological Organization*. Brues, A.M. & Sacher, G. (Eds.), Chicago, Ill.: The University of Chicago Press.

Schimdt–Nielsen, K. (1984). *Scaling. Why is Animal Size so Important*. Cambridge: Cambridge University Press.

Schull, J. (1988). *Are species intelligent?* (manuscript to be published).

Schwartz, J.T. (1988). *The new connectionism*. In *The Artificial Intelligence Debate, False Starts, Real Foundations*. Graubard, S.R. (Ed.), Cambridge, The MIT Press, pp. 85–122.

Sejnowski, T.J. (1986). In *Parallel Distributed Processing*. McLelland, J.L., Rumelhart, D.E. & the PDP Research Group, London, Cambridge, Mass.; The MIT Press.

Sokolowski, R. (1988). Natural and artificial intelligence. In *The Artificial Intelligence Debate*, pp. 45–64. (see Schwartz, J.T. (1988)).

Skarda, C.A. & Freeman, W. (1987). *How brains make chaos in order to make sense of the world*. Behav. & Brain Sci 10, 161–195.

Werner, G. (1987). *Cognition as a self-organizing process*. Behav. & Brain Sci. 10, 183.

Wheeler, J.A. (1974). *The Universe as home for man*. Amer. Scientist 62, 683–691.

Hung-Yul So

Dept. Philosophy, EWHA Women's Univ.
Seoul (Korea)

The Artificial and the Popperian World

Abstract

What K. Popper differentiated as the Third World, or World 3, may be characterized as the world of the artificial. Then, the two remaining worlds, that is, the physical world (World 1) and the mental world (World 2) would belong to the realm of the natural. Now, one of the issues in the AI debate is whether the artificial could simulate the mental. It is a quest for the possibility, first on the conceptual level, of creating a true artificial mind, that may stand against the natural mind of humankind. This artificial mind, then, would obtain its objective existence when it becomes feasible on the engineering level, if ever. And this we may call "the objective mind", in contrast with the subjective mind of World 2. This objective mind, unlike the Hegelian, would be a materialistically realized objective mind, and as such it might function as a model for the social mind, which is an issue in social sciences, as well as in philosophy.

Resumo

Podemos caracterizar como o mundo do artificial aquilo que K. Popper diferenciou como o Terceiro Mundo ou o Mundo 3. Logo, os outros dois mundos, i.e. o mundo físico (Mundo 1) e o mundo mental (Mundo 2) pertenceriam à esfera do natural. Actualmente, um dos tópicos de debate sobre a IA é se o artificial pode simular o mental. Trata-se da procura da possibilidade, primeiramente ao nível conceptual, de se criar uma verdadeira mente artificial, oponível à mente natural da humanidade. Esta mente artificial ganharia então existência objectiva se alguma vez se tornasse exequível ao nível da engenharia. Poderíamos, por conseguinte, denominá-la "mente objectiva", em contraste com a mente subjectiva do Mundo 2. Esta mente objectiva teria uma realização material, ao contrário da hegeliana, e, assim, poderia servir de modelo para a mente social, o que abre uma perspectiva nova tanto nas ciências sociais como na filosofia.

Karl Popper's 'World-3' refers to the realm of the artificial. It is the world made by men and animals. The 'inmates' of World-3 include the non-living structures produced by animals, such as spiders' webs, ants' nests, beavers'dams, or animals' paths in forests, as well as man-made products such as houses, tools, and works of art.(1) But the fact that there are artificial objects made by men and animals is not a problem of interest. The significance of the Popperian World-3 lies in its claim for objective knowledge.

Popper makes a distinction between subjective knowledge and objective knowledge. Objective knowledge, or knowledge in the objective sense, consists of the "logical content of our theories, conjectures, guesses, ...and the logical content of our genetic code."(2) They are objective in the sense that they exist independently, external to our mind. They belong to World-3, whereas the content of subjective knowledge belong to World-2, the mental

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 201/205
1992, APDC & SPF, Lisboa

world. World-1 of course is the physical world. World-3, or the man-made world of objective knowledge, then, attains its objective existence in the form of "the logical contents of books, libraries, computer memories, and such like."[3] In other words, objective knowledge in itself is an epistemic entity. It is a product of man's epistemic activity, or knowledge-acquiring activity. But it obtains its objective, independent ontological status when stored in books or computer memories.

Popper's World-3 is also an autonomous world. "It creates its own autonomous problems."[4] The problems in World-3 "emerge autonomously from the field of new relationships which we cannot help bringing into existence with every action, however little we intend to do so."[5] The problems are autonomous because "they are in no sense made by us; rather, they are discovered by us; and in this sense they exist, undiscovered, before their discovery."[6] Here Popper seems to deny the artificiality of World-3, that is, its being a man-made product. But he wants to have it both ways. "It is possible," he maintains, "to accept the reality or (as it may be called) the autonomy of the third world, and at the same time to admit that the third world originates as a product of human activity."[7] Popper does not try to resolve the apparent conflict or paradox between the autonomy of World-3 and human originality or creativity in producing the contents of that world. Instead, he goes on to say that "in spite and also because of the autonomy of the third world, there will always be scope for original and creative work."[8] Popper's insistence on the acceptability of the apparent paradox suggests an interesting expansion to the concept of 'artificial'. It suggests that something could be artificial, and at the same time autonomous.

The world of objective knowledge, according to Popper, is also the world of 'objective mind'.[9] What Popper means by 'objective mind' here seems to be the autonomy of objective knowledge. He does not mean a mind as a knowing subject. He is quite specific about the limit of World-3 in this respect. "Knowledge in the objective sense is knowledge without a knower: it is knowledge without a knowing subject."[10] Popper's notion of objective mind is nothing more than that of objective knowledge.

However, if we grant a knowing subject to the Popperian world of objective knowledge, what we have in effect is something that is artificial, autonomous and subjective. But a knowing subject with its own autonomous contents of knowledge would certainly be qualified to become a mind. So, what we get here is an 'artificial mind', or to be more specific, the notion of 'artificial mind'.

What we have granted for the notion of artificial mind still remains an issue in the artificial intelligence debate.[11] The debate may be considered in two different levels: the conceptual and logical level and the engineering level. On the conceptual and logical level, we need to raise such questions as: what is a knowing subject?; what is an agent?; what is self-concept? Then, in order to meet with these questions we need to specify some of the typical functions that pertain to the kind of qualifications. For instance, pursuing a goal or an objective, and selecting the means for the given end, or becoming adaptive to the changing circumstances, showing flexibility and creativity, etc. might be such functions.

These functions, then, must be further analysed in terms of appropriate logical concepts, so that they may be implemented by certain computational logic. As for the appropriate logical concepts, we may begin with the kinds of reasoning that are known, in ordinary terms, roughly as 'dialectics', 'abduction', and 'analogy'. By no means they are sufficient or exhaustive for the job, but they seem to be good candidates for the work to begin with, for they all stand for sort of informal reasoning. They are non-extensional, and they allow jumps in reasoning, which would be judged fallacious in any formalized

deductive logic.

Dialectical reasoning enables us to account for the emergence of a new synthesis, a qualitative leap from the quantitative changes necessitated by conflicting forces.

Abductive reasoning allows the logical jump to a generalized abstraction based on insufficient, underdetermined factual information. Scientific hypotheses and speculative assumptions in philosophy, arts, and religion all depend on abductive reasoning.[12]

Analogical reasoning helps us in forming abstract mental models. When the object for modeling has a fixed, mechanical structure, the whole structure could be reduced to the lawful relations of its parts. In such a case, then, analogical reasoning in terms of structural models might be replaced by deductive, causal reasoning based on the fixed lawful relations. But for any object with an open structure abstraction through analogical reasoning is inevitable. As such analogical reasoning allows us to understand the world through structural similarities, and allows us to adjust our mental models to the changing structures in the objective world.[13]

Granted such and other non-formalistic types of reasoning are sufficient to account for the functioning of a knowing subject, the problem of implementing them by computational logic still remains an issue. Man's ability to reason dialectically, abductively, as well as analogically comes from a long process of natural development and evolution. It is a product of natural 'engineering'. But the problem of artificial mind, on the engineering level, is whether human engineering will be able to produce the same kind of functional mind produced by the natural 'engineering'. Here again an optimist would view the engineering feasibility as a good possibility, provided that the functional analysis of artificial mind becomes sufficiently complete in the foreseeable future.

Although when and how we can have a truly artificial mind is an open question, we may, nevertheless, raise questions about why we want to have an artificial mind, or what we expect of it. Questions like these seem meaningful because they influence the future course of development in Artificial Intelligence.

The feasibility of artificial mind itself would be a feat of cognitive science. To simulate a human mind by means of human engineering is in itself a goal of our endeavour in cognitive science. But our interest in cognitive science is broader. Cognitive science is expected to bridge the gap between physical science and social science. What differentiates natural science, meaning physical science, from social science is the factor of human mind. It is the realm of mental phenomena that comes in between the physical phenomena and the social phenomena.

The coming of cognitive science into a mature stage, therefore, would help us see the continuity and the discontinuity among the physical, the mental, and the social phenomena. The relation of discontinuity may be seen as the results of a sort of dialectical leap from the lower phenomena to the upper phenomena. It means qualitative changes, or emergence of new properties, new structures, and new functions. The relation of continuity, on the other hand, means the relation of dependence. The upper level is dependent upon the lower level for its existence. It is ontological dependence. And what is supposed here is of course ontological monism. But the relation of continuity may also be seen as a kind of interrelation. For instance, any function in the upper level depends on the functioning of the lower level. The upper level cannot function unless the lower level functions. The dependence here is again ontological. But because of this very ontological dependence, the supporting function on the lower level is, in a way, dependent upon the function of the upper level. To be supportive means to function in accordance with the functional necessities of the upper level. And the same kind of interrelations applies to the inter-level

relations in properties and structures.

Now, Popper's notion of World–3, initially at least, does seem to refer to the objective knowledge of the physical phenomena only. However, when the continuity, as well as the discontinuity, between natural science and social science is made explicit through the mediating role of cognitive science, if ever, then we could have genuine objective knowledge, in the Popperian sense, of the social phenomena. But objective knowledge, according to Popper, is autonomous in the sense that it produces its own problems following its own logic. The possibility of objective social knowledge, thus, implies that the social phenomena, too, exhibit certain internal logic that works along with laws and principles, just as it is with the physical phenomena. But Popper may think this is going too far beyond what he meant to assert by 'objective knowledge', for his idea is to have an 'open society'.

But the idea of an artificial mind is to have an open mind. Though it is a product of human engineering, it is supposed to be creative, selective, adaptive, and purposeful. The question is whether we could have an artificial social mind. What is a social mind? It must contain objectively represented knowledge of the social phenomena. And it must also function as a mind with the so-called 'general will'. It must be free from subjective interests and biases such as the individual minds show. But even a social mind, insofar as it is short of becoming the universal mind, would inevitably show its limits in interests and biases. An artificial social mind could begin with by modeling or simulating a particular human society. Then it may work on to expand itself toward the universal mind. An artificial social mind, in this way, might be developed as a model of 'the ideal observer' or 'the rational mind'.

Simulating a particular individual human mind by means of an artificial mind might be meaningful if the individual mind is deemed worthy of extended functioning beyond personal death. But the notion of an artificial social mind seems to open up a new world of possibilities for social sciences, and maybe even for humanities as well.

Notes

(1) Karl R. Popper, *Objective Knowledge*, 1972, Oxford University Press, pp. 112, 113.

(2) Ibid. p. 73.

(3) Ibid. p. 74.

(4) Ibid. p. 118.

(5) Ibid. p. 119.

(6) Ibid. pp. 160–161.

(7) Ibid. p. 159.

(8) Ibid. p. 161.

(9) Ibid. p. 153.

(10) Ibid. p. 109.

(11) Imaginative and stimulating discussions on the subject are found in all of the articles edited by Stephen R. Graubard in the Artificial Intelligence Debate, 1988, MIT Press. Especially, *"Real Brain and Artificial Intelligence"* by George N. Reeke, Jr. and Gerald M. Edelman, *"The Prospects for Building Trully Intelligent Machine"* by David L. Waltz, and *"Artificial*

Intelligence and Psychoanalysis'' by Sherry Turkle have been informative and relevant to the ideas developed here.

(12) The nature of abductive reasoning, or abduction explained here is taken from C.S. Peirce, which is quite unlike the now popularized notion of abduction, which caracterizes it as ''inference to the best explanation.'' The popular version seems to be a misinterpretation of the Peircian concept, which might be caused by the mis–representation of the Peircian Pragmatism in general.
J. Buchler, ed., *Philosophical Writings of Peirce*, 1955, Dover Publications, pp. 150–154, 304–305.

(13) David Waltz in *''The Prospects for Building Trully Intelligent Machine''*, explicates a new type of reasoning needed for the kind of AI, which he calls 'memory–based reasoning', that is a sort of analogical reasoning.
Stephen R. Graubard, op. cit., pp. 198–200.

Manuel Liz

Univ. La Laguna, Tenerife
(Spain)

Mental Contents: Attributions and Superveniences

Abstract

In this paper, we stablish some ideal conditions for a correct attribution of semantical content to causal types of mental states. We ask what sort of dependence or determination relation can be between these contents and the required causal types. We examine some concepts of supervenience and define a wide variety of other new concepts. The objective is to elaborate an epistemical, no-fundamentalistic and no-causal relation of determination of mental items by no-mental items that can satisfy the naturalistic demands, leaving the problems of definability and theoretical reduction to empirical research.

Resumo

Estabelecemos neste estudo algumas condições ideais para a atribuição correcta de conteúdos semânticos a tipos causais de estados mentais. Indagamos que modo de dependência ou relação de determinação pode existir entre aqueles conteúdos e os tipos causais requeridos. Examinamos alguns conceitos de superveniência e definimos uma variedade ampla de outros novos conceitos. O objectivo consiste em elaborar uma relação epistémica, não-fundamentalista e não-causal, de determinação de itens mentais por itens não-mentais que possa satisfazer as pretensões naturalistas, deixando os problemas da definição e da redução teorética para a pesquisa empírica.

1. To identify mental contents by semantical properties

Usually, we know what our mental contents and the mental contents of other are according to certain semantical properties that we *assign* to mental states in concrete situations. By this way, we know what our mental contents are because we belong to a linguistic community of competent speakers in a language useful to describe mental states according to their contents and in relation to some special empirical evidences.

For the subject X, the mental state M is the mental state that has a *content identical* to the semantical content of the expression with which X can correctly describe it in his language. The semantical properties of a mental state are the same as the semantical properties of those expressions that can adequately describe such a mental state in a certain language. If C is [*the belief that there is a cat on the mat*][1], X will have C if and only if "X believes that there is a cat on the mat" is true.

The former is a very important triviality. The reasons for which "X believes that there is a cat on the mat" is true are, in principle, pretty indifferent to the causal properties of the mental states, if any, that we can be describing by that expression. The appropriate evidence for the first is of a very different kind than the appropriate evidence for the second.

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 206/213
1992, APDC & SPF, Lisboa

At this point, perhaps, we can be tempted to think that if there were a case in which no empirical evidence can determine certain semantical properties (for example, the reference) of a expression, then no empirical evidence could determine the mental contents describable by means of it, and then there could be mental states with a content necessarily inscrutable, impossible to discriminate with any kind and quantity of empirical evidence available. The consequence is that the semantical explanations could be always indifferent with regard to all empirical and causal knowledge. In particular, with respect to any knowledge about the way we become competent speakers of a descriptive language of mental states according to their contents and in relation to certain empirical evidences. So, the semantical inscrutability becomes a *mental inscrutability*.

The philosophers point out three general cases, closely related, where the last problem can be found. They are the existence of alternative, but observationaly equivalent, theories, the indetermination of the reference of any one referential expression, and the compatibility and correction of different dictionaries with respect to any proference of sentences from the translated language. Habitually, we understand these phenomena as meaning that if someone has one of those theories, references or dictionaries, and not other, there can't be any fact of the matter by means of which we can know it.

I think that all of this involves a fundamental mistake. Although there aren't facts of the matter for the correctness of one of these theories, references and dictionaries more than the other, there must be some fact of the matter, some available evidence, in order to distinguish and to identify them. Whoever can understand the building of these alternative theories, references or dictionaries must also be able to imagine observational situations by means of which he can decide if he has one of these correct theories, references and dictionaries or other. In general, likeness of correction is not sufficient for identity.

Suppose a "proxy function"[2] F(Px)=Gx defined on all the relevant empirical evidence Px for the correction or the asseverability of a determinate theory, reference or translation. If something is P, then it is G too. Now, how to know whether our theory, reference or translation is about things that are P or about things that are G? I think, this question could only make sense if we can know what is to be G independently from knowing what is to be P. If we can't, we can't understand the "proxy function" and G would only be a notational variation or another name for P –or a simple calligraphical negligence. If someone is able to understand "proxy functions", there must be states of the world, describable in a public language, related with his selection of a theory, reference or dictionary among the other built with those "proxy functions". The facts of the matter that permit us *to distinguish* equivalent theories, references and dictionaries are the facts that permit us to build and understand the "proxy functions". But they are not identificable by the adequate empirical evidence for the correction of all those theories, references and dictionaries. In other words, there must be some domain in which those "proxy functions" don't preserve the correction of some other of our theories, references or dictionaries.

Now, with regard to this, we can understand the nonsense of assigning, for example, beliefs expressed in English to a Spanish speaker when in English we can build "proxy functions" on their evidential basis that can't be one by one built in Spanish. In these cases, the mental contents assigned wouldn't be able to have just the same physical stamps and causal properties in the English people (narrow or broad) psychology and in the Spanish people psychology[3].

2. Correct attibutions of semantical content to mental states with the help of causal properties

All these ideas can serve to propose a set of ideal and restrictive conditions for a correct attribution of semantical content to mental states causally characterized:

X will correctly attribute to Y the content P, expressed in the language of X = Def

1) *(Causal condition)* Y is in a mental state of the same causal type, in a certain description, like the X's mental states that would typically have the content P.

– And –

2) *(Condition of semantical determination)* All the "proxy functions" X can make in his language that preserve the correct assertion of "to have the content P", in the most diverse and relevant situations, have one by one semantical contents related in the language of Y (in the sense that Y can perfectly distinguish all the different sets of values these functions can have and, so, there could have other possible causal and determinate descriptions for these contents).

When Y can't be competent in any language able to express contents, or when we don't sufficiently know it, the second condition would change to:

2') All the "proxy functions" X can make ... have one by one related contents with causal types of Y's mental states (in the sense that Y can perfectly distinguish, so much as X, all the different sets of values these functions can have).

If Y is able to distinguish between cat, cat'shadow, etc., and if Y is in a mental state of the same causal type as the X's mental state that in a typical causal way originate the determinate content [*to believe that there is a cat on the mat*], and X know all this, then X will have more than plausible reasons to atribute this content to Y.

If only condition 1 were satisfied, perhaps Y is in a mental state with a content similar to P, but Y can have a lot of other contents originated by the same causal type of mental state. We couldn't know if P is just the *determinate content* of this Y's mental state, only that perhaps it is similar to P up to a point. On the other hand, if only condition 2 were satisfied, the attribution could be correct. Y can have mental states with the determinate content P, but we couldn't know with sufficient guarantee if the mental state to which we attribute the content P really has or not this determinate content.

In agreement with condition 1, X has a *theory*[4] about his own mental contents. This theory classifies mental states (through some causal histories involving things like learnings) according to these states typically generate in X the determinate contents X can distinguish. For X correctly attributes to Y the content P, it is not sufficient that the Y's mental state is of the same causal type as X's mental state when X typically has this content. It is necessary that with this content Y is able *to distinguish* the same things of the world that X is able to distinguish. That Y's mental state may be just of the same causal type as the X's mental state that generate in X the content P, in a certain description, and still do not originate in Y exactly the content P. Condition 2 refines this. The more "proxy functions" X can construct, round the content P attributed to Y, that have for Y associated contents, the more correct will be the attribution of just the content P realized with the causal theory X has about his own contents. Likeness of causal types must be here likeness of contents.

The case is very different when Y isn't competent in any language able to express these contents or when we don't understand it sufficiently. Now, we will only correctly attribute to Y the contents that fulfil condition 1 if we can find in Y tokens of the causal types – of mental states– that generate for us the contents associated with all those possible "proxy

function''. This constitutes our condition 2'. I think unless we know all these relevant mental states of Y, we couldn't correctly attribute him any determinate content we have.

Curiously, the ''proxy functions'' shorten the way for a plausible attribution of contents when the subjects Y have a language. They don't convert the mental in something inscrutable, by means of an inscrutability of the reference, for example. They allow us to refine the attributions of content made through condition 1. The same causal type of mental state could be considerated as originating the same content without knowing any more about the mental states of Y. So, to learn to speak an alien language would be the best strategy to correctly attribute mental contents to other. When we try to attribute contents to ourselves or to our competent fellows in a language, the correct attributions only will depend on condition 1. When the attribution is more difficult, conditions 2 and 2' are necessary. Our conditions always call for, at least, a competent speaker of a language. Perhaps all competence is only an ideal theoretical situation. It means that the classifications X makes about the causal types of mental states that generate their contents also are ideal classifications. They will be causal classifications of mental states according to a typical generation of determinate contents, the contents expressable in a ideal situation of competence and learning. We theoretically describe and build this *ideal situation* knowing new things about the lexicon and the grammar of a language. Things that, though we are speakers of this language, we don't simply know by ''acquaintance''. A final problem: What would happen if Y has a richer language than X? Then, at least a content of X would be distributed in two different contents for Y. But, if who is doing the attributions is X, nothing of that can be known. However, the attributions of X will be correct, although X attributes that content to either one of the two mental states of Y that have the two different contents associated.

I don't believe our scheme is excessively ''parochial''. The only way we have to know causally the states of the world able to generate mental contents is begining to try do it with our own semantical contents. First, we must classify our mental states according to the causal types that we can typically find generating (perhaps in a very broad way) our contents, as we can speak about them in the languages in which we are competent speakers. Only by relation to the causal types of mental states that in ourselves, in a certain description, typically generate some determinate contents, we can correctly attribute them, with the same descriptions, to other subjects.

3. Superveniences of the mental

The supervenience relation constitute a wide family of intended relations of no-causal determination or dependence between sets of properties, facts, events, truths, etc. A very general characterization could be the following:

(S) A supervenience relation from a set A of some of those items to a set B, not necessarilly of the same class of items, is some sort of suprayective function from subsets of B to subsets of A

Usually[5], A is called the *supervenient familly* and B the *supervenience base*. The function could be from item to item, ideal case for the reduction and the definability, or could involve larger subsets of items. Sometimes this will be relative to the way we define the supervenience relation, but not always. When the last happen and those subsets are the same sets A and B, we have the special kind of supervenience that is called *global*. The necessary, or only actual, character of the function that define the supervenience relation serves to distinguish two basic senses of supervenience. One *strong*, in which the function

is thought necessary in all the possible worlds, and other *weak*, in which the function is obtained only in a world (ours, perhaps) but it isn't necessary. Also, this "necessity" or "actuality" could be interpreted in a logical (-metaphysical) or factual (-nomological) sense. By this way, there can be different sorts of *philosophical concepts* of supervenience. We can distinguish *various domains of application* for the items of A and B (the more straight case is when A and B are set of properties or predicates). And, one can replace the notion of "different subsets" by *"(di-) similarity criteria"*, and make the description of A and B relative to larger sets of items, transforming so the supervenience in something *holistic*.

In their various versions, the supervenience relations try to obstruct any proliferation of autonomous metaphysical realms. The basic material reality must determine any other reality by the way of some supervenience. This determination could make untelligible the establishment of correlattions, reductions and definitions between the items related by supervenience, although it is independent of those epistemic succeses. The supervenience has been applied to explain relations of determination between items as diverse as the mental (sensations, desires, beliefs, feelings, the conciousness, etc.) and the physical, the macrophysical common descriptions of the objects and the microphysical descriptions, the relations and the properties, the wholes and their parts, the ethical, aestetical and epistemological properties and certain descriptives properties, etc.

There are important problems for an adequate, and not as general as (S), formulation of the supervenience relations. In some cases, the supervenience could be a symmetric relation or could look very circular. We cannot clearly point out any basic and absolutelly fundamental reality for all the supervenience. The problem of no-standard models arise for any supervenience relation. The insistence in a epistemically independent determination call for very problematic and strong methaphysical assumptions. The narrow/broad divide for the mental puts the supervenience in a very dangerous indetermination about what can be its adequate supervenience base, etc.

There are, I think, two main suppositions of the usual definitions of supervenience, responsible of these problems:

1 - *Fundamentalistic claim*: Something definitelly basic determinates all the rest by means of supervenience relations

2 - *Epistemical independence*: The supervenience relations are ready made and determinates with absolute epistemical independence

Could we define supervenience relations rejecting these two claims, it is, without a definite microphysical basis and with relation of no-causal determination full of epistemical colour?

Let A and B be sets of properties, facts, events, truths, etc. Let D(A) y D(B) be the respective domains of application of A and B. We will make use of an important distinction. The distinction between a guaranteeing criterion and an authorizing criterion[6]. When to satisfy a criterion for t is F is absolutelly decisive to affirm t is F, we have a guaranteeing criterion. It is impossible that something is some way but it doesn't pass the guaranteeing criterion for being this. On the other hand, if the satisfaction of a criterion for t is F only provides a rational warrant for claiming that t is F, not necessary and sufficient conditions, the criterion is an authorizing one. It is possible that t is F and doesn't pass a authorizing criterion, and it is possible too t pass the criterion for being F but that t isn't F. We can represent any theory as proposing at least a guaranteeing criterion for something is some way. Our chemical theories about the water offer us guaranteeing criteria for something to be water. These same criteria allow us to explain why something is or not

water.

Now, we can suggest some new definitions for epistemical and no-fundamentalistic supervenience relations. The main idea is that if A supervenes on B, then there must be some deal between the criteria to apply A-items and the criteria to apply B-items. To apply certain B-items must entail to apply determinate A-items, but not the converse.

1) A 1-supervenes on B =Def There is some suprayective function f(Bi)=Aj from classes Bi of B to classes Aj of A; so that if Aj is authorizally affirmed of some x ϵ D(A), then there is at least one Bi guaranteely affirmed of some y' ϵ D(B), and if one of these Bi is guaranteely affirmed of some other y' ϵ D(B), then Aj is authorizally affirmed of some other x' ϵ D(A) too (where, perhaps x=x')
2) A 2-supervenes on B =Def A 1-Supervenes on B for every partition of A (for every partition of A, there is a partition of B, so that there is a function like f between the obtained classes)

2-supervenience offers a determination stronger and detailler than 1-supervenience. Between both concepts, we could define a wide variety of other superveniences in relation to the number and relevance of partitions of the set A involved in some function like f. There would be two special cases:

(G) A 1-supervenes on B only when Aj=A and Bi=B

(I) A 1-supervenes on B when |{Aj}|=|A| and |{Bi}|=|B|

Case (G) can be viewed as a global-supervenience, case (I) as a item-to-item-supervenience.

Also, we could make other definitions, in relation to the authorizing or guaranteeing criteria involved:

3) A 3-supervenes on B =Def ...(like 1) ... if Aj is authorizally affirmed ... at least one Bi is authorizally affirmed ... (similar to 1 with the adequated changes)
4) A 4-supervenes on B =Def A 3-supervenes on B for every partition of A
5) A 5-supervenes on B =Def ...(like 1)... if Aj is guarantelly affirmed ... at least one Bi is guaranteelly affirmed ... (similar to 1 with the adequated changes)
6) A 6-supervenes on B =Def A 5-supervenes on B for every partition of A
7) A 7-supervenes on B =Def ...(like 1) ... if Aj is guaranteelly affirmed ... at least one Bi is authorizally affirmed ... (similar to 1 with the adequate changes)
8) A 8-supervenes on B =Def A 7-supervenes on B for every partition of A

Superveniences 1, 3, 5 and 7 can be named partial-superveniences and superveniences 2, 4, 6 and 8 total-superveniences. Also, cases (G) and (I) could be defined here.

All these concepts of supervenience could be very useful, I think, to clarify and to make sistematic a lot of no-causal relations of determination. So, superveniences 1 and 2 (and the variety of supervenieces associated with them, like, for example, cases (G) and (I)) are, I propose, the mental typical superveniences. Superveniences 3 and 4 (and associated) could explain the various ways to translate natural languages and their peculiar relations of determination. Superveniences 5 and 6 (and associated) could be adequated to understand theoretical biconditional-relations in the absence of sound natural laws of causal relation. Finally, superveniences 7 and 8 (and associated) could serve to explain some conexions from the scientific to the no-scientific languages, for example, the relation of making sense in no-technical terms of a scientific theory.

Perhaps, there aren't any neat and definitive guaranteeing criterion. If so, all the superveniences will be of type 3 or 4 (and associated). But here, we could distinguish different sorts of more or least authorizing criteria. This is important and it is all we need in a naturalized perspective of supervenience[7].

4. Superveniences 1, 2 (and associated) and correct attributions of semantical content to mental states

We arrive to the main objective of this paper. If we have correctly attributed the content P (according to our previous definition), what sort of dependence or determination relation can there be between content P and the causal states required in the correct attribution? The classical answers (identity theory, some causal relation of production on mental contents, dualism, etc.) aren't completely satisfactories.

We could say that *if the content P is correctly attributed, then P 1-globally-supervenes on the causal states picked out by the condition of causal determination to correctly atribute the content P.* But, P is not a simple and unanallyzable content. P always could be re-described as a class of other identificable contents for which, perhaps, any correct attribution is obtained. 1-global-supervenience is sufficient to maintain and support a philosophical naturalism about the mental. The most important point here is that, in all these cases, 1-global-supervenience doesn't support neither a *general definability* nor a *theoretical reduction* of mental items, because it doesn't entail neither 1) a guaranty of item-to-item biconditional relations for all mental item, nor 2) the combinatorial character present in the total-superveniences and necessary in a theoretical reduction. Only total-supervenience could make the job. More extense and specified are the 1-superveniences obtained, we will be closer to total-supervenience.

Could we arrive to discover total-superveniences for the mental? Here, the decisive obstacles, I think, aren't logical but empirical. Our mental contents aren't as stable as it would be necessary to obtain total-superveniences. The continuous changes and variations in the semantical relations that we employ to describe and to identify mental contents are too much *plastic* (changes is the stereotipes, in the learning-histories to be able to use a determinate word, in the "proxy-functions" we can understand, etc.). Only in very restricted and "modular" cases, total-superveniences would be available. This is an important problem for science, but not for philosophy. It is not a good policy to found philosophical naturalism in the myth of any total science.

Notes

(1) The mental content expressed by us with the help of the meaning of the english words placed into square braquets

(2) In the Quinean sense. Quine (1960).

(3) This same idea is very well argumented in a brief paper of Gomez (1986):

(4) Really, I don't believe that we arrive to have such a theory for determinate contents like "X [thinks of Vienna]", although perhaps so for some other more modest contents. Conditions 1 and 2 are ideal and restrictive conditions. However, through these idealisations we hope to make clear what is to make "causal" theories about the mental and, by this way, to know more about the mental.

(5) I am thinking of the Kim's works. See, for example, Kim (1978, 79, 82, 84, 87 and 88)

(6) This distinction is collected from Rescher (1973)

(7) What has happened with the "necessity" of some others definitions of supervenience, like those of Kim (see note 4)? Here, the necessity of this "weak supervenience" only could mean that a function similar to F is obtained. And the necessity of his "strong supervenience" (relative to possible world) only could have the force that can have the guaranteeing criteria, if any, involved. This is to naturalize the supervenience relation.

References

GOMEZ C. (1986) *"Inescrutablidad de la referencia, lenguajes privados y estructura de la mente"*, Actas del I Simposio Hispano-Mexicano de Filosofia. Salamanca, 1984, vol I, Salamanca, Ed. Univ. de Salamanca

KIM J.(1978) *"Supervenience and Nomological Incommensurables"*, American Philosophical Quarterly, 15

(1979) *"Causality, Identity, and Supervenience in the Mind-Body Problem"*, Midwest Studies in Philosophy, 4

(1982) *"Psychophysical Supervenience as a Mind-Boby Theory"*, Cognition and Brain Theory, vol. V

(1984) *"Concepts of Supervenience"*, Philosophy and Phenomenological Research, vol. XLV, No. 2

(1987) *"'Strong' and 'Global' Supervenience Revisited"*, Philosophy and Phenomenological Research, vol. XLVIII, No. 2

(1988) *"Supervenience for Multiple Domains"*, Philosophical Topics, vol. XVI, No.1

QUINE W.V. (1960) *Word and Object*, Cambridge, MIT PRESS

RESCHER N. (1973) *The Coherence Theory of Truth*, Oxford, Clarendon Press

Ng Tai-Kee

South China Normal - Univ.
Guangzhou, (P. R. China)

Static and Dynamic Structure of Consciousness in Psychology

Abstract

From the fundamental concepts of existence, entity, subject, and object, the author, using mathematics and wave mechanics, shows two figures of static and dynamic structure of consciousness, and further deduces from them, and explain some psychological concepts and phenomena.

Resumo

O autor mostra, a partir dos conceitos fundamentais de existência, entidade, sujeito e objecto e com recurso à matemática e à mecânica ondulatória, duas figuras da estrutura estática e dinâmica da consciência, deduzindo depois dessas figuras alguns conceitos e fenómenos psicológicos, que explana.

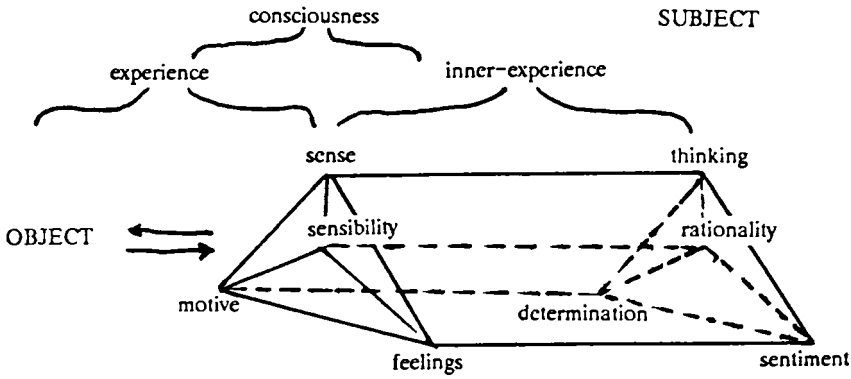
1. Static Structure of Consciousness

In my paper, *Existence and Types of Consciousness*, published in the abstracts of VIII International Congress of Logic, Methodology and Philosophy of Science, volume 5, part 2, sections 7-11, pages 319-321, in Moscow, 1987, by Soviet Organizing Committee, I had shown a static structure of consciousness as a prism. I defined three dimensions of intellect, emotion and volition and separately gave the definitions of sense, thinking, feelings, sentiment, motive and determination. i.e.:

sense - intellectual sensibility;
thinking - intellectual rationality;
feelings - emotional sensibility;
sentiment - emotional rationality;
motive - volitional sensibility;
determination - volitional rationality.

Sensibility is supposed to be the abilities of a man sense-organs and rationality is the ability of the brain of a man - subject. It may be expressed as follow:

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 214/219
1992, APDC & SPF, Lisboa



2. Memory and Intuition

From the above theory, we can define memory and intuition in psychology.

When object relates to subject, and so relates to sensibility and rationality of a man, the relation between object and sensibility constructs an experience, the relation between sensibility and rationality constructs an inner-experience, and the experience and inner-experience construct the man's consciousness. With the construction of experience and inner-experience i.e. the man's consciousness, there must be traces left in his nervous organization simultaneously, no matter how short-term or long-term, weak or strong, discontinuous or continuous, or generalizing gradually and vanishing at last they are. It is the traces-saving that let our rationality code, retrieve and compare for recognition, or associate and reconstruct chunking, clustering or reverberating circuit for recall. Recognition and recall are what we name memory.

As consciousness has three dimensions of intellect, emotion and volition, and memory is crossed consciousness, so memory must have three dimension of intellect, emotion and volition as well, i.e.:

intellectual memory;
emotional memory;
volitional memory.

As to intuition, I agree with Professor H.A.Simon's view. He defines intuition as the memory at a moment; but possibly, more exactly as the occasional memory at a moment, because intuition is different from sensory memory although this one is the memory at a moment too. The very weak, light or quickly vanishing traces are easily forgetten, even can't be coded; but on some appropriate conditions they may be reinforced at a moment and so occasionally can be coded, retrieved or associated, reconstructed for occasional recognition or recall at a moment, namely for an occasional memory at a moment. This is what we call intuition.

As consciousness has three dimensions, memory is crossed consciousness and of course, intuition is crossed consciousness also, so intuition must have three dimensions, i.e.:

intellectual intuition - insight;
emotional intuition - inspiration;
volitional intuition - impulsion.

3. Dynamic Structure of Consciousness

Using wave mechanics, we can get a dynamic structure of consciousness.

Let Ψ v, Ψ e, Ψ i be the wave functions of three dimensions of volition, emotion and intellect separately, then the sum of these three wave functions is just the whole man's consciousness. Supposing Φ vei is the whole consciousness, then

$$\sum_{k=1}^{3} \Psi k = \Phi \text{vei};$$

i.e. Φ vei = Ψ v Λ Ψ e Λ Ψ i.

Obviously, Φ vei is a wave function too. But we have to know:

First, the set of defining wave function Φ vei values is a finite set, because the distance between sensibility and rationality can't be 0 and ∞ too. If it is zero, sensibility and rationality are in one and the man is not different from an animal, and if it is infinity, rationality is apart from sensibility and the man has become quite a God, which are both impossible.

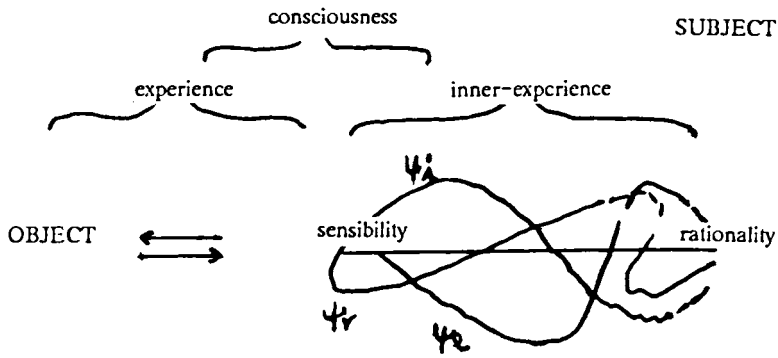
Secondly, in the wave function Φ vei there may be or not a phase difference of its component wave functions of Ψ v, Ψ e and Ψ i. It is said, when the relation between object and sensibility constructs experience and the relation between sensibility and rationality constructs inner-experience, stimulation makes oscillators of sensibility vibrate and travel to rationality along three dimensions possibly in same step or not in same step. In same step are the three component wave functions of Ψ v, Ψ e and Ψ i started on same time and not in same step are not started on same time and is having a phase difference.

Thirdly, not only there may be a phase difference of Ψ v, Ψ e and Ψ i in the resultant wave function Φ vei but the amplitude, frequency, period and phase-velocity of these three component functions may be, or even we may say, must be different also because heredity, physiological factor especially, trace and traces-saving will give them a different influence separately.

Fourthly, trace has one, two or three dimensions of volition, emotion and intellect also. So, it can interrup all coming waves with its correspondent dimension separately to increase or decrease, reflect or refract them not only of amplitude, frequency, period, but of phase-velocity. In our nervous system there is a lot of trace and traces-saving, and of course, the wave function Ψ v of volition, the wave function Ψ e of emotion and the wave function Ψ i of intellect are more complex. They may be harmonic or non-harmonic, periodic or aperiodic, forced or damped, divergent or convergent.

Fifthly, wave has peak and valley. Putting wave in an Euclidean space of three dimensions x, y and z, sensibility in the original point 0 and the distance between sensibility and rationality in x axis extensive to rationality, then amplitude of wave has two values of opposite sign, positive to peak and negative to valley. However, in the discrimination and appreciation of rationality there is a different meaning about these two values. In component wave function Ψ v of volition, positive appears goodness and negative appears evil; in Ψ e of emotion, positive, beauty and negative, ugliness; in Ψ i of intellect, positive, truth and negative, falseness. Man is always towards truth, beauty and goodness.

These may be expressed as follow:



4. Attention and Imagination

From the above dynamic structure of consciousness we can define attention and imagination in psychology.

In wave mechanics, as we have seen, when waves overlap, two cases will appear, one of that is superposition, the other is interference. Superposition has a resultant wave and its amplitude is a vector sum of all component amplitude vectors. In consciousness, when object relates to subject, stimulation makes oscillators of sensibility vibrate and travel along three dimensions to rationality and travelling waves have met traces in the way. If the waves of one, two or three component wave functions Ψ v, Ψ e and Ψ i have overlapped in some positions, there must appear superpositions or resultant waves of superpositions. And if one of resultant waves has a largest positive amplitude vector of all other resultant vectors, this resultant wave will be an irrational or unconscious attention, in which the peak of wave is the center of unconscious attention, the domain of wave is the region of unconscious attention and the boundary of domain is the edge of unconscious attention. When the resultant wave with a largest positive amplitude vector is transferred to other position, the center, region and edge of unconscious attention are transferred to other object.

Obviously, unconscious or irrational attention is based on the process of sensibility to rationality or, as the psychologists name, "bottom-up" process. However, on the other hand, I agree with a view of modern philosophers of science, there is another process of rationality to sensibility, namely, "top-down" process as psychologists call it.

As we know, the set of defining wave function Φ vei values is a finite set, so it is a closed set. As a closed set there is stored energy in its system and under boundary conditions energy is released and makes oscillators of rationality vibrate and travel along three dimensions to sensibility. They also meet many traces in the way. This is the process of rationality to sensibility. Similarly, when the waves of Ψ v, Ψ e, Ψ i overlap, there appears a resultant wave with a largest positive amplitude vector. This is a rational or conscious attention, in which the peak, the domain and the boundary of domain of wave separately are the center, the region and the edge of conscious attention. And if the resultant wave with a largest positive amplitude vector consciously or rationally transfers to other position, the center, region and the edge of conscious attention transfers to others also.

Conscious or rational attention and unconscious or irrational attention are both attention.

As the resultant wave with a largest positive amplitude vector may appear in one, two

and three dimensions, so we can roughly divide attention into:

intellectual attention;
emotional attention;
volitional attention.

As to imagination, as it is lesser related to process of sensibility to rationality, it must mainly come from process of rationality to sensibility, or more correctly, mainly come from conscious attention and be different from conscious attention. The difference between them is whether the weak, light or quickly vanishing traces have been reinforced.

When boundary conditions of the resultant wave function Φ vei have been satisfied, stored energy is released and not only makes oscillators of rationality vibrate and travel to sensibility along three dimensions but also reinforces the very weak, light and quickly vanishing traces and enlarges traces–saving so much as to be able to be more coded, more retrieved for reappearance, or more associated, more reconstructed into new chunking, new clustering or new reverberating circuit for creativeness including illusion although the travelling waves of three dimensions will meet more traces in the way too. This is imagination, and reappearance and creativeness are all made of imagination. If waves of Ψ v, Ψ e, Ψ i overlap, a resultant wave with a largest positive amplitude vector appears, the peak, domain and boundary of domain of that are separately the center, region and edge of imagination as well. Imagination is different from intuition although both of them reinforce the very weak, light or quickly vanishing traces all but imagination is based on the process of rationality to sensibility and intuition is based on the process of sensibility to rationality. Moreover, imagination is related with dynamic analysis and intuition is at a moment.

Similarly, we can roughly divide imagination into:

intellectual imagination;
emotional imagination;
volitional imagination.

About dream, a word I want to say is that the quickly vanishing traces and the vanished ones are not really dead for all. A great deal of them is only sub–trace, not sub–consciousness which I don't like, because Freud put it in a ground of consciousness opposite to our view of consciousness as a whole. It is sub–trace when certain surroundings and boundary conditions are satisfied that may be awaked and be coded, retrieved, associated, and reconstructed as well. I think that this is fundamental on dream. Clearly, dream is based on the process of rationality to sensibility.

5. Ability, Temperament and Character

Person has three parts: ability, temperament and character and we can put them in the dynamic structure of consciousness also.

As we have seen, trace has three dimensions and they can separately influence the waves of three component wave functions of Ψ v, Ψ e and Ψ i. But through repetition for a long time the traces would be deep and fixed, or even connected with each other as a link to coming waves often giving influence in one mode. It is just what we call habit in psychology.

From this we can theoretically say:

when the waves of two component functions of Ψ i and Ψ v meet traces with two fixed dimensions or fixed links of intellect and volition they often give waves the same influences as one mode so that consciousness has a habit in two dimensions of intellect and

volition. It is called person's ability.

Similarly, when the waves of two component wave functions of Ψ v and Ψ e meet traces with two fixed dimensions or fixes links of volition and emotion they often give waves the same influences as one mode so that consciousness has a habit in two dimensions of volition and emotion. It is called person's temperament.

And when the waves of two component wave functions of Ψ e and Ψ i meet traces with two fixed dimensions or fixed links of emotion and intellect they often give waves the same influences as one mode so that consciousness has a habit in two dimensions of emotion and intellect. It is called person's character.

And a man's person just consists of ability, temperament and character.

4

INFORMAÇÃO, COMUNICAÇÃO, REPRESENTAÇÃO & SIGNIFICAÇÃO

INFORMATION, COMMUNICATION, REPRESENTATION & MEANING

P.R. Medina Martins

Inst. Superior Técnico, Univ. Técnica de Lisboa
Lisboa (Portugal)

Comunicação, Conversação e Cibernética

Resumo

Uma breve análise das definições de Cibernética mostra, por um lado, que esta se encontra estreitamente associada com aquilo que podemos designar por conceitos "técnicos" (usualmente ligados a qualquer tipo de medida quantitativa) envolvendo controlo, comunicação, fluxo de informação, etc.; inclui também, no entanto, conceitos "não-técnicos" tais como conhecimento, acção e compreensão relativamente aos quais um suporte quantitativo é difícil de estabelecer. O problema aqui em causa é semelhante ao que, secularmente, tem separado ciências humanas das ciências físicas / quasi-físicas.

Tenta-se esboçar nesta comunicação (através de uma brevíssima análise diacrónica) como certas modificações recentes quer da Cibernética tradicional quer da Algebra clássica que lhe servia de suporte abriram um conjunto de possibilidades susceptíveis de atenuar e mesmo eliminar aquela separação.

Abstract

A short account of Cybernetic's definitions shows their connection on the one hand with "technical" concepts (usually associated to quantitative measures) concerning control, communication, information flux, etc. and on the other hand, with "non-technical" concepts, such as knowledge, action, understanding, which are hardly related with quantitative supports. The problem we face here is similar to the separation that has occurred between human sciences and physics/quasi-physics ones, for a long time.

Our aim is to establish, in a short diacronic analysis, how some recent changes that took place in traditional Cybernetics and classic Algebra opened a set of issues able to atenuate or even to abolish that separation.

1 - Introdução

Numerosas definições foram propostas para "Cibernética" desde 1948, altura em que quer o termo quer o conjunto de ideias a ela subjacentes foram re-introduzidas por Norbert Wiener no panorama científico contemporâneo. Quatro destas definições são especialmente importantes para os objectivos desta comunicação:

- a mais antiga (funcionando ainda como referência) do próprio Wiener o qual apresenta a Cibernética como a "Ciência do Controlo e da Comunicação no Animal e na Máquina";
- a de Mc Culloch encarando a Cibernética como "uma epistomologia experimental associada à geração do conhecimento";

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 223/242
1992, APDC & SPF, Lisboa

- a de Louis Couffignal definindo Cibernética como "a arte de assegurar a eficácia da acção";
- finalmente uma definição muito recente de Maturana que caracteriza a Cibernética "como a ciência e a arte da compreensão".

Uma breve análise destas definições mostra, por um lado, que a Cibernética está estreitamente associada com aquilo a que se podem designar conceitos "técnicos" (usualmente ligados a qualquer tipo de medida quantitativa) envolvendo controlo, comunicação, fluxo de informação etc.; inclui também, no entanto, conceitos "não técnicos" tais como conhecimento, acção e compreensão relativamente aos quais um suporte quantitativo é difícil de estabelecer.

O problema que aqui está em causa é semelhante ao que, secularmente, tem separado ciências humanas das ciências físicas/quasi-físicas.

Tentarei esboçar nesta comunicação (através de uma brevíssima análise diacrónica) como certas modificações recentes quer da Cibernética tradicional quer da Algebra clássica que lhe servia de suporte, abriram um conjunto de possibilidades susceptíveis de atenuar e mesmo eliminar aquela separação.

O artigo compreenderá assim mais três secções fundamentais respeitantes

- aos conceitos básicos da Cibernética clássica (2),
- às modificações que conduziram ao estabelecimento da chamada "Cibernética de 2ª Ordem" ou Cibernética não- clássica (3),
- finalmente a uma secção denominada "As novas fronteiras e os novos problemas"(4) onde aquelas modificações da Álgebra booleana clássica serão sumariamente abordadas.

2 - A Cibernética de 1ª Ordem e o Paradigma Clássico

É sabido que a "Cibernética" de Wiener nasceu da resolução de um problema muito concreto de Engenharia; um problema envolvendo o mecanismo de controlo dos canhões de defesa anti-aérea utilizados na 2ª Guerra Mundial. O grande mérito e impacto que o seu livro teve não resultaram, contudo, da solução em si mesma (que no entanto esboçou aquilo que é hoje a Teoria dos Sistemas de Retroacção e Controlo) mas antes por *através dela* ter estabelecido que a secular separação entre os sistemas "sem história" da Física e os Sistemas Biológicos (onde processos temporais e evolutivos têm necessariamente de ser considerados) não tinha significado. Em particular que os "propósitos", "finalidades", "procura de objectivos" e a "teleologia" característicos destes últimos sistemas eram, todas elas, questões redutíveis a um problema de estrutura interna.

Ao unir desta maneira a Física à Biologia (e inversamente a Biologia à Física) Wiener deu origem a uma verdadeira revolução epistemológica concretizada na emergência de uma *corrente interdisciplinar* por meio da qual um conjunto de técnicas, métodos e abordagens teóricas *gerais* passaram a poder ser aplicados a uma enorme classe de sistemas até aí analisados sob pontos de vista particulares e distintos.

Uma das áreas onde essa corrente teve mais impacto foi a respeitante à simulação efectiva de funções ou processos (fisiológicos, psicológicos etc) característicos dos seres vivos, até aí num estado de imenso atraso relativamente ao já conseguido na Física. Para dar uma ideia desse impacto um panorama do estado da Cibernética quinze anos apenas após o aparecimento da obra de Wiener indica-se a seguir |1|. A listagem (longe de ser exaustiva) vale, essencialmente, pelo número de áreas já exploradas num período tão curto. Salientam-se assim:

a) os sistemas visuais (incluindo reconhecimento, identificação e percepção de formas) de Culbertson, Pitts e Mc Culloch, Rapopport, Osgood e Heyer, Uttley, Deutsch e George, Rosenblueth;
b) os sistemas classificativos (que podem ser aplicados a um ou a vários sentidos e implicam aprendizagem e crescimento) de Uttley, Chapman, George, Tolman, Gordon Pask;
c) os sistemas envolvendo associações (com bases em probabilidades condicionais) de Shannon, Grey Walter, Deutsch, Uttley e Stewart;
d) os sistemas de Broadbent na área do reconhecimento auditivo;
e) os modelos de memória humana (incluindo tempo de atraso e sistemas de armazenagem magnética) na forma descrita por Oldfield;
f) os sistemas que actuam de acordo com "motivações internas" de Uttley e Ashby;
g) as áreas relacionadas com a programação de linguagem natural, construção de regras de referência e processos heuristicos de Turing, Church, Bridgman, Wang, Newell, Shaw e Simon.

Por essa altura a grande maioria desses modelos utilizava componentes mecânicas, electromecânicas e/ou electrónicas na sua construção (motores eléctricos, bobinas, condensadores, relés, elementos sensíveis à luz, temperatura, etc.).

Com a crescente utilização de computadores digitais operada nos últimos vinte anos este "hardware" foi profundamente modificado. Não o foi todavia – e este é o ponto essencial – o *paradigma conceptual* no qual quase todos esses modelos foram (e em grande número ainda são) baseados.

Este paradigma que doravante designarei por paradigma *clássico* |2| emergiu na cultura ocidental nos finais do séc. XVII, inícios do séc. XVIII e durou praticamente sem oposição até finais do séc. XIX. A sua ideia base, ainda hoje não totalmente modificada, é a de que a Natureza (como conceito geral) podia ser encarada como algo que na terminologia contemporânea descreveríamos por "máquina sequencial determinística". O funcionamente desta máquina era suposto obedecer aos postulados da mecânica newtoniana e a descrição das suas acções poderia ser feita por meio de equações diferenciais ou integro-diferenciais.

Sucessivas confirmações experimentais quer daqueles postulados quer deste método de descrição matemático foram alargando sucessivamente o núcleo inicial (aquilo a que hoje denominamos Mecânica de Corpos Rígidos) de tal forma que pelos finais do séc. XIX praticamente todas as disciplinas (Hidrodinâmica, Óptica, Mecânica de Corpos Deformáveis, até à Teoria Electromagnética de Maxwell) utilizavam quer conceitos mecânicos ("força", "massas", "energias" etc.) quer aquele tipo de procedimento matemático.

O sucesso desta descrição foi tal que nos finais do século passado não só se atingiu em Ciência aquilo que A. Rey chamou a "consciência do definitivo" como se supôs também que o paradigma clássico poderia ser generalizável à descrição biológica e psicológica de seres vivos quer isolados quer em grupo.

Não faz sentido nesta comunicação dar ênfase às razões que conduziram ao confinamento do paradigma clássico a certas disciplinas apenas. Importante é frisar que apesar destas restrições (como é sabido o paradigma está limitado à observação e descrição de fenómenos macroscópicos) as suas premissas básicas continuam a desempenhar papel activo em muitas das disciplinas, em particular as ligadas por exemplo à Psicologia.

Algumas destas premissas são particularmente relevantes para os objectivos desta comunicação. Refiro-me em especial à ideia de que:

i) Qualquer tipo de medida, de regulação ou de controle contém, implicitamente,

uma referência a um padrão/sistema de medida denominado *sistema de referência*. Este sistema é em última análise estabelecido por um *observador humano*.

ii) Este observador é (suposto ser) *imparcial, externo ao objecto/sistema sob observação sendo, muitas vezes, a ele também conferida omnisciência*.

iii) Os (possiveis) acontecimentos aos quais o observador assiste (particularmente os que para ele são considerados relevantes para o estudo do objecto/sistema sob observação) são, na generalidade, conhecidos previamente. Ao observador é também atribuida a capacidade de estabelecer hipóteses (determinísticas e/ou probabilisticas) sobre essas ocorrências.

iv) O objecto/sistema sob observação é suposto comportar-se como um *ente impessoal* (um "it" na linguagem inglesa) sendo frequentemente dividido (pelo observador) em partes/componentes tais como "aquilo que transmite" e "aquilo que recebe", aquilo que "causa" estes ou aqueles "efeitos", etc.

v) Uma destas partes (usualmente denominada ambiente) é encarada pelo observador externo/agente das experiências como actuando sobre outra parte; neste caso esta última é vulgarmente denominada *objecto* da observação.

vi) No caso desta distinção não poder ser claramente definida o próprio observador/experimentador pode desempenhar o papel de ambiente. Nestas condições, todavia, as suas acções têm de obedecer a um rígido e *pré-determinado* conjunto de regras.

vii) Para além destas interacções (usualmente representadas em termos de valores de certas "variáveis independentes") todas as outras variações não encaradas como relevantes são *minimizadas*. Através desta restrição procura-se que as condições ambientais sejam tanto quanto possível mantidas constantes (na medida em que podem influenciar os valores das variáveis observadas). Estas últimas variáveis são, por seu lado, encaradas como representando as "variáveis dependentes" do processo.

viii) No caso da experiência ter de ser repetida é exigido que as condições experimentais sejam réplicas exactas daquelas anteriormente estabelecidas. Para além disso as experiências devem ser efectuadas de tal maneira que apenas estejam em causa quer perturbações de uma só variável independente, quer a (consequente) observação de poucas (de preferência uma) das variáveis dependentes.

A representação indicada na Fig.1 condensa de algum modo estes pressupostos.

O observador/experimentador estabeleceu uma distinção entre um "ambiente" e o objecto o qual, para maior generalidade, representa neste caso um "organismo"[1]. O ambiente vai actuar causalmente sobre o organismo de acordo com um padrão de actuação pré-definido.

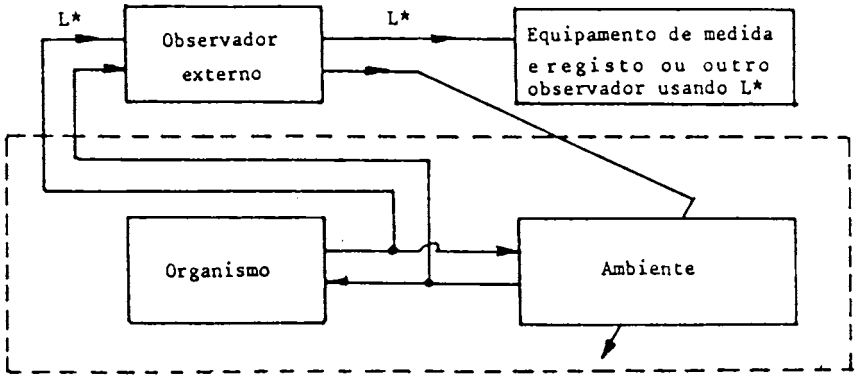


FIGURA 1

Na linguagem do paradigma clássico isso irá implicar um envio de "estímulos" (obedecendo às regras do padrão) e os acontecimentos (ou reacções) observadas são descritos pelo observador (usando uma metalinguagem L)[2] em termos de um "comportamento" englobando "entradas" (ou "reforços") e "saídas" (ou "respostas").

Caso seja possível ao observador é ainda atribuída a capacidade de modificar de tempos a tempos as características do ambiente. Esta situação é representada pela seta que atravessa o rectangulo representativo do ambiente; a seta irá assim desempenhar o papel de um *parâmetro* simbólico de todo o processo.

Uma vez isto admitido os objectivos experimentais do observador serão satisfeitos logo que um número (por ele considerado suficiente) de estimulos-respostas/entradas-saídas seja atingido. Nestas condições o organismo poderá ser descrito quer como um *autómato finito* (em termos computacionais) quer por meio de *funções/matrizes de transferência* (sempre que esses estímulos/entradas ou saídas/respostas variem continuamente no tempo).

Dois aspectos desta representação são cruciais em termos da sua aplicação à Psicologia: o tipo de descrição escolhido pelo observador e o âmbito restrito das suas operações paramétricas. Ambos vão dar origem a que o próprio ambiente possa ser também considerado como um automato finito, em consequência do que (tomando em consideração o que foi dito acerca do organismo como outro automato) duas importantes conclusões se podem extrair.

Em primeiro lugar que a interacção organismo/ambiente pode doravante ser interpretada como uma *troca de símbolos numa linguagem formal*. Segue-se em consequência que

- organismo e ambiente podem ambos ser encarados como *máquinas abstractas* que aceitam/geram símbolos/sequências de símbolos;
- o alfabeto desta linguagem consiste em rótulos descritivos de acontecimentos denominados na metalinguagem L "estímulos" e "respostas";
- aqueles símbolos rotulam também as transições entre estados;
- os estados internos dos autómatos representativos do organismo e do ambiente são símbolos não-terminais da linguagem formal;
- as estruturas destes automatos podem ser encaradas quer como gramáticas formais quer como uma colecção ordenada de regras escritas.

Em segundo lugar (observando agora o sistema externamente) é também possível concluir que o observador pode (através de registos da ocorrência dos estímulos/sequências de símbolos) estabelecer tabelas das combinações de estímulos/respostas, das frequências de acordo com as quais essas combinações ocorrem, etc.; a partir daí estimar também a sua

probabilidade, calcular índices de incerteza relativamente a essas ocorrências, etc. Em resumo, o observador é (suposto ser) capaz de estabelecer uma *teoria* daquilo que está a acontecer – isto pelo menos se o "organismo" diz respeito a um rato, um pombo, a um qualquer subsistema reflexo de um ser humano. Todavia se este "organismo" se referir quer a um ser humano como *entidade global* quer a algumas das suas actividades (particularmente àquelas designadas na generalidade por "pensamento") a aplicação deste tipo de raciocínio pode dar origem ao aparecimento de numerosas dificuldades.

Três dessas dificuldades são aqui de salientar.

A primeira diz respeito ao que foi afirmado acerca de estímulos/respostas. É evidente, com efeito, que de acordo com aquele ponto de vista globalizante o significado de termos tais como "entrada", "saída" "estímulo", "resposta", "parâmetro", etc. se torna extremamente forçado. Por um lado porque há situações que estão claramente fora do âmbito experimental (por exemplo desnutrição total, em termos de Psicologia Experimental pelo menos). Por outro lado porque, com raras excepções (tais como no estudo de crianças muito pequenas ou na privação sensorial do adulto) aquilo que realmente acontece é o experimentador fornecer instruções a um experimentando baseado na crença de que essas instruções irão provocar-lhe um "estado mental" apropriado. O problema é que essas instruções estabelecem uma constância (se na realidade o fizerem) não porque sejam impostas como um quasi-constrangimento no sentido da Física mas – e isto é ponto crucial – porque elas são *interpretadas*, *percebidas* e *aceites*. Devido a este facto dar e receber instruções são situações que estão fora do paradigma clássico. Pertencem mais a *paradigmas normativos* do que a situações funcionais limitadas por condições que o experimentador estabelece independentemente do organismo em si mesmo.

A segunda dificuldade põe em evidência a (falsa) oposição originada no observador/experimentador entre aquilo que é considerado "observação objectiva" e "observação subjectiva" como se estas fossem falíveis (embora extremamente interessantes) e aquelas concretas e merecedoras de confiança (embora muitas vezes não informativas).

Sendo "falíveis" as observações "subjectivas" serão consequentemente excluídas do âmbito dos interesses do observador clássico. O que na realidade acontece é esta exclusão derivar apenas e só apenas de um dos pressupostos do paradigma clássico o qual, neste caso, quase adquire força de lei; que o "organismo" sob observação é um "it", um ente impessoal. Nestas condições tais acontecimentos "subjectivos" só poderão ser analisáveis e descritos por meio de sistemas de referência *não-padronizados* tais como introspecção, rectrospecção, etc.

Finalmente a terceira dificuldade diz respeito quer à descrição pré-definida implícita no paradigma clássico quer à sua abordagem em termos da Teoria de Automatos. Essa descrição irá definir *antecipadamente* os estados de entrada e saída destes automatos; e em conjunto com o que foi afirmado relativamente às restricções paramétricas ambas as condições irão dar origem a que os estados internos do ambiente *não possam ser modificados* em consequência de qualquer acção que o organismo execute fora dos limites da descrição inicial (*novas* acções por exemplo). Nestas condições contudo é possível aplicar ao esquema interactivo da Fig.1 o argumento de Von Foester demonstrando que qualquer sequência comportamental nesse tipo de sistema é *pré-determinada*. Sendo assim proposições do tipo "o sistema organismo/ambiente evidencia adaptações" transformam-se em meras tautologias: o sistema irá comportar-se assim a menos que algumas das condições pré-definidas tenham por qualquer razão sido alteradas sem que o observador o quizesse (caso em que o próprio paradigma de base deixa de ser aplicável). Evolução e/ou aprendizagem por exemplo (num sentido implicando a aquisição/construção de

novidades) são assim situações fora do contexto do paradigma clássico.

3 - O Paradigma Relativista e o Nascimento da Cibernética de 2ª Ordem

Com o objectivo de eliminar estas dificuldades Gordon Pask, Heinz Von Foester e Ranulph Glanville (entre outros) propuseram em 1974/1975 uma modificação ao esquema representado na Fig 1. Essa mudança está graficamente indicada na Fig.2

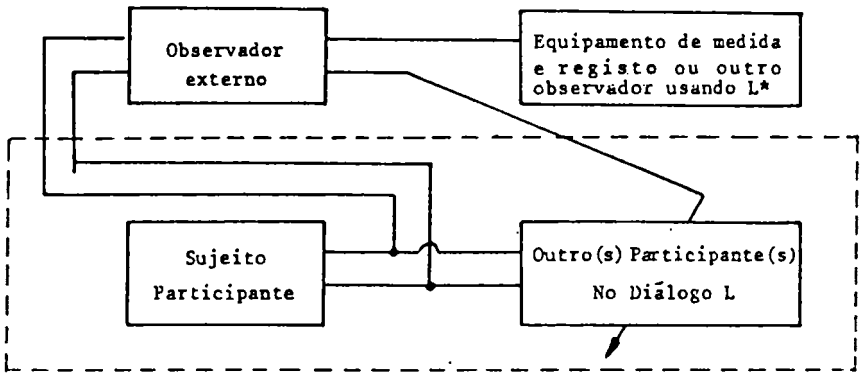


FIGURA 2

À primeira vista esta representação parece ser idêntica à anterior. Também aqui um observador externo utiliza uma metalinguagem L* para descrever o sistema, enunciar e testar hipóteses etc. Contudo, e isto é o aspecto a salientar, as situações que os dois rectangulos inferiores simbolizam são agora completamente diferentes. Na realidade os rectangulos referem-se não só a participantes com também a entidades normativas interactuando com base num contrato experimental previamente negociado com um deles (pelo menos), um ser humano em si mesmo, "qua sentient" e não um simples "it" como na anterior situação funcional.

Os resultados que se puderam (podem) obter desta modificação transcenderam (transcendem) a de uma simples mudança gráfica. Na verdade as consequências que dela se extraíram deram origem a novos métodos de abordagem psicológica, epistemológica, filosófica etc., a novas perspectivas na forma de encarar a comunicação humana (individual ou grupal), a novos processos de simulação de comportamentos humanos por meio de computadores etc. todos eles englobados numa nova visão da cibernética tradicional ligada não à observação, medida, controlo etc. de um processo qualquer mas à observação de um processo de observação, à medida de um processo de medida, ao controlo de um processo de controlo etc.. É a esta nova perspectiva que corresponde a chamada *cibernética não-clássica* ou *cibernética de 2ª ordem*, expressão introduzida por Heinz Von Foester em 1974 |3|.

Algumas destas modificações serão esboçadas nas considerações seguintes.

Começarei por me referir à forma como os problemas indicados em 2/ foram solucionados por meio do esquema da Fig.2.

i) O primeiro desses problemas referia-se ao significado de dar/receber instruções e às suas dificuldades no caso de utilização do paradigma clássico. É evidente agora ser o próprio sujeito sob observação que concorda não só em actuar de

determinada maneira como também em falar uma certa linguagem. Dito de outra maneira: enquanto o contrato se mantiver é o sujeito (não o observador externo/ experimentador) que consequentemente fica responsável pela manutenção das condições em que a experiência está a ser conduzida. Nesse aspecto o papel desempenhado pelo experimentador será mínimo.

ii) Por outro lado, também aquilo que foi afirmado acerca do significado forçado de "estímulos", "respostas", "parâmetros" etc. quer no contexto da Psicologia ("estímulo"=estímulo efectivo) quer no da linguagem semanticamente neutra da Teoria de Sistemas ("estímulo"=entrada causadora de ...) será igualmente sujeito a profundas modificações. É óbvio que tudo o que um experimentador faz *pode* ser interpretado como "estímulo". Da mesma forma qualquer acção correlacionada *pode* também ser encarada como "resposta". Por exemplo ler um poema, ver um filme, dar uma ordem, fazer uma pergunta etc. podem ser estímulos. Todavia enquanto que para o observador tradicional limitado pelas regras do paradigma clássico estas acções são apenas e só apenas "estímulos" para o observador do esquema normativo quaisquer outras interpretações (outras realidades, no caso mais geral) *podem* igualmente ser consideradas. Por exemplo o "poema" pode ser encarado como um poema que o sujeito percebeu (e que eventualmente lhe poderá fazer rememorar um determinado estado emocional), a "ordem" como algo a ser obedecido ou rejeitado, a "pergunta" como algo a que o sujeito sabe/quer ou não sabe/não quer responder.

iii) Nestas condições é evidente também que quer o experimentador quer o(s) participante(s) ligados pelo contrato previamente referido têm agora ao seu dispor um conjunto de regras de interaccão praticamente ilimitado. Assim se o observador partir da hipótese que o sujeito ("organismo, "it") irá responder a um estímulo então poderá ser levado a afirmar tal como no esquema clássico, por exemplo que "o sujeito respondeu com este acto reflexo a uma pressão naquele centro nervoso". Por outro lado, contudo, se adoptar um esquema com base numa linguagem normativa então o experimentador poderá igualmente pôr questões ao sujeito não como "it" mas como *ser humano em si mesma*. Mais ainda: como qualquer conjunto de regras de interacção é teoricamente aceitável (apenas a sua infracção é que tem de ser evitada) segue-se que a separação tradicional entre "observação objectiva" e "observação subjectiva" e os problemas a elas ligados deixam também de ter significado.

Nesse sentido (antecipando um resultado que será analisado com mais pormenor adiante) é de salientar que se o rectangulo do lado direito representar outro ser humano desempenhando quer o papel de observador quer o de uma espécie de "espelho" para os processos afectivos/cognitivos do sujeito do rectangulo esquerdo então o esquema do paradigma relativista pode ser encarado como simbolizando parte (pelo menos) de uma realização psicoterapeutica. O mesmo se poderá afirmar para o caso de os dois rectangulos simbolizarem partes de uma mesma mente, uma actuando primária (embora não exclusivamente) em termos cognitivos outra (mais "primitiva") actuando primária (embora não exclusivamente) em termos afectivos.

iv) Com esta referência ao rectangulo do lado direito é possível examinar outras possibilidades nas quais o esquema normativo evidencia maior generalidade que o esquema clássico. Duas situações extremas e uma situação intermédia serão aqui focadas.

A primeira dessas situações extremas corresponde ao caso em que esse rectangulo simboliza um "ambiente" estruturalmente representado como um autómato finito. Nestas condições o contrato para actuar de determinada maneira é estabelecido apenas e só apenas com o participante (humano) do rectangulo esquerdo. O diálogo preliminar utilizado para negociar esse contrato (no limite "dar instruções") pode ser quer a metalinguagem L* quer outro tipo de linguagem com as características da linguagem natural; mas é evidente também que a manutenção do contrato depende inteiramente da cooperação do participante. E como por outro lado a interacção sujeito/automato é essencialmente do tipo "estímulo"-"resposta" segue-se que o próprio participante (durante a interacção) está a actuar, ele próprio, como automato.

Na outra situação extrema ambos os rectangulos são vistos como simbolizando participantes (humanos) em *conversação*. A sua interacção neste caso não é uma simples troca "estímulo" - "resposta"- "estímulo" -... mas antes um *dialogo* numa linguagem L que pode ser diferente da metalinguagem L* utilizada pelo observador. Obviamente a escolha de L faz parte do contrato previamente negociado; em geral as interacções por meio de L podem tomar qualquer forma podendo até ser feitas através de dispositivos mediadores (tais como um teclado). Apesar disso contudo L deve ter muitas qualidades de uma linguagem natural; em particular numa interpretação pragmática e semântica para além da sua sintaxe. É também evidente por outro lado que a responsabilidade na manutenção do contrato original é distribuida pelos participantes podendo ambos utilizar expressões com base em L para discutir um determinado assunto ou externalizar operações mentais normalmente de caracter privado.

Existem vários (possíveis) inconvenientes nesta última situação. Em particular

- pode não existir diálogo propriamente dito;
- no caso de existir, qualquer dos participantes pode ser incapaz de o interpretar sem ambiguidade; (recorda-se como caso limite uma situação de delírio num esquizofrénico, por exemplo) etc.

Estas dificuldades podem ser evitadas se o observador externo (adoptando um ponto de vista intermédio) impedir de algum modo uma conversação totalmente "livre" atribuindo a um dos participantes (humano) um papel de agente "entrevistador" ou de "professor". Ou já numa fase mais avançada, substituir esse próprio agente humano por um processador (não exactamente um automato finito) e fazer com que este sistema artificial execute as suas instruções como participante. Particularmente importantes neste caso são as situações em que o participante (humano) visa objectivos que é incapaz de atingir sozinho (situação em que de algum modo funciona como "estudante") desempenhando o outro participante (artificial) o papel de seu "professor".

Nestas condições o esquema da Fig. 2 desempenhará o papel de um sistema de aprendizagem/ensino; não necessáriamente a de um ensino unilateral (em que o "professor" se limita a ajudar o estudante humano a atingir os seus objectivos) mas antes a de um ensino *mútuo* no qual, através de sucessivas interacções com vários estudantes (humanos), o professor (artificial) pode, ele, próprio ir adquirindo mais e mais conhecimentos. Sistemas deste tipo baseados no antigo Caste de Gordon Pask, Scott e Kaligourdis |2| |4| têm sido desenvolvidos desde 1975 nos EUA, Reino Unido e mais recentemente na Holanda.

Para além das suas características computacionais (que neste artigo não têm interesse referir) o que estes sistemas interactivos contêm de relevante é uma filosofia de *comunicação* subjacente a qual, sob muitos aspectos (epistemológico, psicológico, forma de encarar os próprios computadores e consequente forma de modelar e simular as actividades/comportamentos humanos, etc.), se opõe frontalmente a muitos dos métodos

vulgarmente utilizados hoje em dia.

Alguns dos conceitos-chave desta nova abordagem indicam-se em seguida.

i') O primeiro corresponde a uma visão do ser humano que, de algum modo, contrasta com a visão clássica na qual dois seres humanos são usualmente considerados como entidades físicamente distintas. Do ponto de vista da Cibernética não-clássica esta condição deixa de ser necessária. Importante é a distinção entre um ser humano (doravante denominado um indivíduo M) e as suas *perspectivas* (designadas indivíduos P). A um indivíduo M podem corresponder muitas perspectivas que por vezes até podem ser mutuamente contraditórias, divergentes e opostas. Consequentemente a esse indivíduo M irão corresponder também vários indivíduos P. Por outro lado, organizações sociais, comunidades religiosas, escolas de pensamento, grupos políticos etc. compostos por muitos indivíduos M podem ser caracterizados por poucos (no limite um) indivíduos P na medida em que esses indivíduos M compartilham poucas (no limite uma) perspectivas dominantes comuns.

ii') A aquisição/construção dessas perspectivas está obviamente dependente das vivências de cada indivíduo M. Consequentemente essa aquisição/construção está sempre ligada à forma como cada indivíduo M foi observando/interpretando/construindo o seu mundo fisíco, individual e/ou social. Para a Cibernética não-clássica a tese que prevalece em todos esses processos é, portanto, não a tese realista vulgarmente seguida mas antes uma tese idealista no sentido de Dick Gregory |5|, |6| de que os observadores estão ligados ao mundo através dos seus próprios actos de observação. Como tal essa observação é sempre auto ou outrem-referida ("self or other referenced but never it-reference"). Consequentemente não é possível determinar o significado de algo através de um referente externo pré-existente mas, pelo contrário, apreendendo o contexto no qual esse algo existe ou no qual pode ser visto[3]. Sob esta perspectiva conhecimento e verdade não serão portanto realidades objectivas mas antes matéria de concordância num certo universo de discurso. Daí a espantosa importância dada ao aspecto *conversacional* das interacções entre indivíduos M.

iii') Cada indivíduo M interactua com outros indivíduos M enviando/recebendo *mensagens*. Contudo de acordo com ii') a linguagem por meio da qual estas mensagens podem ser descritas não deve ser interpretada de acordo com a visão tradicional (em que essas mensagens quase desempenham o papel de telegramas) mas antes como *construções* produtivas/reprodutivas.

Significa isto que indivíduos M inseridos em esquemas conversacionais não são meros observadores uns dos outros (como a visão clássica postula) mas antes participantes no sentido anteriormente definido.

Este aspecto é crucial e deriva directamente da observação quotidiana. Como Glanville salienta "não há processo nenhum que nos permita entrar dentro da mente dos outros. Em consequência aquilo que cada um de nós faz é construir modelos daquilo que outrém nos diz (através da conversação), re-iterar estes modelos como sendo representações daquilo que perceberam do que perceberam de ... numa regressão teoricamente interminavel até que se atinja um ponto em que cada um dos intervenientes já construiu significados mutuamente partilhados e mutuamente aceites" |7|.

Afirmar nestas condições que dois ou mais indivíduos M quaisquer são participantes numa conversação irá consequentemente implicar uma troca do

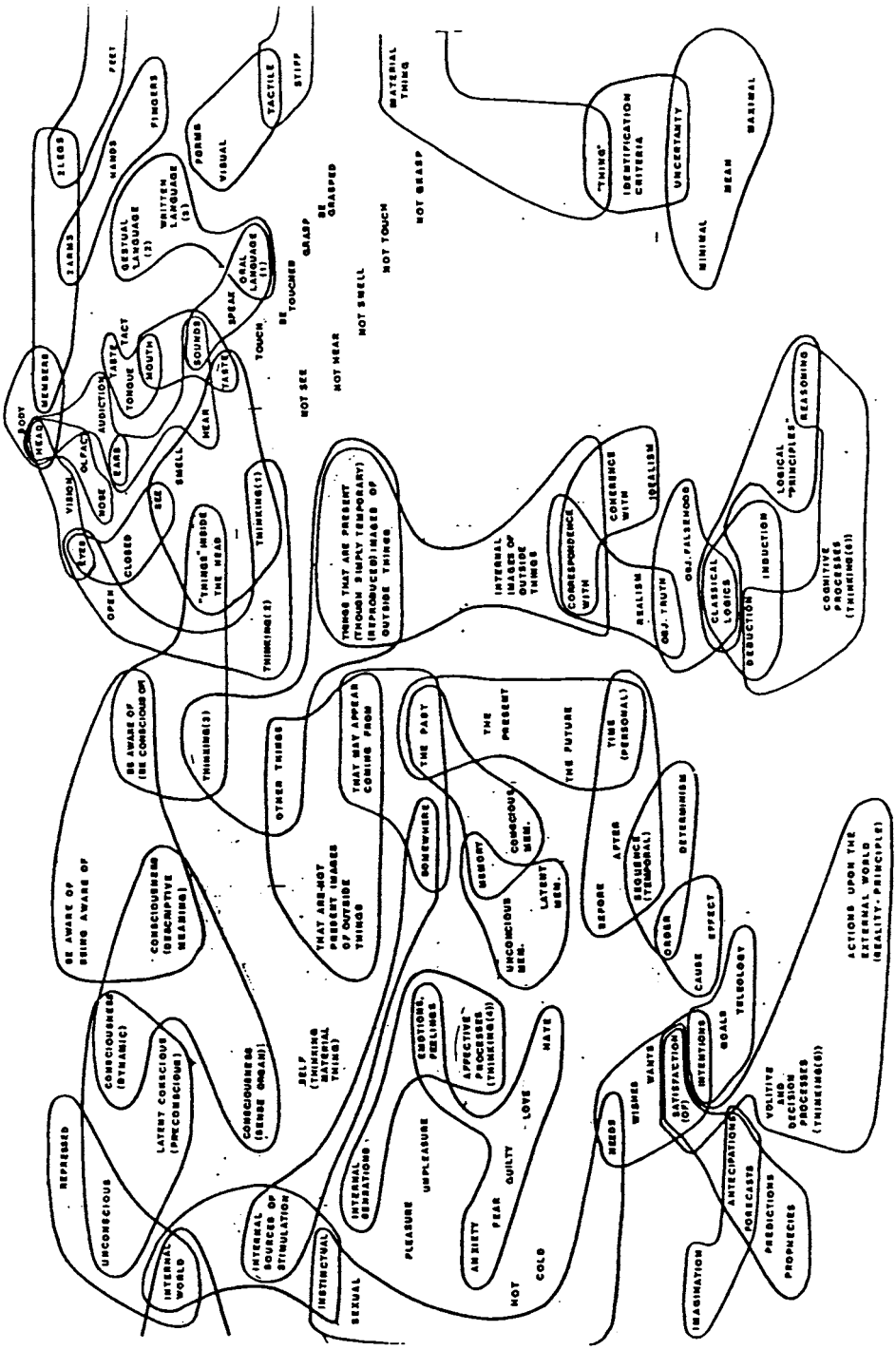

FEET
ZLEGS
HANDS
FINGERS
ZARMS
WRITTEN LANGUAGE (3)
GESTUAL LANGUAGE (2)
FORMS
VISUAL
TACTILE
STIFF
ORAL LANGUAGE (1)
SPEAK
TOUCH
BE TOUCHED
GRASP
BE GRASPED
NOT TOUCH
NOT GRASP
NOT SMELL
NOT HEAR
NOT SEE
MATERIAL THING
"THING"
IDENTIFICATION CRITERIA
UNCERTAINTY
MINIMAL
MEAN
MAXIMAL
BODY
MEMBERS
HEAD
AUDICTION
TASTE
TACT
TONGUE
MOUTH
SOUNDS
TASTE
HEAR
OLFAC
EARS
SMELL
VISION
NOSE
SEE
EYES
OPEN
CLOSED
"THINGS" INSIDE THE HEAD
THINKING(1)
THINKING(2)
THINGS THAT ARE PRESENT (THOUGH SIMPLY TEMPORARY) (REPRODUCED) IMAGES OF OUTSIDE THINGS
INTERNAL IMAGES OF OUTSIDE THINGS
CORRESPONDENCE WITH
COHERENCE WITH
IDEALISM
REALISM
OBJ. TRUTH
OBJ. FALSEHOOD
CLASSICAL LOGICS
DEDUCTION
INDUCTION
LOGICAL "PRINCIPLES"
REASONING
COGNITIVE PROCESSES (THINKING(6))
BE AWARE OF (BE CONSCIOUS OF)
THINKING(3)
OTHER THINGS
THAT MAY APPEAR COMING FROM
THE PAST
THE PRESENT
THE FUTURE
TIME (PERSONAL)
BE AWARE OF BEING AWARE OF
CONSCIOUSNESS (DESCRIPTIVE MEANING)
THAT ARE NOT PRESENT IMAGES OF OUTSIDE THINGS
SOMEWHERE
MEMORY
CONSCIOUS MEM.
LATENT MEM.
UNCONCIOUS MEM.
BEFORE
AFTER
SEQUENCE (TEMPORAL)
DETERMINISM
ORDER
EFFECT
CAUSE
TELEOLOGY
GOALS
INTENTIONS
SATISFACTION (OF)
WANTS
WISHES
NEEDS
ACTIONS UPON THE EXTERNAL WORLD (REALITY-PRINCIPLE)
VOLITIVE AND DECISION PROCESSES (THINKING(5))
ANTICIPATIONS
FORECASTS
PREDICTIONS
PROPHECIES
IMAGINATION
CONSCIOUSNESS (DYNAMIC)
LATENT CONSCIOUS (PRECONSCIOUS)
CONSCIOUSNESS (SENSE ORGAN)
SELF (THINKING MATERIAL THING)
EMOTIONS, FEELINGS
AFFECTIVE PROCESSES (THINKING(4))
HATE
LOVE
GUILTY
FEAR
ANXIETY
HOT
COLD
UNPLEASURE
PLEASURE
INTERNAL SENSATIONS
REPRESSED
UNCONSCIOUS
INTERNAL WORLD
INTERNAL SOURCES OF STIMULATION
INSTINCTUAL
SEXUAL

FIGURA 3

significado que cada conceito tem para cada um relativamente ao topico central da conversação até que (através de sucessivas interacções/trocas) cada um seja capaz de entender esse conceito não apenas de acordo com os seus próprios termos e as suas próprias perspectivas mas também de acordo com os termos e as perspectivas do outro.

iv') Segue-se que o conceito de verdade desta concordância conversacional é uma verdade de *coerência* e não uma verdade de correspondência (embora como é óbvio esta última não seja excluída factualmente). Cf. Fig. 3.

Tradicionalmente estas relações são traduzidas por uma matriz de adjacência booleana em que à existência de uma ligação entre a palavra x e a palavra y corresponderá um valor 1 na matriz e à não existência de relação o valor O. Operações efectuadas sobre esta matriz permitem definir, por exemplo, o que são os nós iniciais da rede (o equivalente a conceitos "elementares"), nós finais, nós convergentes (correspondentes à condensação de um conjunto de termos), nós divergentes (implicando que um certo conceito X pode estar ligado a vários conceitos Y, Z, W ... e traduzindo consequentemente possíveis ambiguidades) etc. Grupos de nós que satisfaçam mutuamente as propriedades de reflexividade, simetria e transitividade irão permitir, por outro lado, especificar classes de equivalência as quais servirão de base para as classificações anteriormente referidas, para o estabelecimento de hierarquias, etc.

Através da composição destes operadores elementares podem igualmente definir-se operações mais complexas. Por exemplo atribuindo uma "extensão" unitária a um conceito elementar (nó inicial) e calculando a sua propagação e soma (nos nós convergentes) será também possível justificar a razão porque a noção psicológica de "coisa" é mais "extensa" que a noção de "objecto", este mais "extenso" que o de "móvel", este que o de "cadeira" etc.[4]

Muitos mais procedimentos deste tipo seriam possíveis de indicar. Não é este contudo o ponto importante. Importante é sim, salientar que quer a teoria dos grafos/hipergrafos na qual elas estão baseadas quer outras teorias que a adoptaram ou adoptam como processos de representação (e aqui incluo quer os aspectos mais tradicionais da Teoria Geral dos Sistemas de Klir ou Mesarovic quer a Teoria dos Automatos finitos, determinísticos ou probabilísticos, quer ainda a teoria das redes de Petri recentemente "redescoberta") têm tido como último suporte lógico-matemático os postulados da Algebra booleana clássica. Nos últimos anos contudo, e muito em particular nas duas últimas dezenas, quer físicos, quer matemáticos, quer engenheiros, projectistas de sistemas de controlo, peritos em Cibernética, Inteligência Artificial Lógica, Linguística, Psicologia etc. começaram a ser defrontados com um conjunto de problemas relativamente aos quais esses postulados se mostraram (mostram) inadequados e/ou insuficientes.

Alguns exemplos irão ilustrar esta afirmação.

i'') Começarei pela Física referindo-me a um problema que é usualmente encontrado na análise de uma classe particular de sistemas dinâmicos denominada *sistemas não determinísticos*. Em termos muitos gerais a questão que está subjacente a esta classe de sistemas é a de que durante a sua evolução temporal um determinado estado presente pode dar origem a um conjunto de *possíveis* estados seguintes quer teórica quer factualmente. Dito de outra forma: como este tipo de sistemas envolvem incertezas na sua estrutura e/ou comportamentos não é possível em termos de modelação matemática (e consequente simulação por meio de computador) saber se o sistema é, na *realidade*, determinístico ou não. Um dos exemplos mais flagrantes de situações deste tipo corresponde à descrita

na Fig.4.
A regra através da qual os pontos X_k são gerados consiste em definir a intersecção da vertical de um dado ponto X_0 inicial com o triângulo ABC inscrito no quadrado e a partir daí traçar uma horizontal que ao intersectar a diagonal do quadrado vai definir um novo X_{k+1} repetindo-se em seguida todo o processo.

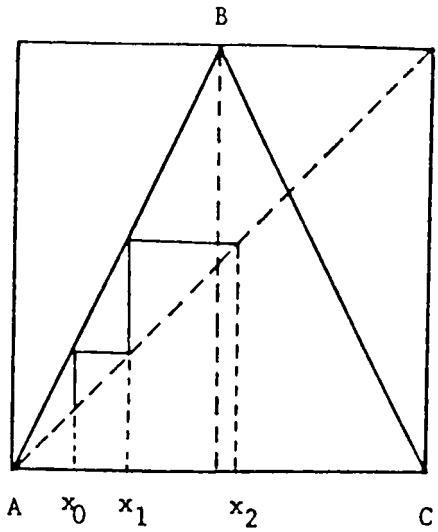


FIGURA 4

Se se associar a cada ponto X_k um valor 0 se X_k < que 1/2 e 1 se X_k > que 1/2 então nestas condições a sequência de pontos gerados por um processo que é visívelmente determinístico torna-se matematicamente indistinguível de um processo de Bernoulli. Ou seja: se se supuser que a sequência de estados de 0s e 1s corresponde à sequência de estados do sistema *não há processo matemático* (à parte a observação do mecanismo gerador interno do próprio sistema) de determinar se o seu comportamento (externamente observado) pode ser modelado por meio de regras determinísticas ou através de um processo estocástico |11|.

ii'') Um problema análogo (típico dos chamados sistemas não lineares) indica-se na Fig.5.

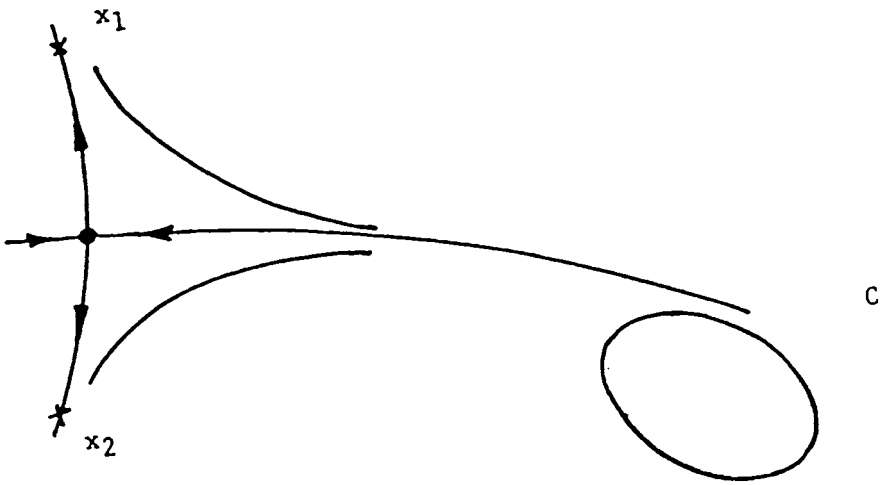


FIGURA 5

Trata-se de um sistema dinâmico com dois pontos de equilíbrio X_1 e X_2 e uma separatriz que vai intersectar um ponto de sela e tende para uma orbita fechada C. Uma vez que qualquer ponto de C adere à região de atracção de X1 e X2 a evolução do sistema a partir de um determinado valor inicial X0 pertencente a C é praticamente indeterminada. Ou seja, uma vez mais um sistema determinístico pode dar origem a uma total incapacidade de previsão da sua evolução. A situação é mais grave ainda porque estas incertezas associadas à descrição matemática de um sistema (particularmente em sistemas não lineares) são tradicionalmente "solucionadas" introduzindo sucessivas hipóteses simplificativas. É usual um projectista de um sistema de controlo linealizar esse sistema (um dos processos de o modelar e simular em computador) através de 10 ou mesmo 15 hipóteses simplificativas. O problema é que se a incerteza associada à descrição do protótipo não for bem compreendida (e como se viu há situações de incerteza estrutural) o seu modelo vai dar origem a incertezas ainda maiores; em particular, se o comportamento nominal do modelo não for o mesmo do do protótipo e o modelo for estruturalmente instável.

iii")O terceiro exemplo que vou citar tem características aparentemente diferentes. Tal como na situação anterior também envolve uma órbita fechada; essa órbita todavia refere-se agora a um (ou vários) ciclo(s) fechados em grafos complexos (análogos por exemplo ao da Fig.3) com grande número de nós (palavras) e consequentemente grande número de relações.

A questão que aqui está em causa é bastante semelhante a um problema examinado em Investigação Operacional, denominado na generalidade "problema do caixeiro viajante". Trata-se, em resumo do planeamento de um circuito fechado de tal forma que os seus nós constituintes sejam obrigatoriamente percorridos (pelo menos) uma vez tendo em atenção certas condições de

optimização. Determinar circuitos deste tipo em grafos com um número relativamente pequeno de nós é uma operação para a qual existem algorítmos relativamente eficientes. A questão é que estes algorítmos requerem na maior parte dos casos uma análise exaustiva de *todas* as potenciais soluções as quais, no caso do problema anterior, equivalem ao cálculo de n! sendo n o número de circuitos possíveis. Se se reparar que por exemplo que 6! = 720 e que 9! = 362.880 é fácil compreender como a inclusão de um simples nó (ou palavra) adicional vai originar um aumento astronómico das possibilidades a analisar. Imagine-se o que seria fazer um cálculo semelhante para o vocábulário normal de um adulto medianamente culto! Apesar disto contudo este tipo de raciocínio baseado em combinatorias é não só frequentemente utilizado na simulação de processos mentais[5] como também (e este é o outro ponto importante a reter deste exemplo) porque muitos desses próprios processos parecem obedecer a descrições tipo "círculo fechado". São conhecidas com efeito muitas reacções psicológicas (usualmente de carácter patológico) em que o indivíduo ao fim de percorrer um determinado caminho (subentenda-se "raciocínio" ou um outro comportamento) volta invariavelmente às afirmações ou comportamentos iniciais repetindo-se o processo vezes sem fim.

iv")Com esta referência a processos mentais retomo de algum modo aquilo que afirmei anteriormente acerca da Cibernética não-clássica considerar os seres humanos em si (como entidades globais) e não como "it" das observações. Só que nestas condições ao referir-me a esses processos, ao que na generalidade é designado por *pensamento*, novas questões aparecem.

Com efeito "pensamento" é em si mesmo um termo altamente ambíguo. Thomson |14| por exemplo associa-lhe seis significados diferentes:

- aquilo que os psicólogos denominam *pensamento autístico* (relacionado com fantasia, com o "sonhar acordado", com a vaga sensação de algo incontrolado que está em permanente mudança interna);
- como sinónimo de *lembrar*, *rememorar*, *evocar* (que pouco tem a ver com "memória" no sentido computacional mas antes como reprodução/construção[6]);
- no sentido de *imaginar* (diferente da fantasia na medida em que o imaginar é suposto ser uma construção com uma coerência que é modelada/derivada da observação de acontecimentos/coisas tais como normalmente ocorrem/aparecem);
- como sinónimo de paradigma/*padrão de comportamento* explícito ou subjacente (por exemplo quando alguém está a executar uma determinada tarefa prática e se lhe diz "Pense no que está a fazer!");
- num sentido ligado a *crença*, ao "acreditar em" (quando alguém num grupo social está a afirmar "o que pensa" acerca de política, religião ou dos últimos resultados obtidos pela sua equipe de futebol, "pensar" aqui é sinónimo de "acreditar");
- num sentido ligado ao *raciocínio*, à *reflexão* (meditação), ao "cogitatio" cartesiano, àquilo que antigamente era associado a intelecto. Contudo, mesmo de acordo com este último significado (aquele que mais frequentemente é evocado e que aparentemente é mais simples por parecer implicar "apenas" o uso de regras inferenciais) "pensamento" continua a conter profundas ambiguidades. Ryle |15| por exemplo ao analisar a actividade intelectual de um cientista não só constatou que ela subentende um conjunto de actividades subsidiádias (tais como cálculos, comparação de diversas hipóteses, utilização de diferentes

estratégias na manipulação dessas hipóteses, perguntas e respostas quer pessoais quer sociais, detecção e discriminação de discrepâncias e contradições, tomadas de decisão etc) como verificou também que essa própria actividade é altamente individualizada. Como tal, fortemente dependente quer da história, experiência e aprendizagem pessoais quer dos próprios contextos nos quais ele pode estar inserido.

Como Ryle concluiu, "pensamento" aparece assim como uma actividade *polimorfa* englobando sempre (mesmo no "pensar" científico) fantasias, imaginações, evocações, paradigmas (aqui entendido como padrão de comportamento mental) crenças e raciocínio.

Acerca da simulação deste último tipo de actividade, aquilo a que poderia chamar simulação de processos cognitivos puros (se é que eles existem como tal), nada há a acrescentar relativamente ao que a Inteligência Artificial já conseguiu. Mas o que dizer por exemplo acerca das crenças, não das crenças epistémicas que de um modo ou outro são redutíveis àquele tipo de processos mas antes das crenças deônticas as quais parecem envolver mecanismos e processos que (na generalidade) designamos por "afectivos"? Que dizer também acerca da simulação de todo este universo dos afectos, sentimentos etc. ainda praticamente inexplorado em termos de simulação e que (para além de tudo o que contém como área de investigação em si mesma) está, por exemplo, tão intimamente ligado ao reforço e motivação nos processos de aprendizagem? Que dizer finalmente do aspecto construtivo da memória humana que nos fornece imagens de nós próprios a entrar num quarto ou a nadar numa piscina?

v") Para terminar esta série de exemplos queria referir mais duas questões nas quais mais uma vez o tratamento, modelação e simulação de processos com base na Algebra booleana clássica parece mostrar-se insuficiente e inadequada.

A primeira destas questões provem da já conhecida descrição dada por Konrad Lorenz acerca das suas experiências de impressão (imprinting) com patinhos depois de ter substituido a mãe pata:

> "Eu estava a felicitar-me pela obediência e exactidão com que os meus patinhos me seguiam quando levantei os olhos de súbito e vi a cerca do jardim coroada por fila de rostos de uma palidez funérea: um grupo de turistas colocara-se ao longo da cerca e observava-me de olhos arregalados. Os patos estavam ocultos pela relva alta e tudo o que os turistas viam era o meu comportamento inexplicável e de facto, aparentemente louco" |16|.

Os aspectos que aqui quero frisar não se referem nem à Etologia de Lorenz nem à pragmática da comunicação que a afirmação tão bem ilustra mas antes ao epíteto de "louco" atribuído a *si mesmo* pelo próprio Lorenz e provavelmente também pelos turistas que observavam "um homem barbudo, rastejando agachando-se, espiando entre os arbustos, percorrendo veredas em forma de oito, olhando por cima do ombro e grasnindo sem interrupção" |16|. Na realidade ao fazer essa atribuição Lorenz (e os turistas) estão a levantar dois problemas-tipo que a Cibernética não-clássica tem vindo a abordar há vários anos.

O primeiro desses problemas-tipo refere-se à questão da *auto-referência*[7] (self-reference), ou seja a padrões, esquemas paradigmas *anteriormente* interiorizados, que são evocados em dadas situações e servem como sistemas de *comparação* (normalmente através de processos de rectroação) no julgamento

de situações presentes e no planeamento de acções futuras.

O segundo problema-tipo está relacionado com a distinção entre o "louco" e o "normal" e envolve, no fundo, os critérios utilizados para estabelecer o que é patológico e não-patológico *como se* estas classes fossem disjuntas no sentido da Algebra Clássica i.e. "se pertence à classe dos loucos então não pertence à classe dos normais" e vice versa. Qualquer psiquiatra sabe, todavia, que este tipo de classificação é erróneo. Isto porque tais termos, bem assim como um imenso conjunto de outras palavras, expressões e regras de inferência utilizadas vulgarmente em linguagem natural[8] se refere a elementos ou conjuntos cujos limites *não são* bem definidos. Em linguagem técnica são termos *vagos* (fuzzy) e o seu estudo deu origem a uma extensão da Algebra booleana clássica denominada *Teoria dos Subconjuntos Vagos*.

A teoria começou a ser formulada na década dos 60 por Zadeh muito embora a expressão francesa "ensemble flou" já tivesse sido utilizada em 1951 por Menger ao estabelecer (com base numa interpretação probabilística) aquilo que é hoje a operação max-prod de uma relação transitiva "fuzzy". A ideia de Zadeh era também bastante semelhante à interpretação semântica de "vagueza" (vagueness) apresentada por Black em 1937 a propósito daquilo que ele denominou "perfis de consistência", os antepassados das actuais funções de pertença (membership functions).

Não faz sentido apresentar nesta comunicação os fundamentos da Teoria dos Subconjuntos "fuzzy". Todavia para uma compreensão global da estrutura deste artigo três desses aspectos não podem deixar de ser salientados. Em primeiro lugar desejava frisar que a algebra "fuzzy" (e a sua lógica subjacente) é uma algebra multivalente que (na interpretação de Zadeh) compreende um conjunto de valores *entre* 0 e 1 e não apenas dois valores 0 ou 1 como na algebra booleana clássica. Esta pode, portanto, ser considerado um caso limite, uma particularização daquela. Em segundo lugar gostaria também de salientar que nesta algebra "fuzzy" está implícita a utilização de um *referencial* (daí a designação de *subconjunto* fuzzy). Esse referencial pode variar no tempo ou no espaço (um espaço de configurações no sentido de Lanczos); mas qualquer que seja essa variação é *sempre* em relação a esse referencial que um elemento ou subconjunto é *numericamente* avaliado ou medido (através de funções que dão o grau de pertença dos elementos/subconjuntos relativamente ao referencial[9]). Em terceiro lugar - e este é claramente o ponto mais importante a salientar - porque praticamente todas as situações-tipo anteriormente exemplificadas caiem dentro do âmbito da Algebra "fuzzy".

Com efeito

- todas as palavras e expressões citadas em v") correspondem a funções ou operadores "fuzzy" típicos;
- a ideia de referencial (fixo ou variável) também citada em v") é como afirmei parte integrante dos fundamentos teóricos dessa algebra;
- a questão da subjectividade e das crenças deônticas corresponde a situações cuja descrição em linguagem natural é feita utilizando expressões do tipo "grande", "pequeno" "bastante" etc. redutíveis portanto ao primeiro item;
- finalmente os problemas referidos em i") e ii") englobando incertezas e ambiguidades estruturais e/ou comportamentais de certos sistemas dinâmicos e as (consequentes) dificuldades na sua modelação e simulação dizem respeito quer ao conceito de medida |-fuzzy introduzida por Sugeno |17| |18| |19| quer a chamada Teoria dos Controladores Linguísticos |20| |21|, uma técnica "fuzzy" permitindo simular em computador não só sistemas não lineares (sem os inconvenientes dos processos de controlo clássico) como também motores de

inferência utilizando regras de tipo "Se X_1 é muito grande e X_2 é bastante pequeno e X_3....então Y é medianamente pequeno" ou "se X_1 aumenta pouco ou X_2 diminui muito então Y é muito grande" com X_1, X_2, X_3 etc. representando entradas e Y saídas.

Se se comparar este tipo de descrição (dinâmica) com a que é dada acerca do funcionamento de certos mecanismos mentais (relacionando por exemplo a agressividade com constrangimentos internos ou sociais) compreender-se-ão facilmente as tremendas potencialidades da Teoria dos Subconjuntos Vagos na modelação e simulação desses processos.

Com isto não quero de forma nenhuma afirmar que os problemas focados (ou outros com eles directa ou indirectamente relacionados) estão já resolvidos por meio da abordagem "fuzzy". Na minha opinião o que elas fornecem é, sim, um suporte *teórico* mais vasto e promissor que o da Algebra e Lógicas clássicas, uma base onde aquelas questões aparentam inserir-se melhor que noutras abordagens utilizadas hoje em dia[10]. Na passagem da teoria à prática (em que "prática" corresponde aqui à implementação *efectiva* de modelos mentais) há contudo um universo de problemas ainda por solucionar.

Uma questão que nesse aspecto me aparece crucial é a questão do *tempo*, o mesmo tempo que separava as áreas da Física e da Biologia e que a Cibernética de algum modo uniu. Do nascimento à morte, da transformação de animais altriciais, nidicolas, imaturos e indefesos ao nascer até à fase adulta, todos os seres humanos de todas as épocas da humanidade podem encarar-se como "máquinas" temporais quer a nível biológico (maturação, ritmos circadianos puramente endógenos ou em interrelação com os ritmos exógenos naturais, ciclos de satisfação/insatisfação de necessidades orgânicas assimiladoras e excretoras, etc.) quer a nível psicológico.

O tempo está com efeito subjacente ao desenvolvimento aqui incluindo por exemplo a capacidade de estabelecer distinções entre acontecimentos ou de lhes conferir significado (o que pressupõe no mínimo a análise, modelação e simulação das estruturas afectivas e cognitivas bem assim como da sua articulação e evolução temporal).

O tempo está íntimamente ligado à génese do objecto permanente (psicológico, físico e no sentido psicanalítico) e às relações de causalidade/finalidade que se podem estabelecer entre esses objectos. Tempo está também envolvido na aprendizagem (não só na avaliação dos sucessos como também naquilo que "falhou" no passado e, consequentemente, servirá no futuro como meio de evitar novos insucessos), na construção antecipada de "situações" que posteriormente serão projectadas para um futuro (condicionadas às vezes pela experiências passada). O tempo pré-existe por outro lado em qualquer troca conversacional na busca (e modificação) das auto-referencias e perspectivas, na variação do grau de confiança no interlocutor, na extensão ou redução da área de interesse em torno do tópico de conversação, nas reacções comportamentais do tipo "como antes" (regressões) em vez de "como agora", etc. O tempo estabelece, finalmente, a grande distinção teorética e conceptual entre as abordagens com base nas lógicas modais e as lógicas "fuzzy", uma das razões fundamentais que deram precisamente origem ao aparecimento da Teoria dos Subconjuntos Vagos |22| |23|.

Tal como afirmei anteriormente não é minha intenção clamar pela "verdade" ou falsidade das abordagens "sem tempo" das lógicas modais ou "com tempo" das lógicas "fuzzy". Na realidade estando a exploração, modelação e simulação do mundo mental apenas no início (nas palavras do Prof. Murray Eden da Universidade de Maryland "numa fase equivalente à teoria dos quatro humores" utilizados na Idade Média para explicar a Fisiologia Humana) seria erróneo da minha parte dar enfase a este ou aquele processo

apenas em detrimento dos outros.

Para evitar esses erros é que a Cibernética introduziu a Teoria das Perspectivas; e é afinal para conversar sobre elas que aqui nos encontramos.

Notas

(1) É de notar que a distinção organismo/ambiente se mantém independentemente da natureza de observações posteriores; os rectangulos delimitantes destas duas entidades funcionam assim como verdadeiras fronteiras simbólicas dessa distinção.

(2) Embora utilizada de acordo com a teoria de Automatos quer com as regras da Psicolinguística matemática esta "linguagem" é, contudo, uma entidade sintática apenas, não uma linguagem no sentido vulgar de linguagem natural.

(3) Compare-se esta tese idealista subjacente a toda a Cibernética não-clássica com as mudanças operadas na Física desde o aparecimento dos fenómenos quânticos.

(4) Operações e procedimentos deste tipo estão na base quer da resposta à questão "O que é X?". (que em linguagem natural corresponde à utilização de duas operações sequenciais correspondentes a primeira a um deslocamento para um nó de extensão superior seguida por um deslocamento para os nós de extensão inferior que X engloba) assim como à bem conhecida reacção, estudada em Psicologia Experimental, referente à exclusão de palavras que não são coerentes com uma dada sequência. Pense por exemplo naquilo que "está a mais" ou não pertence às duas sequências |10|:

Arranha-Céus - Catedral - Templo - Oração

Catedral - Oração - Templo - Arranha-Céus

(5) Ver por exemplo os processos de estimulação inventiva de F. Zwicky |12| ou a operação de saturação de uma rede de autoria de Peter Clark |13|.

(6) O leitor tente recordar a última vez que entrou no seu quarto, no seu gabinete ou nadou numa piscina e analise aquilo que a sua "memória" lhe forneceu.

(7) Examinada particularmente por Ranulph Glanville, Maturana e Varella.

(8) Por exemplo velho, novo, muito, pouco, grande, pequeno, bastante, quase, mais, menos, mais ou menos, assim assim, estar próximo de, estar longe de, ser parecido com etc.

(9) Compare-se esta afirmação com o que foi dito a propósito de Lorenz, dos turistas ou das perspectivas na Teoria da Conversação.

(10) Em particular as lógicas modais que, simultaneamente com as lógicas "fuzzy", estão a ter presentemente um enorme desenvolvimento.

Bibliografia

|1| GEORGE, F. *"The Brain as a Computer"*, Pergamon Press 1973

|2| PASK, G. *"Conversation, Cognition and Learning"*, Elsevier 1975

|3| von FOESTER, H. *"Cybernetics of Cybernetics"* BCL Report 73; 38

|4| PASK, G. and SCOTT, B. "CASTE: *"A system for exhibiting strategies and regulating uncertainties"*, Int. J. Man-Machine Studies 1973

|5| GREGORY, D. *"Philosophy and Practice in Knowledge Conversation"*, Human Productivity Enhancement Vol 2, Praeger 1987

|6| GREGORY, D. and PASK, G. *"Conversational Systems"*. Zeiner ed it. 1983
|7| GLANVILLE, R. *"Conversation Theory: a terse summary"*, April 1987
|8| PASK, G. *"Developments in Conversation Theory*. Part II: *Protologics, their origin and their development"*, Jan 1987
|9| PASK, G. *"Conversation Theory and Protologics"*, April 1987
|10| MAYER, R. *"Thinking and Problem Solving: an Introduction to Human Cognition and Learning"*, Scott, Foresman & C° 1977
|11| de GLAS, M. *"A mathematical Theory of Fuzzy Systems"*, Fuzzy information and Decision Processes, North Holland 1985
|12| KAUFMANN, A. *"An introduction to Fuzzy Subsets Theory"*, Vol IV, Masson 1977
|13| CLARK, P. *"Saturation, Bifurcation and Analogy in Protologics: a position paper"* fev 1985
|14| THOMSON, R. *"The Psychology of Thinking"*, Pelican Books 1977
|15| RYLE, G. *"The Concept of Mind"*, London (Hutchinson) 1949
|16| WATZLAWICK, P. and BEAVIN, J. and JACKSON, D. *"Pragmatics of Human Communication: a study of interactional pathologies and paradoxes"*, W.W. Norton 1967
|17| KLIR, G. and FOLGER, T. *"Fuzzy Sets, Uncertainty and Information"*, Prentice Hall 1988
|18| DUBOIS, D: and PRADE, H. *"Fuzzy Sets and Systems: Theory and Applications"*, Academic Press Vol 144, 1987
|19| KLIR, G. *"Where do we stand on measures of uncertainty, ambiguity, fuzzyness, and the like?"*, Fuzzy Sets and Systems 24 (1987) North Holland
|20| SUGENO, M. *"Industrial Applications of Fuzzy Control"*, North Holland 1985
|21| DUBOIS, D. and PRADE, H. *"Théorie des Possibilités"*, Masson 1987
|22| MEDINA MARTINS, P. *"Towards Historical Experimentation: Cybernetic models of historical evolution"*, Proceedings–International Congress on Cybernetics and System Research, Wien 1988
|23| MEDINA MARTINS, P. *"Preliminary contribution for the area of Learning, Knowledge Acquisition and Knowledge Representation"*, ESPRIT PROJECT 1988

Rui M. Boavista Vieira Marques
Tokyo Univ., Tokyo (Japan)

Information Theory - Communicative Data

"If you get careless or go romanticizing scientific information, given it a flourish here and there, nature will soon make a complete fool out of you."
Robert M. Pirsig
"Zen and the art of motorcycle maintenance", page 101

Abstract

Contemporary science is undergoing a radical change towards adaptability, multiplicity, plural uniqueness, temporality, and complexity.

Technological innovations, particularly in the fields of communication and information, has created new feedback loops between humans and utensils at the various levels of social, psico and universal relative positionning.

Aware of the already difused borders between classicaly defined miles-away fields of science, we introduce comparative models of information transfer, by reduction of redundancy.

The first application is a real-time digital information compression scheme where adaptiveness to different types of input data is stressed, and by second we undertake a social communications model and evaluate entropy considerations for japanese environment absence of "glasnost".

Resumo

A ciência contemporânea está a prosseguir uma mudança radical no sentido da adaptabilidade, da multiplicidade, da unidade plural, da temporalidade e da complexidade.

As inovações tecnológicas, particularmente nos campos da comunicação e da informação, introduziram em retroacção novos laços entre homens e utensílios nos vários níveis — sociais, psicológicos, universais — do posicionamento relativo.

Conscientes de que as fronteiras entre dominios cientificos classicamente bem delimitados são já difusas, apresentamos modelos comparativos de transferência de informação por redução das redundâncias.

A primeira aplicação consiste num esquema de compressão da informação digital em tempo real no qual a adaptação a diferentes tipos de input de dados é pressionada e, seguidamente, propômos um modelo de comunicações sociais e avaliamos aspectos de entropia no ambiente japonês de ausência de "glasnot".

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 243/252
1992, APDC & SPF, Lisboa

Summary

Introduction

- From Shannon Theory to Universal Modeling
- Model Complexity: Ziv–Lempel, Kolmogorov–Chatin

Downstream Application

Adaptive Data Compression Using Self–adjusting Data Structures;
- A free parameter model of data compression

Upstream Application

Japanese Communications Model
- A circular parameter model of data elasticity
- Complexity and Efficiency

Open Problems/Postface

Introduction

Shannon's fundamental theorem of coding states that, given messages randomly generated from a model, it is impossible to encode them into less bits (on average) than the entropy of the model.

A message can be encoded with respect to a model, using either Huffman or arithmetic coding. The former method is frequently advocated as the best possible technique for reducing the encoded data rate, but recent studies on other adaptive methods are presenting better results. However still, it will be used as a benchmark reference and briefly, given that each symbol in the alphabet must translate into an integral number of bits in the encoding, Huffman coding achieves minimum redundancy. In other words, it performs optimally if all symbol probabilities are integral powers of 1/2, but that's not the most often case.

In arithmetic coding, a message is represented by an interval of real numbers between 0 and 1. As the message becomes longer, the interval needed to represent it becomes smaller and the number of bits needed to specify that interval grows. Sucessive symbols of the message reduce the size of the interval in accordance with the symbol probabilities generated by the model. The more likely symbols reduce the range by less than the unlikely symbols and hence add fewer bits to the message. As known, the effectiveness of any model can be measured by the entropy of the message with respect to it, usually expressed in bits/symbol.

In universal modeling, the measure of complexity assigned to the models, is equationed by a theorem which states that models that use any kind of alphabet extension are inferior to the best model using no alphabet extensions at all.

The need of a universal modeler encoder comes from the out–standing situation where no well defined statistical information source to which the code can be tuned, is available. In broad terms modeling involves a determination of certain source–strings events and their contexts, which uniquely describe the source string. Rissanen & Langdon [1] regard the model as consisting of two parts: 1) the structure which is the set of events and their

context, and 2) the parameters which are the probabilities assigned to the events. The structure is intended to capture the redundancies in the entire set of source strings under consideration, while the parameters are tailored to each individual string separately. This way a degree of "universality" can be achieved.

In the non-adaptive case, complexity of the model is determined by the number of parameters, and in adaptive ones, although parameters need not to be sent to the decoder, because there is an algorithm calculating values for each symbol along the string, the decoder number of registers is likely to be the same.

Alphabet extention is a process in which the original source is parsed into substrings, which themselves become the statistical events to be encoded. As known, a restriction on number of parameters is necessary, because any model's capability to provide compression must be weighted against its complexity, of which the number in its description is a reasonable measure.

The Lempel-Ziv algorithm consists of a rule of parsing strings of symbols from a finite alphabet into substrings or words whose lengths do not exceed a prescribed defined integer and a coding scheme that maps these substrings sequentially into uniquely decypherable codewords of fixed length. The strings are selected so that they have nearly equal probability of occurence. As a result, frequently occuring symbols are grouped into longer strings, whereas infrequent symbols appear in short strings. This strategy is effective at exploiting redundancy due to symbol frequency, character repetition and higher-usage patterns.

Under these circunstances, as we try to determine complexity of a finite sequence, Lempel and Ziv proposed a new approach to the early works in the area that linked the notion of complexity of a given sequence to that of an algorithm by which the sequence is suposed to be generated (Kolmogorov [2]). It consists in, rather than evaluate the complexity of a finite sequence from the point of view of simple self-delimiting learning machine, adds a new word to its memory every time it discovers a substring of consecutive digits not previously encountered. The size of the compiled vocabulary and the rate at which new words are encountered serve as basic ingredients in the evaluation of the complexity of the sequence. With this new understanding and the restriction of the encoders and decoders to finite-state-machines was possible further development and refinement of Kolmogorov/Chaitin pioneering ideas to a concept of finite-state-complexity which in practice corresponds to the largest number of distinct phrases whose concatenation forms the initial string.

As referred in the previous paragraph, Kolmogorov and Chaitin's [3] studies of the state complexity of Turing machines were of much interest, proposing the use of the lenght of the shortest binary program, when fed into a given algorithm will cause it to produce a specified sequence, as a measure for the complexity of that sequence with respect to the given algorithm. The intuition that the formalism programs could be concatenated and some comparison with the properties of entropy function could be conclusive in that, there is a persuasive analogy between the entropy concept of information theory and the size of programs. Recent other concepts as uniform complexity and program-size complexity defined as the size in bits of the shortest self-delimiting program for calculating strings have been introduced. This last concept was applied to the notion of optimal universal computer and nomenclature as information-theoretic complexity or as algorithmic information theory began to appear in the current field literature.

The latest improvement was supplied by developments in the area of arithmetic coding where complexity is considered as an extension of enumerative procedure calculation.

However in order to improve speed and robustness of adaptive schemes, coding arbitrarily distributed integers is often called for in use with more sophisticated models of text, image or other data. Consider for instance the localy adaptive data compression scheme of Bentley et al. [5], in whick the encoder and decoder cache the last N different words seen. A word present in the cache is transmitted by sending a cache index and one not present is transmitted by sending a new–word marker followed by the characters of the word. This is an excellent model for text in which words are used frequently over short intervals and then fall into disuse. We call that particularity, "locality of reference" and in next chapter we introduce theoretical and practical results of a fast and robust locally adaptive compression scheme.

Downstream application:

ADAPTIVE DATA COMPRESSION USING SELF–ADJUSTING DATA STRUCTURES [6]

"I wonder if my readers know the colour of that darkness seen by candlelight"
Jun'ichiro Tanizaki
"In praise of the shadows", page 34

Most compression techniques are not flexible enough to process different types of redundancy. Our method dynamically adapts to the redundancy characteristics of the data, requiring only one pass over the data being compressed. A data compression scheme that expoits locality of reference, such as occurs when words are used frequently over short intervals and then fall into long periods of disuse, is described and implemented. If the message ensemble to be transmited exhibits locality of reference (i.e. if the local frequency of words change dramatically within the message), the scheme performs better than Huffman coding because a word will have a short encoding when it is used frequently and a long encoding when it is used rarely. Furthermore the codeword will be halved monotonically if the same word will sequencially be brought–up, effect that enables a smother and much moderate behaviour than comparable move–to–root or similar schemes.

We also explore contextuality by implementing a modified splaying function, using a binary search splay tree type [6], in which if words from one subtree of the code tree are repeatedly referenced, splaying will shorten the codewords of all words in that subtree.

The scheme is based in a simple and fast running–time heuristic for self–adjusting operations on a binary search tree. It describes a defined word scheme that uses a technique from other field domain, that deals with defined words self–adjusting non–sequential search, in which we wish to maintain a dictionary of words so that frequently accessed words are gradually well positioned near the most accessible place: in our case the root of a binary tree.

This form of compression is "transparent" or noiseless: the expanded data is an exact replica of the input data, and the compression apparatus is given no special program information; such as data type or usage statistics. A major depart from standard splay heuristics is the non–lexico ordering and a major improvment is done on performance characteristics.

From the results, we derive that it performs in same cases better that dynamic Huffman coding and also better than a good implementation of the related move–to–front heuristics. The present scheme has many implementation advantages: it is simple, allows fast encoding, requires only one pass over the data to be compressed and memory requirements

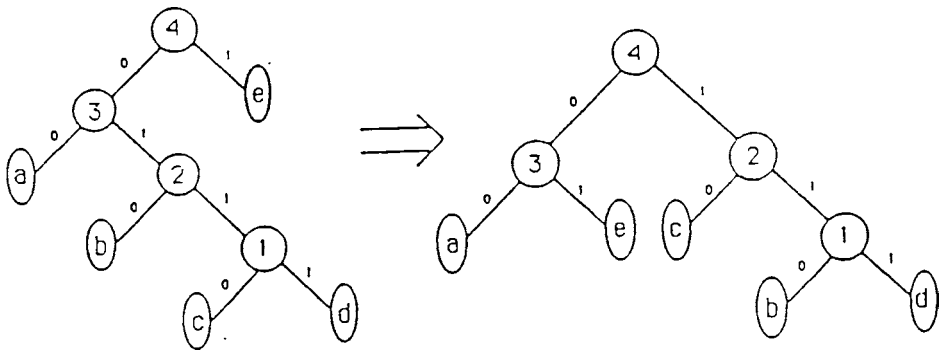
are very short compared with other recent adaptive methods.

Theoretical & Practical ANALYSIS OF PERFORMANCE

A code is dynamic if the mapping from the set of messages to the set of codewords changes over time. The use of a splay tree – a self-adjusting form of binary search tree – will enable the data structure to be in a arbitrary state, but during each operation, a simple restructuring rule intended to improve efficacy of future operations will be applied. It consists in moving the accessed leaf of the tree to the root, or to a position closer to the root by performing a sequence of rotations along the original path from the leaf to the root.

In the figure, accessed leaf with contents "c" is moved to relative closer position to the root. As codeword constitution corresponds to path identification, if next sucessive occurence is "c", the codeword has halved from "0110" to "10". In order to relieve amortized operations task (reduce number of rotations), the lexicographical order will not be preserved and an auxiliar data structure will have to be used, in our case, a hash table.

To analyse performance characteristics of our modified splay algorithm, we do some comparisons with move-to-front heuristics (after access, leaf is moved straight-up to root), dynamic Huffman and standard splay. A table of results and a graph with an example of compression are presented.

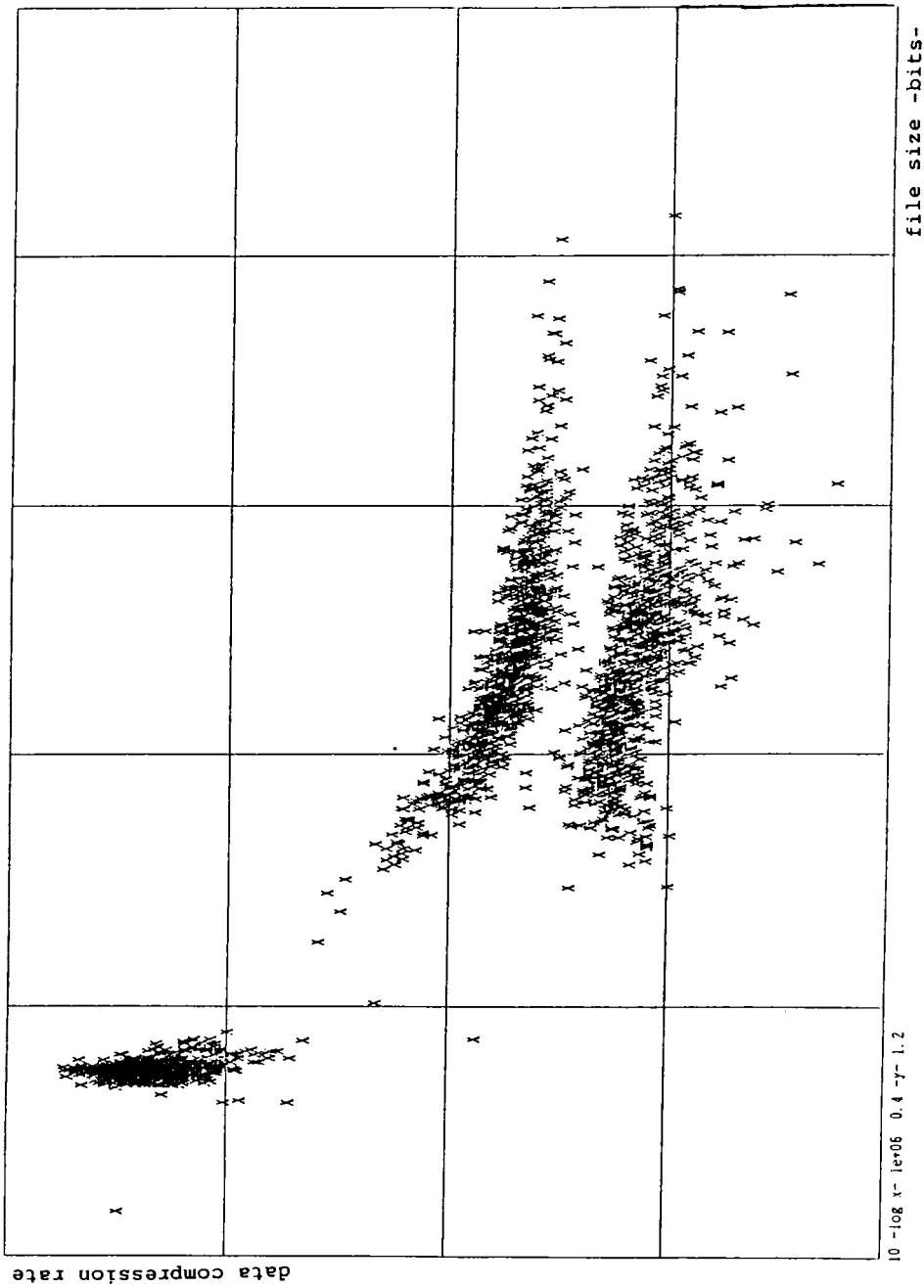


Tree structure self-adjusting operation, after access on data contents "c". Tree is automatically balanced and codeword lenght is approximately halved.

Complete Table of Results								
					Compression Ratio:%			
File	Bytes	Bits	$\mathcal{H}$	Length	Mod-Splay	MTF	Splay	Huff.
Text1	428	3424	1808	2119	61.8	64.5	-.-	62.6
Text2	4214	33928	17955	19904	58.6	59.7	-.-	55.7
Text3	16435	131480	70467	76194	57.9	58.1	-.-	54.6
Nroff1	23630	189040	95423	104965	55.5	55.9	-.-	51.9
Nroff2	49080	392640	197970	215676	54.9	55.3	70.0	51.6
C sour1	1308	10464	6453	7401	70.7	69.8	-.-	66.5
C sour2	12305	98440	62600	72464	73.6	67.1	68.1	65.0
C sour3	41808	334464	220685	243432	67.7	67.2	-.-	66.7
Object1	25383	203064	141271	131218	64.6	66.7	-.-	70.6
Object2	40960	327680	188944	184270	56.2	60.7	79.1	58.6
Object3	114688	917504	653634	629570	68.6	70.7	-.-	72.1
Image	129623	1037057	355034	258804	24.9	23.5	23.1	35.1

Nroff files are T_EXpre–format files.
Text files are english files collected at UNIX/usr/man/cat files.
C source files are normal density contents C language source files.
Object files are normal binary files and binary C compiled files.
Image file is a typical human image face raster file.

Comparison of our algorithm –Modified Splay– with Move To Front heuristics, Standard Splay and Dynamic Huffman.
Compression Ratio is defined to be the output/input bits ratio and H is the entropy classically defined.



A practical example of compression on a UNIX system, including three different input data types; Nroff pre-formated files, Unix manual files and small script files.

Upstream application:

JAPANESE COMMUNICATIONS MODEL

"Pourquoi: le Japon?
Parce que c'est le pays de l'écriture"
Roland Barthes
"L'empire des signes"

For a point of departure it could be interesting to remark that the proofs and diversified interpretations of Shannon's Capacity Theorem were all different from Shannon to Feinstein, Wolffwirts, Fano or Gallager. In Cover's various papers on multiaccess chanel (asyncronous capacity), we can read "the capacity region for the discrete memoryless multiple-access channel whithout time synchronization at the transmitters and receivers is shown to be the same as the known capacity region for an ordinary m-a channel".

Recent progress in chaotic dynamics are driving our stimulation to new exciting fields. Studies on Aloha packet-switching techniques, which shows intrinsically unstable behaviour because it is a multi-access contention protocol, it's nowadays being analysed under the name of catastrophe theory. In fact, an increase in control parameters of the m-a communication systems results in a sudden increase in packet collisions in the communications channel and hence a sharp decline in the actual throughput. When trying to utilize the channel for many burst-mode users, controling input-trafic and choosing appropriate retransmission rates, it's possible to encounter a catastrophe point which is the trigger for gross changes in the ordinary performance measures such as throughput rate and average delay. If that, stabilize a m-a channel is also a good exercise of entropy/capacity determination question.

Not a long time ago, it was created by Paul Virilio, the "Accidents Museum", proposing a more precise study of development behaviour by analising some dramatic consequences of uncontrolable situations of natural disaster, man-provoked entropy inflation situations or contextual accidents.

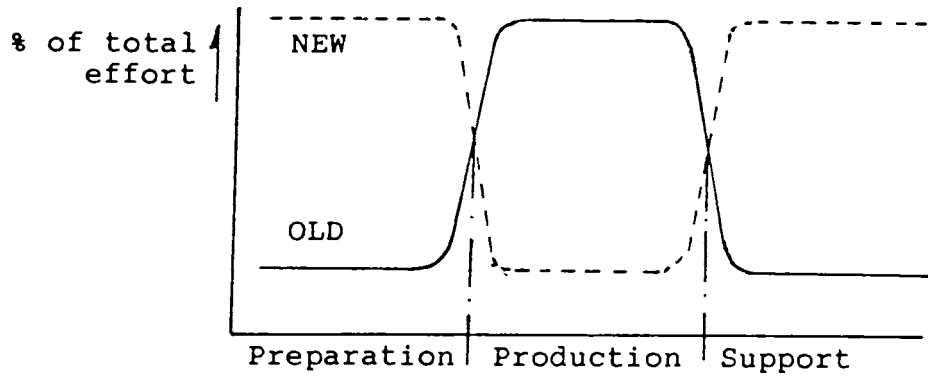
With all these modern manancial scientific tools, the author felt rather challenged to analise the so-called japanese management system by trying to identity at various levels the communications model and its axioms.

Some suspection arouse when one of our japanese colleague research fellows amazingly commented: "Boolean Algebra alphabet is very much related with the Yin/Yang Zen duality". But where to begin when we try to define a metacommunications system: not only words have meaning but everything as well. A "high context environment" needs a disciplined high-intuitive auditor crowd, suitable to a multi-data source information.

It came to our mind to do some comparative considerations about japanese Momoyama Period – the most artistic and creative times this country has ever passed through – and the Discoveries Period of late renaissance in Europe, namely the portuguese arrival to Japan (finally the "winds of God" had concede some permit...). Not only the timming overlap but also in what is concerned with the medieval times ideal profile of a noble man; he should master sword and pen. Portuguese navigators were a bit of scientists for the time and Christian priests were much knowledged and engaged into political power struggles. All this influence came to take part also in the broke-up of relations between Tea Ceremony Founder Sen-no-Rikyu and Shogun Hideyoshi. This historical episode is analised by japanese experts as a case of negative work of "ishin-denshin" (intuitive understanding, without use of words or signs). The ability of "reading-between-the-lines" is a bypath from our traditional oral/writen homogenity kingdom. Somehow, we feel

that the japanese use language as a secondary communication apparatus, a kind of tighten-up of compression parameters.

When information is compressed, the goal is to reduce redundancy, leaving only the information content (on average). This could mean to reduce orally to entropy level, emphasizing and leaving space for different unusual gestuality or some patterns of silence. In other words, leaving the oral level with an ambiguity existence threshold and stuffing that apparent emptiness with flowerished gestuality, silenced assumptions, traditional understanding or taken-for-granted rules. As a possible framework, we will use a figure introduced by Sheridan, Vamos and Aida [7], for industrial production phase versus automatization:



** Fig.: Old (pre-automation) and new (post-automation) allocations of human worker effort. (* by the courtesy of Professor Shuhei Aida, University of Electro-Communications, Chofu, Tokyo.)*

It represents that modern technologies phase of preparation such as research and design effort, and the final phase of product support such as marketing, sales, distribution and maintenance effort, has increased with factory automation.

Related with japanese suitability to modern technologies particulary robotics and neuro-systems, we return to the figure model and undertake some similar approach, for others fields of techno-social environment. Just to mention the more noticeable:

EDUCATION: where it's need 9 years of scholarity to understand a daily newspaper, but where adult will manage to read extremely low-key comics. Visual and oral components are involved in a circular effective pattern of inter-disciplinarity.

Stage-setting for DECISION-MAKING (problem-solving or even scientific discoveries) where a kind of obstinate fixation at the first preparation phase will subdivide it, into a deeper three layered block: nemawashi (consultation) / taishutekigoi (consensual efficacy) / mokuyaku (tacit contract, agreement).

ROBOTICS: Robots in Japan are seen as a metaphor for the relationship between people and technology, mainly by its accessibility and broad public acceptance. This does not mean that every single local has mastered or understood technology, but that it is perceived as something familiar, useful, practical and friendly: contemporary japanese attitudes towards robots were largely formed by comics and toys.

Postface

When evaluating complexity of a model, we realized that with some particularly arranged data was possible to compress it under values of classic entropy. Also for japanese communications model considerations, we were able to detect a situation where at least oral communication works under a particular problem-formulation.

Amazingly, we reach the understanding that the heart of complexity and chaos is simplicity itself, or all the other way around. It has to do with the nature of equations and what is known as initial conditions and boundary conditions.

With the resulting modern types of mathematical analysis, a wide variety of phenomena as telecom packet-switching, economic trends, insect population changes, natural and human behaviour forecast were undertook by scientists. But will it be fair to say that it is a modern approach to understand reality when backwards history we detect in ancient Greek form of tragedy a high-sense of universal propensity to chaos, and in other geographical coordenates about what were speaking the chinese Taoist philosophers?

Bibliography

[1] Rissanen and Langdon, *"Universal Mod. and Coding"* IEEE-IT, 1/81
[2] Kolmogorov, *"On Log. Foundations of I.T."* Prob. Inf. Transm., 7/69.
[3] Chaitin, *"Theory of Program Size F. Identical I.T."* J. ACM, 7/75.
[4] Bentley et al., *"Locally adaptive D. C. Scheme"* C. ACM, 4/86
[5] Sleator and Tarjan, *"Self-adjusting B. S. Trees"* J. ACM, 7/85.
[6] Marques, *"Adaptive data compression using self-adjusting data structures"*, M.sc thesis, Univ. Electro-Communicat. Tokyo, 2/89.
[7] Sheridan, Vamos, Aida, *"Adapting automation to man, culture and society"*, Automatica, 11/83.

Mukesh J. Patel

Centre for Cognitive Science, Univ. of Edinburgh,
Edinburgh (UK)

Relationship between Meaning and Representation: an Experimental Approach

Abstract

When presented with a simple text that describes two individuals with four properties, as well as encoding the intra-individual links between properties subjects also encode inter-individual links in their mental representation (Stenning, Patel and Levy, 1987). In order to generalise these findings, in this experiment the temporal order in which a reader is presented with the properties describing two individuals is different for each one. The results indicate that though there is no change in readers' comprehension, their representation strategy is affected by the temporal order of non-lexical higher order semantic information. Multiple regression model based on reading time data is presented together with certain recall errors patterns which show how the changes in readers' representation strategy are determined by the combined effect of the semantic information needed to encode the inter-individual links and *of the temporal order of incoming information.*

Resumo

Verifica-se que, quando confrontados a um texto simples que descreve dois indivíduos com quatro propriedades, os sujeitos tal como codificam na respectiva representação mental ligações intra-individuais entre as propriedades, codificam também ligações inter-individuais (Stenning, Patel e Levy, 1987). Com vista à generalização destas verificações, a ordem temporal como na experiência um leitor é confrontado às propriedades que descrevem os dois indivíduos é diferente num caso e noutro. Os resultados indicam que, embora a compreensão dos leitores não sofra modificações, a sua estratégia de representação é afectada pela ordem temporal da informação semântica não-lexical de ordem mais elevada. Um modelo de regressão múltipla, baseado nos dados do tempo de leitura, é apresentado juntamente com certos padrões de erros de lembrança que mostram como as mudanças de estratégia de representação dos leitores são determinadas pelo efeito combinado da informação semântica necessária à codificação de ligações inter-individuais e da ordem temporal de chegada da informação.

1. Introduction

In this paper the relationship between meaning and representation is explored. In general we address the question, how is the information in a text describing two individuals, such as "A tall, rich, Swiss vet", and "A tall, poor, Welsh chef", encoded (represented) in a mental representation? Our particular concern is with the associative

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 253/259
1992, APDC & SPF, Lisboa

intra-individual links (ie, the binding of the individual "chef" with the property "Welsh"), and inter-individual links (ie, the representation of whether the individuals are the same – both "Welsh" – or different on a particular dimension) that a reader encodes in order to preserve the information about the properties which describe each individual. How is this sort of semantic information organised in the representation? After a brief review of results from previous studies, we present results from our recent study, which indicates that the encoding of this type of information varies according to the temporal order in which it is available to the reader. It is argued that the representation of the meaning of a simple text is partly determined by non-lexical higher order semantic information.

Using simple texts such as the one in the paragraph above which describes two individuals on four property dimensions, Stenning, Shepherd and Levy (1988) present a multiple regression model based on reading time data which indicates that apart from encoding the lexical items readers also spend a significant amount of reading time (and therefore processing) on encoding various associative links between individuals and properties. For example, readers encode the fact that one individual is "a chef" and that he is "Swiss", and another is "a vet" and "Welsh", as well as the fact that on this (nationality) dimension the two properties are different (mismatched). This claim was further supported by recall error data which shows that some of the error patterns tend to occur more frequently than would be expected if the representation did not include any such associative inter-individual links. For example, readers typically make higher than expected polarity errors on a mismatched property dimension by recalling both nationalities correctly but assigning it to the wrong individuals. Hence, readers know that the nationalities of the individuals were different but fail to recall the link which binds the relevant property to the correct individual. Information regarding the links between individuals and properties will be referred to as *match structure*. In another study (Stenning, Patel and Levy, 1987) it was found that not knowing to which individual the next sentence will refer to did not significantly affect readers strategy to represent match structure. This is particularly significant because referential unpredictability renders the match structure a lot more difficult to encode.

The use of simple texts imposes a limit on the possible sorts of associative links that can be encoded in a representation. Each text has eight sentences which describe two individuals on four property dimensions. The first property dimension (introducer) is always mismatched the other three are either matched (ie, the same lexical item describe both individuals) or mismatched equally often. Texts, therefore, can be presented in eight different patterns of match structure. The constraint on the possible sorts of associative links allows us to analyse the data in a great detail and with a greater degree of confidence. It facilitates the development of models (based on the observed reading time and recall error data) that account for some of the encoding processes involved in the construction of mental representation. It also provides insights on the nature of mental representations and how it is partially determined by higher order semantic information such as match structure.

Our main concern is with the encoding and representation of match structure information, and with how its representation is affected by factors such as temporal order in which sentences within each text are presented. Though, without doubt lexical information plays a crucial role in the construction of a representation, its effect on the representation of the associative links themselves can be argued to be relatively unimportant. For example, knowing that "a vet is tall" and that "a vet is short" does not

alter the nature of the link which binds the property dimension (independent of whether it is "tall" or "short") to the individual. Results of the present study support this assumption.

This study was designed to explore the nature of the interaction between match structure of property dimension and the temporal order in which it is presented to describe the individuals. In both previous studies mentioned above, the properties describing each individual were always presented in an identical order. Hence, a reader always lernt about the nationality of both subjects before he/she learnt about the next property of both individuals.[1] The fixed temporal order of property dimension greatly facilitated the encoding of match structure. The question we wish to address is how far is the strategy of encoding match structure information dependent on a fixed property dimension order? If there is no dependency then are representations of the same text presented in different property dimension orders different, and how does temporal order effect the representation of intra and inter-individual associative links? There are number of ways in which the property dimension order of each individual was different. Hence, a reader might know that "a vet is tall" and that "a chef is rich" *before* he/she knows whether the "chef" is "tall" or "short". It is evident that this difference in property dimension order effects the availability of match structure information. Note that match structure information is *constant* in all texts: what varies is the order in which it is available for encoding.

2. Design and methodology

Texts in nine different property dimension orders (format), each in eight different match structure patterns were presented to 20 subjects. Hence, 56 texts made up one complete design. Each subject read 144 texts. The introducer (profession) was always mismatched and presented in the first two sentences of every text. Individuals were described on four property dimensions – profession, nationality, stature and temperament (apart from profession, not necessarily in that order). Properties describing each individual were randomly selected from four cohorts (one for each dimension), each containing 6 contrasted pairs of equal numbers of mono- and bi-syllabic words. The reference to individuals was predictable, in that alternate sentences referred to each individual. Text were presented on a VDU screen one sentence at a time in a self-paced reading task, followed by a menu-aided recall task. Reading times for each sentence were recorded together with the recall errors data collected at the end of each text presentation. Since lack of space limits us to only a brief outline of the experimental design and procedure, suffice to say that other relevant design variables were taken into account and controlled.

3. Results and discussion

Our interpretation of the results make two basic assumptions. First, that reading time data reflects some sort of mental processes involved in the construction of a representation of incoming information (text comprehension). Hence, longer reading times are assumed to involve either more processing *per se* or more complex processing. Second, that the pattern of recall errors is a function of both, the information represented, and, its *organisation* as determined by temporal order in which match structure information is available.

The first important result is that sentential reading time is not significantly affected by lexical items presented in any temporal position (Patel, 1989). Thus there is no significant

difference in reading time when, for example, the third sentence is "the chef is rich" or "the chef is tall". What is important, however, is the temporal position of the property dimension independent of its lexical value. Typically, for both individuals encoding the third property invariably take longer than encoding the second one, which suggests that at least some of the processing is determined by the number of possible associative links that a reader can encode. Since lexical items of the second, third and fourth property dimensions did not significantly effect reading times, the data of nine different formats was collapsed into three *format groups* which preserved the temporal order of properties describing each individual. They were as follows:

Temporal position	1	2	3	4	5	6	7	8
format group 1	I1-prop1	I2-prop1	I1-prop2	I2-prop2	I1-prop3	I2-pro3	I1-prop4	I2-prop4
format group 2	I1-prop1	I2-prop1	I1-prop2	I2-prop3	I1-prop3	I2-pro4	I1-prop4	I2-prop2
format group 3	I1-prop1	I2-prop1	I1-prop2	I2-prop4	I1-prop3	I2-pro2	I1-prop4	I2-prop3

In the above "I" refers to the individual (first or second) and "prop" to the property dimension (1 to 4) presented at that temporal position. Format groups reflect the order in which match structure information became available to the reader. Table 1 shows the mean reading time observed at each temporal position (1 to 4) for each individual.

Table 1

Mean reading times (secs.) of individuals' properties by format group (n=2880)

		Individual-1				Individual-2			
Property		1	2	3	4	1	2	3	4
Format group:	1	1.26	1.70	2.24	2.56	1.44	1.62	1.98	2.31
	2	1.23	1.67	2.24	2.56	1.34	2.05	2.48	2.71
	3	1.26	1.75	2.33	2.83	1.41	2.12	2.91	2.63

The relatively low reading times in format group 1 probably reflect the relative ease with which the match structure information can be encoded since the property dimension order is identical for both individual. However, encoding the same information albeit presented in a different temporal order (as in format groups 2 and 3) requires more reading time. This is particularly true for the second individual at temporal position 4 (2.71 secs.) and 3 (2.19 secs.) in format groups 2 and 3 respectively. Clearly, the difference in the temporal order in which match structure information becomes available to the reader has the effect of increasing the processing load at certain temporal positions. Below we consider the sorts of processes that help explain reading time differences between similar texts presented in different formats.

A general multiple regression model was developed on the basis of the observed reading time data (see Patel, 1988, for further details on regression analysis technique). The model contains a number of variables which take into account match structure information available for encoding at each sentence in each format group. At any sentence in a text in any format the information can be characterised in three ways; either the property can be encoded as matched or mismatched vis-a-vis the property of the other individual on the same dimension, or it remains unresolved (neutral) if the other individual's property is not

known. The model presented here further includes a variable which accounts for the unusual rise in reading time noted in the paragraph above. A closer look at the data shows that most of the increase in reading time occurs when readers had to encode a matched property. Since previous studies show (eg, Stenning et al, 1988) that encoding a matched property dimension typically requires less processing than a mismatched property dimension or an unresolved property we postulated that readers were probably encoding different sorts of associative links. This variable was called "unresolved–match". All four variables are cumulative.

The model[2] shown in table 2 predicts an increase in reading time as the number of mismatches (eg, mis1), matches (eg mat1) and unresolved (eg neut1) increase. It also predicts a high coefficient for unresolved–matched properties (0.64 secs.). The combined predicted coefficients of the variables account for 95.3% of the non–error variance in the reading time data.

Table 2

Summary of regression model predicting reading time from match structure

Variable	Coeff(secs.)	Standard Error
Intercept	0.97	.04
NEUT1	0.28	.03
NEUT2	0.74	.03
NEUT3	0.96	.06
MIS1	0.40	.04
MIS2	1.00	.04
MIS3	1.35	.05
MIS4	1.51	.10
MAT1	0.21	.03
MAT2	0.30	.04
UNRES–MATCH	0.64	.04

In general the model supports the view that wherever possible readers' representation strategy utilise match structure information. However, the inclusion of unresolved–match also predicts that at certain temporal positions readers seemingly fail to encode the inter–individual link that a property is matched on that dimension. In which case what sort of link(s), if any, does the reader encode? We will argue that when a reader finds it difficult to encode match structure information he/she makes more intra–individual links. Analysis of recall errors tends to support this position.

Recall error patterns in this study show a degree of consistent differences between format groups which can be explained in terms of when (ie, at which temporal position) in the text match structure information becomes available. We have already mentioned how error patterns suggest that match structure is encoded in a representation of simple texts. At a very simple level errors can be classified in three basic categories; *single, multiple* and *joint*. During recall a reader can either make an error on one individual (single) or more than one error on the same individual (multiple) or one error each on the same property dimension of both individuals (joint). If a reader has encoded the match structure information (that a particular property is matched in this case) then the likelyhood of

making joint errors will be high. Alternatively if the reader tend to encode the property by making more intra-individual links then the likelyhood of multiple errors involving that property would be high. In order to illustrate this point we will confine ourselves to reporting the error patterns observed for texts presented in format group 2.

Our reading time regression model predicts that at sentence 8 when the fourth property of the second individual matches with the second presented property of the first individual readers tend not encode the inter-individual link. The significantly low frequency (see table 3) of joint errors on prop2 (0.2%) as compared to prop3 (1.5%) and prop4 (2.1%) supports this prediction (X^2=26.1, df=2, $p < 0.001$). However, multiple errors involving prop2 (2.9%) of the second individual (presented last) are not significantly lower than those involving prop3 and prop4 (2.9% and 3.3% respectively) which strongly suggests that readers are encoding a lot more intra-individual links; they seem to bind the fourth property of the second individual not only with the introducer (eg, that "the chef is tall") but also with the second and the third properties (eg, that "chef is tall and Swiss", etc). Similar sorts of error patterns were also observed at appropriate temporal positions in format group 3 (Patel, 1989).

Table 3

Percentage recall errors on each property presented in format group 2

		prop1	prop2	prop3	prop4
Joint errors (matched)	Both individual	-	0.2	1.5	2.1
Multiple error	Individual 1	1.4	0.7	2.2	1.9
	Individual 2	1.9	2.9	2.9	3.3

4. Concluding remarks

Overall, our results provide further evidence in support of the methodology developed to study the role of higher order semantic information during encoding processes involved in the construction of a representation. This study in particular shows that this type of information is crucially dependent on the order in which it becomes available for encoding. It needs to be stressed that the amount and the sort of match structure information in a typical text is invariant. A reader's comprehension at the end of the same text, as far as the fact that it describes two individuals with four properties each is concerned, presented in all three format groups is going to be the same. And yet our results show that the representation of the same text differs in terms of intra- and inter-individual associative links according to format (or according to the temporal order in which each individual is described). Depending on format, therefore, the same text obtains different mental representations. The interesting question then is when the representation proves to be faulty or inadequate (ie, when readers fail to recall both individuals correctly) is the reader's *comprehension* (or lack of it) determined by the temporal order (format) in which the information was encoded? Results from the present study seem to suggest something of that nature. In which case the meaning of the text – as opposed to the meaning of individual lexical items abstracted away from context – can be said to be crucially dependent on sentential temporal order. However, the present study lacks the required detailed

observations needed to settle the issue regarding the possible connections between meaning and temporal order. In order to do so will require a great deal of further consideration and study. Our study shows that any account of the relationship between meaning and representation will need to account for the effect of both, higher order semantic information (such as match structure) *and*, the temporal order in which it is presented.

Notes

(1) In some respects the Stenning, Patel & Levy (1987) study is slightly different but this does not render out general point invalid.

(2) Though in defining the variables it was assumed that each sentence contains a constant amount of information, given that the model is data driven, it was not predicted that their contribution to reading times would be a linear function. The underlying assumption being that encoding processes are not the same each time but of the same sort for the same type of match structure information (eg, mismatch). Hence, during the last stages of multiple regression modelling each level of these variables was replaced with a "dummy" variable where, for example, "mis1" and "mis2" refer to mismatch value of one and two respectively.

References

Stenning, K., Shepherd, M. and Levy, J. (1988) – *On the Construction of Representations for Individuals from Description in Texts*. Language and Cognitive Processes, 3(2) 129–164.

Stenning, K., Patel, M.J. and Levy, J. (1987) – *The 'Binding Problem' in human memory: some effects of referential discontinuity on the construction of representation for individuals*. Research Paper No. 18, Centre for Cognitive Science, University of Edinburgh.

Patel, M.J. (1989) – *The effect of temporal order of text presentation on representation*, Working Paper, Centre for Cognitive Science, University of Edinburgh.

Jean–Claude Perez

Centre Européen d'Intelligence Artificielle, I.B.M. France
Montpellier (France)

Peut–il y avoir Mémoire sans Représentation?*

Mots–clefs:
Réseaux neuronaux,
Modèle neuronal "Chaos Fractal",
Mémoires holographiques,
Auto–apprentissage,
Représentation des connaissances.

Termos–chave:
Redes neuronais,
Modelo neuronal "Caos Fractal",
Memórias holográficas,
Auto–aprendizagem,
Representação dos conhecimentos.

* Ce texte est extrait du livre de l'auteur: *"Une troisième voie vers l'intelligence artificielle: entre hologrames et chaos"* (à paraître 1991, Ed. Masson, Paris).

PEUT–IL Y AVOIR MÉMOIRE SANS REPRÉSENTATION? Poser une telle question peut paraître, à priori, absurde. L'ambition principale de l'informatique ne consiste t–elle pas à représenter ''de mieux en mieux'' l'information?

Curieusement, je me suis posé, sérieusement, cette question lorsque j'ai réfléchi aux problèmes de sécurité et de ''piratage'' informatiques.

En particulier, j'identifiai une application nouvelle et fort intéressante du ''détecteur de nouveauté CHAOS FRACTAL'': la mémorisation holographique des ''mots de passe''.

Quel est le problème? L'une des principales sources de piratage informatique résulte du fait que, quelque part sur les disques systèmes, dans les zones affectées à l'administration du système, il existe des tables dans lesquelles sont stockés, en clair, TOUS les mots de passe relatifs à chacun des utilisateurs.

Quiconque parvient à pénétrer dans cette zone privilégiée et sensible, détient alors tous les mots de passe qui lui permettront ensuite d'accéder à l'intégralité des applications et utilisations du système.

Une idée naturelle mais très couteuse en ressources et en performances consiste à ''encrypter'' ces codes. L'encryptage revient à transformer un code lors des communications ou manipulations puis à le restituer, sous sa forme originelle, lors de l'utilisation.

La forme la plus sophistiquée d'encryptage se nomme ''encryptage à clef révélée'' (ref Hellman–1981). Cette technique est inspirée de la théorie des nombres (calcul des congruences). L'utilisateur, lors du décodage, doit combiner deux clefs: l'une est secrète,

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 260/270
1992, APDC & SPF, Lisboa

l'autre est publique (d'où le terme de "clef révélée"). La propriété majeure est la suivante: si la vérification d'un code est rapide, au contraire le retour arrière consistant à décrypter l'information codée pour reconstituer le code originel sans disposer de la clef secrète, est un problème de la classe NP-complet. Cela signifie que le piratage d'un tel code par le calcul est une opération quasiment impossible. En ce sens, le codage est pratiquement irréversible si on ne dispose pas de la clef secrète.

Un exemple d'un tel fonctionnement serait possible en utilisant une seule fractale du modèle "CHAOS FRACTAL":

$$X(N+1) = 1 - MU.X(N)**2$$

Le message en clair (transmis) serait une valeur quelconque de XO telle que son image par la fractale conduise, après un nombre "N" d'itérations récursives à déterminer, au message X(N) à encrypter.

Le message encrypté X(N) serait la valeur de l'attracteur après "N" itérations récursives.

La clef révélée serait le nombre d'itérations "N" à exécuter.

La clef secrète serait la valeur de MU (nombre réel compris entre zéro et deux).

La fonction chaotique étant totalement déterministe et reproductible, l'application de "N" opérations récursives sur XO et pour un point de fonctionnement MU, reconduirait au message en clair X(N).

Au contraire, dans le type de mémoire holographique que nous allons proposer, le processus deviendra totalement IRREVERSIBLE car le passage d'une seule fractale à un réseau de fractales inter-connectées fait perdre toute réversibilité (l'opération de communication entre neurones fractals est totalement déterministe mais irréversible).

En fait, bien que ce système assure une extrême sécurité informatique, on ne peut plus, paradoxalment, parler ici d'encryptage car, dans ce type de mémoire, il n'y a plus de représentation de l'information (!), et donc plus d'encryptage de l'information...

Voici une description du processus:

-Imaginez une mémoire holographique numérique matérialisée par un tableau de nombres réels MU (ref Perez-1989 chapitre 10).

-Le contenu de cette mémoire étant accessible à quiconque, on peut donc parler de "mémoire holographique révélée" (par analogie avec les "codes à clef révélée").

-Considérant cet hologramme comme une sorte de "boîte noire", nous la nommerons H (paradoxe de la représentation: bien que révélée, la boîte ne contient aucune information, elle constitue plutôt un opérateur, une fonction). On peut parler de "boîte noire" dans la mesure où on ne s'intéresse qu'au comportement de cette boîte face à des stimulis externes. Jamais on ne s'interessera au contenu de la boîte...

-Soumettons H à un code à apprendre (stimuli correspondant à un nouveau mot de passe). Appellons ce code C1.

-La boîte noire H va réagir au stimuli C1 en trouvant un nouvel équilibre du réseau fractal qui se concrétisera par une nouvelle configuration de l'hologramme H. Nous la nommerons Ha(C1) pour "hologramme issu de l'apprentissage de C1".

-On a maintenant le comportement BINAIRE suivant:

=>Si on soumet Ha à tout mot de passe erronné Ci

tel que Ci différent de C1,

alors:

Ha évolue vers un autre hologramme Hs (H stimuli)

tel que Hs (Ci) différent de Ha(C1).

C'est-à-dire que l'hologramme cherche à évoluer en s'adaptant à Ci.

Si le réseau était en mode apprentissage, nous le laisserions s'adapter, ce nouvel hologramme serait maintenant la mémoire à la fois de C1 et de Ci, soit H(C1, Ci).

Comme le réseau est en mode reconnaissance, nous ne laissons pas évoluer l'hologramme, qui demeura donc sur sa valeur Ha(C1).

=>Si, au contraire, nous soumettons Ha au code appris C1,

alors:

le mot de passe valide C1 sera reconnu, et cela se traduira par une non variation de l'hologramme:

Hs(C1) = Ha(C1).

Nous venons donc de bâtir une boîte noire capable de répondre à des stimulis "mots de passe" suivant deux comportements DIFFERENCIÉS, selon qu'elle soit soumise à un mot de passe appris ou bien à un mot de passe erronné.

Ce qui vient d'être dit est généralisable à un ensemble de mots de passe ou à une séquence dynamique de mots de passe pouvant comporter plusieurs milliers d'éléments.

En ce sens, on peut se demander où mettre la boîte noire:

-Dans l'ordinateur: la boite noire se comporterait alors comme une sorte de serrure pouvant être ouverte par plusieurs clefs, par exemple par la clef de l'utilisateur, mais aussi par la clef de l'administrateur du système. Mais ce second accès est-il nécessaire? Oui, pour des raisons de "dépannage", non, pour des raisons de sécurité, car une telle possibilité augmenterait les possibilités de piratage du système (clef "passe-partout").

-Hors de l'ordinateur: la boîte noire constituerait alors une sorte de "serrure externe" qui "viendrait se faire ouvrir par une clef fixe (et inconnue) située dans l'ordinateur à accéder". Cette seconde forme est fort originale, mais elle implique que la phase d'apprentissage d'un code se fasse sous le contrôle de l'ordinateur à accéder, sinon, n'importe quelle serrure pourrait apprendre n'importe quel code par simple contact. Ce mode de fonctionnement rappelle quelque peu celui des lymphocites de la biologie: chaque lymphocite est une sorte de serrure itinérante dans laquelle peuvent venir se fixer diverses clefs (virus ou autres lymphocites), ouvrant alors un processus dynamique de communication entre lymphocites...

Qu'est-ce qui est nouveau dans ce système de sécurité holographique? Il y a un double fait remarquable:

1/L'hologramme H est "révélé et public".

2/Nulle part, dans H, on n'a "stocké" une quelconque trace de l'apprentissage de code C1.

Cette nuance est fort abstraite: on ne peut pas dire qu'on ait REPRESENTÉ le mot de passe C1 dans la boite noire holographique H... puisque cette représentation n'existe pas!

Par contre, nous avons bâti un processus dynamique et actif qui a assimilé, qui s'est adapté au code stimuli C1, et qui, plus tard, sera capable de RECONNAÎTRE ce code au sens du "rappel" et du "déjà-vu".

La question qui maintenant va se poser est la suivante: quelle part représente ce type de mémoire vis-à-vis du phénomène général de mémorisation? Est-ce là une sous-propriété secondaire et marginale, ou bien le "rappel" et le "déjà-vu" constituent-ils une propriété fondamentale et peut-être suffisante capable d'engendrer l'intégralité des processus de mémoire?

L'important avec ce détecteur de nouveauté, c'est que "l'information n'est nulle part représentée"!

De ce fait, le problème de "piratage" de tels systèmes devient caducque et sans object!

Alors, si ce mode de mémoire s'avérait suffisant, nous démontrerions que LA

MÉMOIRE PEUT "SE DISPENSER" DE TOUTE REPRÉSENTATION!

En d'autres termes, un système peut mémoriser le monde extérieur sans le représenter symboliquement, mais, simplement, par une relation d'activité dynamique, la reproduction des mêmes stimulis engendrant un comportement invariant (on ne peut plus dire, comme on le dirait pour un système à représentation de l'information, "les mêmes causes produisant les mêmes effets", ou "les mêmes informations d'entrée produisant les mêmes résultats de sortie"...). En fait, un tel système est capable de différencier la routine vis-à-vis de la nouveauté, il y parvient par une différence dans sa dynamique interne.

Afin de vous aider à apprécier le "saut quantique" que permettrait un tel système dans le domaine du "piratage informatique, voici une analogie:

Imaginez deux personnages qui apprennent, se comportent, et observent, selon la même dynamique. Ils apparaissent donc identiques à tout observateur extérieur. En quelque sorte, il s'agit là de deux "clones" humains. Cependant, il existe une différence entre ces deux humains: le premier peut parler, expliquer et justifier ses comportements, tandis que le second reste muet et inapte à expliquer, décrire ou justifier ses comportements. Pourtant, tous deux communiquent avec le monde extérieur dans la mesure où ils s'adaptent et réagissent aux stimulis et perturbations de ce monde extérieur. La différence est que le second d'entre eux ne peut être étudié et analysé qu'en termes de "boîte noire".

Supposons que, lors d'un interrogatoire policier, on interroge ces deux humains sur "ce qu'ils ont vu".

-Sous la contrainte, le premier tentera d'expliquer, de justifier, il communiquera son "modèle", sa "version" des évènements.

-Au contraire, le second ne sera capable que de réagir à des stimulis, par exemple, il "reconnaîtra" un accusé présumé, il reconnaîtra des lieux, des indices, mais jamais il ne pourra "expliquer", communiquer un scénario possible. En fait, contrairement au premier, ce second personnage n'aura bâti aucune représentation interne de l'information, ou, plus exactement, s'il en a bâti une, jamais il ne sera capable de la transmettre et de la justifier, car elle n'est pas transmissible, car elle n'est pas analysable.

Aujourd'hui déjà, on peut ressentir ce problème lorsqu'on compare les comportements respectifs des systèmes experts d'une part, et des réseaux neuronaux d'autre part: les premiers sont capables d'expliquer pas-à-pas leurs inférences et déductions tandis que les seconds "trouvent" de manière globale mais sont incapables de justifier et d'expliquer leurs décisions.

Au travers de cette "parabole du muet", on comprendra bien quelle serait l'avancée, sur le plan de la sécurité informatique, si de tels systèmes voyaient le jour, ce qui est dès aujourd'hui possible.

Cependant, ce qui est efficace pour les mots de passe l'est-il aussi pour toute autre information? Non, puisqu'il n'y a plus réversibilité de l'information! Le "candide" avancera les arguments suivants: "Vous codez les mots de passe mais pas la totalité de l'information utile... Pourriez-vous, avec votre mémoire holographique, encrypter l'intégralité de l'information... C'est-à-dire... Supprimer toute information du système! Soit encore, créer des systèmes actifs mais muets!

Quel paradoxe: ce système serait capable de produire des comportements relativement "intelligents" alors que toute représentation interne de l'information resterait à tout jamais inaccessible... Et pourtant, tant l'information interne que les opérateurs internes resteraient "publics" et "consultables"...

En fait, ma phrase est imprécise, je la reformule:

"Il deviendrait impossible de bâtir un "pont", une relation, entre la représentation du

monde externe et la dynamique des informations et opérations internes au réseau holographique. Ce pont existe mais sa "non-linéarité" et son irréversibilité le rendent à tout jamais inaccessible à un quelconque observateur extérieur! Ce système, bien qu'il puisse communiquer avec le monde extérieur, ne pourra jamais communiquer sa représentation interne des connaissances."

Reconnaissance de l'information et restitution de l'information:

En fait, le problème de la mémorisation est généralement décomposé par les psychologues en deux sous-ensembles (ref Morton-1985):

-Le "rappel" ou "reconnaissance" qui consiste à attribuer la qualité de "déjà-vu" à un stimuli.

-La "restitution" d'une information qui consiste, pour certaines conditions de contexte et de stimulis, à "renvoyer" vers le monde extérieur une information "antérieurement perçue".

On peut classer les types de mémorisations en trois catégories, chacune sera ensuite analysée en fonction de ces deux critères de mémorisation; nous distinguerons:

-La mémorisation triviale, qui consiste à stoker l'information "en clair" ou d'une manière "faiblement codée", ce sera le cas des codes usuellement utilisés dans les ordinateurs (ASCII, EBCDIC, etc.).

-La mémorisation encryptée, ce type de transcodage est plus complexe mais il demeure REVERSIBLE. On peut donc, à tout moment, restituer l'information en clair.

Les deux mémorisations ci-dessus sont du type "représentationnelles", il existe toujours une relation, une correspondance biunivoques entre le monde extérieur et la représentation symbolique interne. C'est le cas des informations stockées sur disques, des bases de données relationnelles, des systèmes experts, et, plus généralement, de tous les systèmes "à bases de connaissances".

-Au contraire, le troisième type de mémorisation est constitué par la classe des systèmes, naturels ou artificiels, du type de notre mémoire holographique. Là, l'information n'est plus représentée, et ces systèmes sont irréversibles: l'encryptage, si l'on peut parler d'encryptage, est incapable de restituer l'information stimuli, par contre, il saura la reconnaître, si elle ré-apparaît.

Tout cela se résume par le tableau suivant:

Type de mémoire:	Reconnaissance	Restitution
Stockage en clair	oui	oui
Encryptage	oui	oui
Holographique	oui	???

Dans la dernière case, en bas à droite, nous avons remplacé le "non" par des points d'interrogation, ce qui contredit ce qui vient d'être écrit plus haut.

Nous allons, en effet, tenter de répondre à cette question:

"Comment un système ou organisme, seulement capable de reconnaître l'information déjà vue, pourrait-il restituer, plus tard, cette information sans JAMAIS l'avoir stockée sous une forme explicite?".

En d'autres termes, un système peut-il restituer une information sans la stocker (je sous-entends: "sans la stocker sous forme d'information, ce qui n'interdit pas de la stocker

sous d'autres formes plus opératoires telles qu'activité, comportement dynamique, eyc...'').

En fait, nous touchons là au coeur du paradigme de mémorisation, je citerai un passage de *''L'homme neuronal''* (ref Changeux 1983):

Changeux affirme là son désaccord avec Bergson, je cite:

> ''Le cerveau, machine à penser? Bergson, dans ''Matière et Mémoire'', écrivait que le système nerveux n'a rien d'un appareil qui servirait à fabriquer ou même à préparer des représentations''.
>
> ''La thèse (je cite Changeux) développée dans ce chapitre est l'exact contre-pied de celle de Bergson. L'encéphale de l'homme que l'on sait contenir, dans l'organisation anatomique de son cortex, des représentations du monde qui l'entoure, est aussi capable d'en construire et de les utiliser dans ses calculs.

D'après J.P. Changeux, *''L'homme neuronal''*.

Ainsi, pour Bergson, il n'y a pas de représentations dans le cerveau, tandis que Changeux affirme exactement le contraire.

Je pense qu'ils ont tous deux raison, en fait l'ambiguité vient du sens attribué au mot représentations: en commençant par le language, en continuant par l'écriture, en finissant par l'informatique, l'homme a toujours construit ses progrès sur une représentation symbolique du monde extérieur. Cette représentation est plus INFORMATIONNELLE qu'OPERATIONNELLE. C'est-à-dire, cette représentation s'articule toujours autour du concept statique d'information et rarement autour du concept, plus dynamique, d'opération. Même un programme informatique, qui pourtant est un opérateur, est représenté, se stocke et se transmet sous forme d'information. On ''photographie'', on codifie, on représente, les ÉTATS statiques, plutôt que les COMPORTEMENTS dynamiques.

Dans ces types de représentations, le temps est presque toujours absent. Je rejoindrai là Prigogine qui, dans *''Entre le temps et l'éternité''* (ref Prigogine 1988) montre comment l'homme a ignoré l'irréversibilité du temps dans tous les modèles de la Physique et de l'univers qu'il ait bâtis: ou bien il ignorait complètement le temps en se limitant à des représentations causales ou probabilistes, ou bien il ne considérait le temps qu'élevé ''au carré'', ce qui ''rendait le temps réversible'', c'est le cas des équations de la relativité générale d'Einstein.

Toute cette culture axée sur la modélisation et les ''sauts de cause en effet'', d'état, en état, nous laisse aujourd'hui croire que l'important est l'état, la stabilité, l'information statique, plutôt que la situation, le contexte, la dynamique. Tout cela a conduit à ''linéariser'' le monde et à écarter toute non-linéarité, en la considérant comme une singularité, comme un cas d'exception. Tout comme nous nous sommes fait croire que le monde était essentiellement linéaire, nous croyons aujourd'hui que l'essentiel est ''l'information descriptive d'un état''.

Nous avons ignoré le temps, la dynamique et le caractère essentiallement non linéaire de ce qui nous entoure. Voici le processus:

Pour décrire l'univers et le monde qui nous entourent, nous devons en rechercher les stabilités, car la dynamique et le non-linéaire échappent à toute description et à toute représentation. Ainsi, comme l'esquisse Prigogine (ref Prigogine, 1988), en ''poursuivant'' toujours cette stabilité, l'homme a imaginé des modèles successifs de l'univers:

> ''Les Grecs tenaient pour évident que seule la perfection du cercle pouvait donner son sens à celle des mouvements célestes.
>
> C'est donc en fonction de l'idéal circulaire qu'il jugèrent les mouvements qu'ils observaient.
>
> Dès lors, depuis Ptolemee jusqu'à Copernic, l'astronomie se voua à ''sauver les phénomènes'', à construire de manière ad hoc les trajectoires complexes, de cercles

> et d'épicycles, qui permettraient de rendre compatibles l'idéal et l'observation. Mais la Physique depuis Galilee jusqu'à aujourd'hui, a elle aussi été guidée par un idéal de perfection, celui que Leibnitz baptisa du nom de "principe de raison suffisante". À la manière du cercle grec, l'égalité réversible entre cause et effet a constitué une contraine quasi-invisible, fil conducteur qui mène de la physique galiléenne à la mécanique quantique et à la relativité, comme le cercle mène de Ptolemee à Copernic."...

...Puis...

> "Kepler renouvela l'idéal d'intelligibilité de l'astronomie: il osa "briser le cercle" et créer les mathématiques qui permettent de comprendre les phénomènes que l'astronomie traditionnelle se bornait à sauver. Briser le cercle de la raison suffisante, créer un language mathématique qui rende intelligibles les processus et les évènements irréversibles que la physique traditionnelle se borne à "sauver" par des approximations phénoménologiques, c'est finalement à ce défi que son intérêt pour la question du temps a conduit l'un d'entre nous." (sous-entendu, Prigogine).

Extrait de *"Entre le temps et l'éternité"*
(ref Prigogine 1988)

Ainsi, associé au piège de la représentation, il y a le piège de la stabilité.

L'exigence de stabilité de la représentation par l'information nous a fait oublier l'existence du mouvement, du temps, et des comportements dynamiques. On a déformé le sens du terme "représentation", il y a la place pour d'autres modes de représentations que la "représentation par l'information explicite", c'est en ce sens que tant Changeux que Bergson ont raison car ils n'attribuent pas le même sens au mot "représentation": pour Bergson, il s'agit plutôt du sens "informationnel" de la représentation, au contraire, pour Changeux, "représentation" signifie une sorte d'activité dynamique, ce n'est donc plus de l'information... La représentation, au sens de Bergson, serai accessible à un observateur extérieur, au contraire, la représentation, au sens de Changeux, serait inaccessible à tout observateur extérieur. Tous deux tomberaient alors d'accord, puisque Bergson affirme que "sa" représentation ne peut exister dans le cerveau, tandis que Changeux rétorque que "sa" représentation existe dans le cerveau...

Essayons maintenant de répondre à notre question centrale:

"Ce nouveau type de représentation "comportementale" permet-il de produire une "restitution" de l'information vers le monde extérieur?

Si non, ce type de mémoire est incomplet car il ne couvre pas l'une des deux opérations fondamentales de la mémoire.

Si oui, ce type de mémoire est complet puisqu'il assure les deux fonctions fondamentales de la mémoire: le rappel et la restitution.

Venons-en au problème de la restitution d'information. Sans avoir l'ambition d'en expliquer les mécanismes, nous essaierons plutôt d'esquisser ce que pourrait être une "représentation sans information", nous soulèverons en particulier les questions suivantes:

-Quelle est la relation entre image observée et image produite?

-Quels seraient les mécanismes d'une représentation dynamique?

-Serait-il possible de ne pas stocker les représentations?

A l'issue de cette discussion, peut-être commencerons-nous à entrevoir qu'il peut exister un mode de représentation qui ne soit pas de l'information...

Sur la relation entre image perçue et image produite...

La question dont nous discuterons ici concerne la nature du "pont", de la "relation", qui existent entre l'image perçue du monde extérieur et l'image produite vers le monde extérieur. Le terme "image" prend ici un sens général, on pourrait également utiliser des termes plus généraux, qui ne soient plus restreints à l'univers visuel: par exemple, le terme anglo-saxon "pattern", on peut aussi utiliser les termes généraux objet, concept, etc...

Le but que nous poursuivons est de réduire le mode de représentation par de l'information, jusqu'à, peut-être, la rendre... inexistente.

Donc, dans ce couplage entre objet perçu et objet produit, je crois pouvoir distinguer trois cas:

-Le couplage "informationnel".
-Le couplage "aléatoire".
-Le couplage "auto-référentiel".

Voyons-les tour à tour.

1/Le couplage "informationnel":

Il existe ici un lien formel et combinatoire entre le concept perçu et le concept produit. Dans tout système informatique, ce lien, nous pourrions dire "liant", est l'information: c'est le même atome d'information (par exemple, le "nom d'un client"), qui sera perçu (saisie puis enregistrement), pour être, plus tard, reproduit, tel quel (sur une "facture", par exemple).

L'unique évolution qui ait eu lieu à ce niveau-là, c'est que la relation entre perception et production est devenue de plus en plus combinatoire, passant d'une "restitution" brutale de l'information stockée à une restitution de plus en plus décisionnelle, résultant d'une composition, d'une complexification, d'un certain assemblage: ce fut d'abord la notion de programme, puis, aujourd'hui, la notion de système expert. La connaissance devenant de plus en plus "atomique" et "en miettes d'informations", s'est réduite et parcellisée au niveau de la représentation, cela afin de parvenir à plus de flexibilité au niveau des lois d'assemblage lors de la restitution. Ainsi, bien que la complexification et la combinatoire augmentent sans cesse, ces machines restent des systèmes à "base de connaissances" et à "traitement de l'information".

Malgré cette "explosion combinatoire des comportements possibles", le plus petit atome de représentation reste, toujours, une information, et, toujours, on peut reconstituer le chemin qui relie l'information perçue à l'information produite. Ce chemin est d'ailleurs réversible dès qu'on conserve la trace des étapes successives (exemple d'un système expert capable de "justifier" et "d'expliquer", pas-à-pas, ses déductions successives).

2/Le couplage "aléatoire"

Tout à fait aux antipodes du couplage informationnel, on peut imaginer "qu'il n'existe aucun lien" entre le concept perçu et le concept produit, cela bien que ces deux concepts soient correllés!

Prenons un exemple: à la fin du livre Perez 1989, je montre l'image "d'un clown" qui est produite par des applications successives de l'automate cellulaire de CONWAY ("jeu de la vie") sur une population initiale formant une sorte de "U", et se réduisant à 9 cellules seulement (3 x 3 cellules voisines). Ce qui est curieux, c'est que, jamais, il n'y a

eu une quelconque volonté de "dessiner un clown"... Pourtant, cette figure présentée, indépendamment, à des dizaines d'enfants, conduit toujours au même verdict: tous y reconnaissent une tête de clown. Ce n'est là qu'un accident du hasard, une sorte de "mirage"...

On peut alors imaginer le scénario suivant:

-Des circuits de mémoire sont sensibles (au sens du rappel) à la perception de l'image, externe, d'une tête de clown.

-A partir d'une sorte de "germe" (petit ensemble d'états simultanés de quelques neurones, ce terme, proposé par R. Thom est suggéré par J.P. Changeux au niveau de l'amorce d'activités cérébrales), ici, le germe est la configuration "U" initiale, il va se produire "une sorte d'embrasement cellulaire" conduisant à la production d'une image interne "clown".

-Cette image interne "clown" étant fortement correllée avec l'ancienne perception externe d'un clown, il y a "rappel".

-Cette correllation accidentelle entre perception externe et production d'image interne peut être assimilée à une sorte de "résonance". C'est du reste ce que suggère J.P. Changeux (Changeux 1983):

> "Ces objets mentaux particuliers, ces ébauches ou pré-représentations, existent avant l'interaction avec le monde extérieur. Elles résultent de la recombinaison de groupes pré-cablés de neurones ou d'assemblées de neurones et de ce fait leur diversité est très grande..."

puis...

> "L'épreuve de la réalité consiste en la comparaison d'un concept ou d'une image avec un percept. Le test pourra être réalisé par l'entrée "en résonance", ou au contraire "en dissonance", de deux assemblées de neurones confrontées".

Extrait de *"L'homme neuronal"*, J.P. Changeux

-S'il y a "intention" de communiquer vers l'extérieur, ne serait-ce que pour que le système montre qu'il "sait" redessiner "ce" qu'il perçoit, le système est capable de commander des organes moteurs par ce "pattern" clown qu'il sait produire. Il n'est pas question, ici, de discuter de cette "intentionalité", pas plus que du passage de la production d'une image interne à la production, matérielle, de cette image, en "dessinant", par exemple.

Ce type "d'intelligence accidentelle" est fort élégant, mais il pose immédiatement les deux questions suivantes:

a) La diversité potentielle des "accidents" entre images du monde réel et images internes produites par hasard est-elle assez grande pour "couvrir" toutes les images perceptibles du monde réel?

b)Bien que le "clown" interne ne soit pas mémorisé, puisque issu d'un processus dynamique, il faut bien qu'il ait eu "représentation" de l'image initiale, du "germe U"... On réduirait alors la représentation sans toutefois la faire disparaitre...

Nous allons tenter de répondre à ces deux questions en proposant le troisième type de couplage: le couplage auto-référentiel. Ce type de couplage constituera, en fait, la suite naturelle de ce qui vient d'être spéculé ci-dessus.

3/Le couplage "auto-référentiel":

Paradoxalment, les réponses aux deux questions ci-dessus sont des spéculations autour de la notion d'infini:

a) Une réponse possible à la première question est la notion d'infini des possibles: de même qu'un attracteur trouvera toujours, parmi l'infini de ses états potentiellement possibles, la bonne réponse à une perturbation imprévue, on peut penser de même que:

Un système à très grande diversité sera capable de produire une infinité d'images "analogues" à l'infinité des images qu'il pourrait percevoir du monde extérieur! C'est là, l'un des paradoxes de l'infini!

Afin de nous en persuader, considérons les potentiels respectifs de créativité dans chacun des trois cas suivants:

-Un automate cellulaire de CONWAY de 100x100 cellules.

-Une population de neurones holographiques CHAOS FRACTAL.

-La population des neurones du cerveau.

-La population des neurones de CONWAY: L'automate cellulaire de CONWAY comportant 100x100 cellules est déjà capable de produire, à partir d'un "U" initial, une bonne dizaine de figures rappelant des figures du monde extérieur (clown, coeur, etc.). On peut imaginer le potentiel de créativité engendré par les milliers d'autres figures "germes" initiales possibles. Pourtant, chaque neurone est ici fort "rustre", il ne possède que 8 synapses (8 voisins) et il ne travaille que sur des signaux binaires (0/1), tant au niveau de ses états d'entrées (synapses) qu'au niveau de ses états de sortie (états de l'automate).

-La population des neurones holographiques CHAOS FRACTAL: Considérons maintenant un réseau, tel que celui qui sera discuté à la fin du livre, qui serait composé de 100x100 neurones holographiques CHAOS FRACTAL: ici, chaque neurone CHAOS FRACTAL est un petit réseau holographique de 3x3 fractales (tel que décrit dans Perez 1989 chapitre 10). Ainsi, chaque neurone holographique comporte 9 synapses analogiques et un état de sortie 0/1 de type "rappel" ou "déjà-vu". Ce neurone est ainsi capable de reconnaitre, au sens du rappel, 10000 configurations différentes du groupe 3x3 synapses. Un seul de ces neurones suffirait pour résoudre le problème des mots de passe vu précédamment ou pour vérifier la validité d'un code client parmi 10000... On manque alors d'imagination pour prendre conscience du potentiel de créativité d'un réseau qui serait composé de 100x100 neurones holographiques. Tout ce qu'on peut dire, c'est que son potentiel de créativité serait infiniment supérieur à celui du "jeu de la vie créateur de clowns accidentels"...

-La population des neurones du cerveau humain: Que dire alors d'un réseau qui comprendrai entre 10 milliards et 100 milliards de neurones, chaque neurone recevant des entrées analogiques issus de 1000 à 10000 autres neurones?

Il est peut-être possible qu'un réseau aussi diversifié soit capable de produire, spontanément, et "accidentellement", toutes les "représentations internes" analogues (ou assez voisines) aux milliards d'images qu'il pourra percevoir du monde extérieur.

C'est là "un paradoxe de l'infiniment diversifié".

b) Venons-en maintenant à la seconde question: un tel système doit, à un certain moment, en un certain endroit, posséder des représentations minimales, celles des "germes d'images" telles que ce petit "U". Je crois qu'on peut résoudre ce problème par un second paradoxe de l'infini: tout comme ce "germe U" avait émergé du hasard (ref Perez, 1989), ce germe pouvait, lui-même, être issu d'autres activités neuronales... qui pouvaient, elles-mêmes être de lointains résultats du germe. Cette "pirouette" conceptuelle consiste à résoudre le "problème du début" par l'artifice de l'auto-référence, c'est-à-dire "du système clos sur lui-même". La situation est un peu analogue à ce qui se passe en cosmologie, suivant que l'univers soit imaginé ouvert et "avec bords", ou bien "clos", à l'image d'un tore, donc ne possédant ni bords, ni débuts, ni fins: bien que fermé, un tel

univers est infini, en particulier, le nombre de trajets ou cheminements qu'on peut y faire est infini de même que le problème de l'origine (big-bang) peut être "esquivé" (ref Hawking 1989).

Dans un réseau neuronal, cet univers infini mais clos, au sens de l'auto-référence, serait un réseau où toute représentation interne est le fruit d'une dynamique, d'une activité issues d'autres représentations internes qui sont le fruit... On reconnait là la boucle infinie des systèmes auto-référentiels qui ne peuvent se définir, donc se représenter qu'à partir d'eux-mêmes et de leur propre auto-représentation.

Donc, tant Bergson que Changeux ont raison car la notion de représentation serait "sans objet" dans un système auto-référentiel; nous verrons dans Perez, 1990, exposant le paradigme de "l'holo-ordinateur", comment, dans de tels systèmes, information et opérateurs sont fusionnés, auto-couplés et indissociables. De tels systèmes échappent à toute description ou modélisation représentationnelles... Pourtant, ils font preuve de mémoire, mais d'une mémoire dont la représentation échappe à tout observateur.

Je pense qu'on peut fort bien clore ces spéculations par cette citation de Fodor, rapportée par Changeux à propos de l'apprentissage.

> "Je montrerai non seulement qu'il n'y a pas de théorie de l'apprentissage mais que, dans un certain sens, il ne peut pas y en avoir."

J. Fodor (1980), rapporté par J.P. Changeux
"L'homme neuronal" (ref Changeux 1983)

Bibliographie

CHANGEUX 1983: *"L'homme neuronal"*, J.P. CHANGEUX, Ed. Fayard 1983.

HAWKING 1989: *"Une brève histoire du temps"*, S. HAWKING, Ed. Flammarion 1989.

HELLMAN 1981: *"Les mathématiques de la cryptographie à clef révélée"*, M. HELLMAN, dans *"Les progrès des mathematiques"*, Ed. Belin 1981.

MORTON 1985: *"Headed records: a model for memory and its faillures"*, J. MORTON et al, revue COGNITION vol. 20-1, June 1985, Elsevier New-York.

PEREZ 1989: *"De nouvelles voies vers l'intelligence artificielle"*, J.C. PEREZ, Ed. Masson 1988/ 1989, ISBN: 2-225-81815-0.

PEREZ 1990: *"La nouvelle cybernétique des hologrames fractals"*, J.C. PEREZ, à paraître Ed. Masson, 1990.

PRIGOGINE 1988: *"Entre le temps et l'éternité"*, I. PRIGOGINE et I. STENGERS, Ed. Fayard 1988.

Klaus D. Dutz

Westfälische Wilhelms-Univ., Inst. für Allgemeine Sprachwissenschaft,
Münster (Germany).

Models of Models of Meaning. Representation, Knowledge and Communication in Cognitive Linguistics and in Leibniz

Abstract

This paper deals with some of the philosophical implications of the modern cognitive linguistics' view on mental representation(s) and knowledge. We specially discuss methodological problems of adequcy and transferability of "language-of-thought" models (as J.A. Fodor proposed them) and "experimental realism" (as being coined by G. Lakoff). Herein, the theories developed by G.W. Leibniz, concerning representatio *and* expressio *take a certain space as 'models of proof'. It will be demonstrated that certain Leibnizian models, in a special although differently structured form, prove to be lost "predecessors" of today's cognitive models of science. From this, a fallacy in some semiotic conceptions may turn out, and it may finally be shown that cognitive Linguistics preserves a 'Cartesian paradigm', while the question rises whether a 'Leibnizian paradigm' shows up to be more adequate for the development of a cognitive semiotics and communication theory.*

Resumo

Este estudo trata de algumas das implicações filosóficas contidas no ponto de vista da linguística cognitiva moderna sobre a(s) representação(ões) mental(ais) e o conhecimento. Especialmente, serão discutidos os problemas metodológicos de adequação e transferibilidade dos modelos de "linguagem-do-pensamento" (J.A. Fodor) e de "realismo experimental" (G. Lakoff). As teorias desenvolvidas por G.W. Leibniz sobre "representatio" e "expressio" serão abordadas como "modelos de prova". Demonstra-se que certos modelos leibnizianos são, numa forma especial embora diferentemente estruturada, os "percursores" esquecidos dos modelos cognitivos da ciência hodierna. Pode ressaltar daqui uma falácia em algumas concepções semióticas e, por último, ser mostrado que a linguística cognitiva preserva um "paradigma cartesiano", suscitando-se a questão de saber se um "paradigma leibniziano" não será o mais adequado ao desenvolvimento da semiótica cognitiva e de uma teoria da comunicação.

In cognitive sciences, it is most common to talk about models. To talk about models is dangerous. There are very few terms being as much misunderstood as this one. So, now I am asking you to take the boat together with me and follow me through a terminological swamp. We shall see some strange model alligators looking from the one side to our boat, and we will have to ask us whether they have already retired or are as virulent as they are

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 271/277
1992, APDC & SPF, Lisboa

looking, and we shall look at some model fishes, not knowing if they still are palatable. Sometimes, we have to ask ourselves how all these fish may live together in this swamp, and I shall try to give a certain reply.

But first of all, let us ask why we should make this trip. As I do not know your reasons, I give mine: starting from linguistic and semiotic historiography, we certainly find us challenged to compare the different meanings of models and the different models of meaning. It goes for the understanding of what we really mean by linguistic, semiotic, or communicative knowledge about the forms of scientific recognition of the field. This is important, for insofar as we do not know the background of our reflection, we do not even know the legitimation of our scientific reasoning in front of the background of the science itself.

We all know the questions about: what is reasoning? How do we make sense out of our experience? What is a conceptual system, and how is it organized? Is there a conceptual system common to all men? If so, what kind is this system? If not, what is the difference, and what are the universals of human reasoning? All these questions are as old as human thinking is, but some answers seem to be new.

First, the answer of 'cognitive science', or cognitive scientists. They seem to fight against the assumption that meaningful concepts and rationality are *transcendental*, in the sense that they are abstract and disembodied. It may happen that it is embodied in organism, but then it is not necessarily so. There remains the question what kind of necessity is meant in this context, but I leave it open for a moment. The opposite is that reason has a bodily basis. This, of course, brings along with it the question of what kind of bodily – and let us now say: human – condition and process produces that reason in question. The nature of the thinking organism and the way it functions in its environment are of central concern to the study of reason.

Second, the answer of 'artificial intelligence research'. Its central aim always is to find models of reasoning, which are adequate, fulfilling the central items and producing a *representation* of the object or process in question, which then, in itself may function as another model for a artificial process. This thought is as old as the thought about human thinking and reasoning, but, of course now, we have for the first time the opportunity to bring it to birth with the help of grown technology. One might by the way, suspect a central mistake in this argumentation, because grown capacities in technical handling do not necessarily mean grown possibilities of finding new insight into the real human processes of reasoning. These are other conditions of necessity, implying necessary and sufficient arguments in state of a scientific argument, but I will return to this later, too. Again, it is the question of the state of models, we have to think about.

Third, there is the science of semiotics, who tries to find its own history, and by reconstruction of semiotically relevant theories of former centuries, it always faces theoretically important features of models, which, somehow miraculous, always seem to fit into the needs of today's science, at least, today's semiotic. Historiographical work in semiotics seems to have to do something with 'self-fulfilling prophecies', one has to think, as it goes without saying. But let us stop here for one thought. *WHY* does semiotics in its history produce such results? Is there the suspicion that historiography produces the results already been asked for in the actual discussion. If so, what builds the "pink glasses" of historiographers? There might be another kind of necessity, and this is the next, to which I have to return later.

Our starting point was the seemingly difference between 'old' and 'new' thought about the epistemological sources of reasoning. George Lakoff, in his study named impressingly

Women, Fire, and Dangerous Things, gives an outlook of the premisses, we have to face, and also gives us a terminological string, when he is naming the non-cognitive view *objectivism* – there is, of course, another 'objectivism', to be found in 13th-century Arabic philosophy, but it has nothing to do with our author's argumentation – and the so-called "new views" of *experimental realism* – which gives strong remembrance to medieval *realism*, and this, not without allusions to it in the writings of Lakoff, but to be correct, he himself does not hint to this – and which he also calls *experimentalism*.

The premisses of 'objectivism' are that:

- Thought is the mechanical manipulation of abstract symbols.
- Mind is an abstract machine, manipulating symbols essentially in a way a computer does, that is by algorithmic manipulation.
- Symbols (e.g. words and mental representations) get their meaning via correspondences to things to the external world.
- All meaning is of this character.
- Symbols that correspond to the external world are *internal representations of external reality*.
- Human mind is a *mirror of nature*, and correct reason mirrors the logic of the external world.
- Thought, therefore, is *abstract* and *disembodied*. It is independent of any conditions of human body, human perceptual system, etc.
- Machines that do no more than mechanically manipulate symbols that correspond to things in the world are capable of meaninful thought and reason.
- Thought is *atomistic*, in that it can be broken down into simple symbols used in thought, which are combined into complexes and manipulated by rule.
- Thought is *logical* in the narrow technical sense.

Some of these theses seem to be common, some of them are strange. But first, let us have a look at the other side. What is the view of *experimental realism*?

- Thought is *embodied*, the core of our conceptual systems is directly grounded in perception, body movement, etc.
- Thought is *imaginative*, in that concepts go beyond the literal mirroring, or *representation*.
- Thought has *Gestalt properties* and is thus not atomistic.
- Conceptual structure can be described using *cognitive models* that have the above properties.

Again, there is this mixture of common and strange theses. Unfortunately, we are blocked by the author's composition of pros and cons, because up to this time we were used to take them in another form which is cutting through this confrontation diametrically.

What, in fact, is Lakoff heading for? Obviously, to give linguistic and philosophical background to the so-called *prototype-theory* of Eleanore Rosch and her associates. He therefore aims to make clear, via cognitive categorization, that prototype effects occure in linguistic structure. To him, linguistic structure makes use of general cognitiv apparatus, equalized with category structure. Linguistic categories are kinds of cognitive categories.

At this point I like to change the direction of my talk a little bit and add to the thoughts of Lakoff those of Jerry A. Fodor, who, in his study about *Psychosemantics*, has his own thoughts about the problem of meaning in the philosophy of mind. He also takes the position of embodied thought, connecting it with what he calls the *Language of Thought Hypothesis*, based on Intentional Realism. His questions are: could beliefs and desires be material? Could they be immaterial? And, how can one know that we have any? To him,

the answers are obvious: It is 'yes' to the first one, because whatever has causal powers is ipso facto material, and therefore the answer to the second question is of course 'no'. And for the third case, his point is that there are no alternative theories than the commonsense belief/desire psychology, which is explaining the facts. And now, following his arguments, we arrive at the same point as we did in Lakoff: Mental states are not, however, the only things that represent. The other candidates are *symbols*. And symbols as well as mental states have, must have, a *representational content*. Always and again, as I would like to cite Fodor, every time a philosopher of language turns a corner, he runs into a philosopher of mind who is pounding the same beat.

The main business of both, then, seems to be the problem of *representation* itself. Fodor calls it the "metaphysical question of the place of meaning in the world order". How can anything manage to be *about* anything? Again, representation, as we have heard by Lakoff, above. But, it is not a different kind of approach which Fodor takes, concerning this crucial term of representation? It is obvious, that the first takes the term as a mental device, speaking of *internal representations* of external reality, which leads to the concept of mirroring of reality in mental concepts. That would be the so-called objectivism, and so far contraposed to *experimental realism*, as there thought is understood as "imaginative", beyond representation. Obviously, our both authors talk about different terminological items, both named representation. Fodor's position is, that there are various things to do usefully when your car gets a ping in its cylinders, but to quantify imaginatively over the engine is not among them. You need a story about the engine to explain how the car behaves. It is rational strategy you need, and it is about representation of objects, of action, and of behavior. If you have trouble with the thing, get it fixed.

So now, as we focused on *representation*, how shall we get it fixed? Of course there are lots of philosophical ways to do so, and not only one seems to lead to solution. Unfortunately the most of them (taking the philosophy of language path) seem to take that direction, which goes saying, "If there are deep and difficult problems about representation, then we won't have any representation". And others, taking the path of philosophy of mind, add: "If no representation means no belief/desire psychology, then *we* won't have any of *that*". Personally, I do not like this 'troublemakers-not-allowed'-strategy, but then I have to ask where the possible answer is.

What I propose is to open the old box of history of philosophy and semiotics, and have a short look what we might find about the story of representation there. Again we see, that is very common to talk about representation there. Welcome back to the terminological swamp.

For Cartesian philosophers, the question of 'representation' is a question, how the mind behaves. For them, the mental sphere seems to be organized quite clearly in states of adequate representation of reality – remember the 'objectivist's view' of Lakoff. What bothers them, is language, for somehow language does not behave like reality and mind, respectively. So, representation is *not* part of language, but language is only serving the function – as one among other functional devices. So let us hear both sides. What is the opposite view?

There is another philosopher, Gottfried Wlhelm Leibniz, who takes it. His two central concepts in this are *expressio* and *repraesentatio*. It was already Norman Kretzman in the late 60ties, who noticed that these two terms are building a point of intersection between ontology or epistemology and language philosophy in his thinking. To Leibniz, both terms are to be seen as two sides of one coin. Of course, the represented thing expresses something, and this is the *expressio* of it. Vice versa, the *expressio* represents something

and this is the *repraesentatio*. Let us have a look at this examples, which are taking a somewhat abstract view. On the one hand there is the map, which represents the sole ground, the talk, which stands for the thoughts and truths (you see, Leibniz still is taking the objectivist's line), but then it becomes interesting to see what he is adding: (For historiographical and terminological clearness, I give the Latin terms of expressed things and things expressed:)

caracters	numeros
circulus magnus	circulus parvus
ellipsis	circulus
omnis effectus	causa plena
facta cuiusque	anima eius
idea circuli	circulus
ellipsis	circulus

This does not sound very much on the objectivist's line, but seems much more to come near what we heard about experiemental realism – or a language of thought hypothesis. What is common with a circle and an ellipse? And why is there the same relation between an ellipse and a circle als thing expressed and expressed thing, and vice versa? Are we not used to say that the one thing is a – semiotically speaking – representamen of the other. And now, Leibniz is telling us that this relationship may be mutually converted? And what is about the whole effects expressing the whole causes, while the cause, via *idea* is reconverted into the expressed thing itself?

Obviously, we are wrong if we try to start from the examples and try to find the common ground. This common ground must be a principle, and this principle itself lies in the relation between thing expressed and expressed thing, between the process of representation and the human mind.

As we see from the example of idea, the simple case of analogy, or even, similarity, cannot be meant by Leibniz. Customary, Leibniz's thought on representation is interpreted in the frame of 'structural isomorphy'. The normal critique, as customary as the opinion, refers to the concept of a isomorphic relation like a projection one-to-one. This is not quite correct, since a isomorphic relation deserves certain additional criteria: First, that it builds a relation between sets, in which each element of one set is mutually appointed to each element of the other set, second, that this isomorphic projection is possible insofar, as we keep the operations between the elements of the first set and their corresponding predicates also in the projection into the other set, and third, that we have look to the fact that if each element of the first set has a certain predicate, it has to get a certain equivalent predicate in the second set. Therefore, we are not free to speak of isomorphic sets generally, but only of isomorphically structured sets in respect of certain relations between subject and predicate in concern with set-framings.

Using 'representatio' and 'expressio', Leibniz ever and ever tells us his famous example of the relation between a town and its different aspects in perspective. You remember the town, you remember the relations, you literally recognize this town under each condition, although the sight changes in every respect: What now is the front of a house, then is the back, you see a street not from the east, but from the west, and so on. But one thing never changes: it is the relation of the house in respect to the street, in respect to its garden, it is the relation of one street to the other.

Therefore, Leibniz tells us, it is not the content of the perceived elements of sensations, nor the sensations or the fact of perceiving them itself, which gives us this impression of this certain town, but there must be a special and single power in the acting of the human

mind, which forms this relation, and that it must be ruled by a certain law, which he calls *lex expressionum*. And, as we have already said, this is only a set of conditions which are vice versa able to build a *lex repraesentionis*.

There are two sources which feed this thoughts of Leibniz. The one sidecar is to be found in the tradition of the concept of man as *mikrokosmos*, who embodies all knowledge in himself – opposite to the already cited frame of "objectivism" –, and which can be traced back via Heraklite, Empedokles to Parmenides of Elea, Anaxagoras, and, making a jump via Aristotle – see the second book of *De Anima*– to Boethius, Thomas Aquinatis and Giordano Bruno. The second source obviously is to be found in the discussions between Leibniz and Spinoza, concerning the status of the term *idea*. Leibniz defines *idea* as something, "quod in mente nostra est". But, scientifically speaking, this is not satisfactory up to this, because someone, having a method of scientific reasoning to approach the external objects, does not have selfwithstanding the *idea* of these objects. With Leibniz's own words: "Necesse est aliquid in me, quod non tantum ad rem ducat, sed etiam eam exprimat". For epistemological reasoning, we do not only need the mind to lead us to the objects, but also something in the mind, which also expresses and mutally represents the objects.

So, in respect to the historical time and its terminological properties, Leibniz's aim to find principles of human reasoning leads to a similar result as the one we have seen on Lakoff. But, I think, with a certain difference. Leibniz always is in regard of his historical sources, as we have tried to point out before. He knows about the epistemological premisses of discussion, and he knows about the discrepances between them.

Before I continue, I like to take up one contraposition in advance, the one asking me why I did not give respect to Locke, or to *sensualism*. On this, I would like to take the part of George Berkeley, who reintroduced in his *Treatise Concerning the Principles of Human Knowledge* an argument wellknown since the scholastics into philosophical discussion: his question, familiar to us, is whether sensa data are logically depending on the objects of sensation, or whether the sensation itself produces the epistemological basis. At least he is able to give an example of logical contradiction in Locke'opinions about sense data reception, since Locke is referring to the fact that singular sense data always represent singular objects, and the proposition of existence of their singularity necessarily implies the existence of the sense data. This is, of course, a tautology.

Now, let us return to the argumentation from which started. What was it about the difference between *objectivism*, and *experimentalism*? The point of view of Lakoff is that there is a contradiction between the assumption of disembodied thought models and the premiss, that there are cognitive models on their own, which, undisguisingly floating through our swamp, make the conception of embodied cognitive models necessary. The objectivist's view, he said, incorporates also the cognitive mirror–model, while the view of experimental realism shows up that there are *Gestalt* – properties at work. Now, as we have seen, Leibniz tells us, with good arguments, that there are in fact the premisses of objectivism, internal representation of external reality, disembodied thought, and production of meaning through via symbolis via correspondences to things in the external world, that lead directly to some other premisses of experimentalism, the aspect of supersumative power of concepts, and of relation instead of statical conditions in cognitive area.

This is what I wanted show, when I told you at the beginning that the criteria were crossing diametrically through both sides of seemingly oppositive concepts. A certain look at historical development of the discussion in fact gives this impression. Now, turning to Fodor, we realize that his approach only at first seems to be similar to that of Lakoff. But,

as we have seen, Fodor does not give some strange and undeclarable aspects the task to function as epistemological devices, but tells us that there is representationality of the object of science, which is in question. Fodor, on his hand, takes the view of an "objectivist" and uses the epistemological devices of a medieval nominalist and is, therefore, not falling into the trap of realism. Lakoff, on the other hand, combats 'realism', and falls into the traps of a surface nominalism, because his pros and cons are mixed up between all parties.

Now, what about the necessities I talked about a few minutes ago? It is a common saying in modern cognitive linguistics that some argument or hypotheses must follow "necessarily" from the other. If we leave the rhetoric circumstances apart, we realize that there is a somehow a-historic understanding of necessity. The first necessity, cited as argument of cognitivist against the belief that meaningful concepts are transcendental, goes that thought *necessarily* is embodied. But the objectivist's argument on thought aims to logical matter, while the opponents take the ontological part, which is at least senseless, as we have seen in Leibniz. So, there is no real contradiction, but only a change in argumentational level. The second necessity mentioned went to the address of artificial intelligence research. It goes to the aspect that a model-building-system always is forced to take the necessary properties of the reflected basis. But this does not mean, on the other hand, that it necessarily does so, only that it is this by hypothesis. This, of course seems to be very trivial, and one wonders how often this aspect remains unreflected. There, finally, is the supposed necessitation in writing historiographies of a certain science, under conditions which are off of this context. But we should keep in mind that handling with epistemological models implies the look back to the results of the history of science, and that our understanding of it always is only an interpretation.

With this, I have to leave the swamp for today, I have caught only a few fishes, and the alligators all went by. But I understand, that this trip cannot be finished in one day and in one talk.

Danilo Marcondes de Souza Filho

Dep. de Filosofia, Pontifícia Univ. Católica
Rio de Janeiro (Brasil)

A Questão das Linguagens Artificiais no Pensamento Moderno

Resumo

O propósito deste estudo é discutir algumas das implicações filosóficas dos projectos de formulação de linguagens universais correntes durante os séculos XVII e XVIII. Estas linguagens são consideradas do ponto de vista da sua relevância científica e comunicativa, assim como enquanto tentativas de formulação de linguagens artificiais susceptíveis de processar informação de modo mecânico. Consideram-se com brevidade algumas das razões que levaram ao fracasso destes projectos.

Abstract

The aim of this paper is to discuss some of the philosophical implications of the projects of formulating universal languages current in the XVIIth and XVIIIth centuries. These languages are considered from the point of view of their scientific and communicative relevance, as well as attempts at formulating artificial languages which would process information in a mechanical way. Some of the reasons for the failure of these projects are then briefly considered.

As tentativas científicas de criação de linguagens artificiais que permitam o processamento mecânico de dados e informações, tornando-o mais rápido e eficaz, contribuindo assim para o progresso da ciência, que caracterizam o desenvolvimento da informática em nosso século, não são, ao contrário do que se pode supor, um projecto novo, mas fazem parte da tradição cultural do Ocidente desde pelo menos o final do Pensamento Medieval. Sobretudo a questão da relação entre pensamento e linguagem tem importância decisiva para as tentativas de construção de "máquinas de pensar", levando à discussão sobre a natureza do pensamento e de sua expressão linguística. Esta questão é fundamental exactamente na medida em que essas máquinas operariam com a linguagem de forma a se aproximar cada vez mais da maneira como o pensamento opera. A necessidade de uma representação linguística adequada do pensamento é, portanto, um dos problemas centrais nesta discussão.

Já no *Crátilo* de Platão encontramos a ideia da necessidade de eliminar os limites e imperfeições das linguagens naturais, em grande parte responsável pelo papel secundário, do ponto de vista filosófico, atribuído às investigações sobre a linguagem no pensamento clássico. Com efeito, se os signos linguísticos são meramente sinais convencionais, conclusão a que Platão chega após reconhecer a impossibilidade de demonstrar a existência de uma relação natural entre o sinal e o objecto por ele denotado, então estes signos não

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 278/283
1992, APDC & SPF, Lisboa

podem ter nenhuma função relevante no processo de conhecer a realidade, servindo apenas para o registro e a expressão deste conhecimento que o antecede totalmente. Segundo esta visão, portanto, é o pensamento e não a linguagem que deve ser objecto da análise filosófica, já que o pensamento sim seria capaz de captar a essência do real. Essa concepção platónica marca profundamente o desenvolvimento da tradição filosófica em seu modo de considerar a linguagem.

Com isso, o estudo da linguagem é praticamente relegado aos domínios da *retórica*, como sistematização de recursos estilísticos, e da *gramática*, como classificação e descrição dos elementos componentes de uma língua empírica, que adquire carácter normativo ao procurar padronizar o uso da língua, estabelecendo os critérios e regras para o emprego correto dos termos e para a construção adequada de sentenças. Essa preocupação com o estudo sistemático da gramática, iniciada pelos gregos (Dionísio Trácio, *Téchne Grammatiké*, c.100 a.C.) continuada pelos latinos (p. ex. Varro, *De lingua latina*, 116–127 d.C.; e Prisciano, *Institutiones grammaticae*, c.500 d.C.), terá um desenvolvimento especial no pensamento medieval com o surgimento das *gramáticas especulativas*. O projeto medieval da gramática especulativa representa, de fato, o que podemos entender como o primeiro passo na direção de uma consideração autenticamente filosófica da linguagem, encontrando-se assim na origem do projeto de construção de uma "linguagem filosófica" que tornar-se-á generalizado nos sécs. XVII e XVIII e que constitui nosso principal objeto de estudo neste ensaio.

As gramáticas especulativas distinguem-se das meramente descritivas e normativas por pretenderem constituir uma ciência, no sentido aristotélico sobre a linguagem, isto é, um corpo de conhecimento necessário e dedutivamente demonstrado, formulando uma **explicação**, no sentido causal, de seu objeto de investigação. Enquanto a gramática clássica é meramente descritiva e normativa, estabelecendo quais as construções corretas em uma língua, surge com as gramáticas especulativas a necessidade de explicar e sobretudo de justificar esta idéia de correção e seus respectivos critérios. Certamente, há tantas gramáticas quanto há línguas, porém só pode haver ciência se houver um princípio unificador, é desta forma que surge então a busca de uma ciência da linguagem, de certos princípios gerais subjacentes a todas as línguas, que permitam a formulação duma lógica geral da linguagem e estabeleçam a ontologia da linguagem. Considera-se que as línguas representam o mundo, refletindo-o como um espelho (daí o termo *speculativa*, de *speculum*), tratando-se então de investigar os modos desta representação especular, os *modi significandi*. A maioria desta gramáticas inspira-se em interpretações dos tratados aristotélicos *Categorias* e *Da Interpretação*, sendo a mais famosa a *Grammatica Speculativa* de Tomás de Erfurt (c. 1325).

É precisamente a concepção da existência de uma linguagem universal subjacente a todas as línguas, cujo germe se encontra nestes tratados medievais, que inspirará o projeto seiscentista e setecentista de formulação duma *linguagem artificial* que constitua exatamente uma codificação, através duma notação adequada, desta linguagem universal e de sua estrutura lógica, permitindo assim eliminar e superar as diferenças entre as diversas línguas naturais, bem como suas limitações e imperfeições.

Passemos então a examinar especificamente este projeto de construção artificial de uma linguagem universal, que torna-se, segundo alguns historiadores, um autêntico lugar-comum entre filósofos e cientistas nos sécs. XVII e XVIII. Pretendo considerar aqui sobretudo o aspecto linguístico deste projeto, procurando explicitar a concepção de linguagem por eles pressuposta e suas implicações filosóficas quanto à questão da representação do conhecimento. As seguintes questões orientarão basicamente esta breve

tentativa de análise:

1) Qual a *concepção de linguagem* subjacente a esses projetos?

2) Quais os *objectivos* da construção destas linguagens?

3) Qual o *sentido filosófico* destes projetos?

4) Quais as razões principais de seu suposto *fracasso* e consequente *abandono* neste período histórico?

Não pretendo, entretanto, procurar responder a essas questões uma a uma, isoladamente, mas sim considerá-las em bloco, devido à sua inter-relação.

As tentativas de construção de linguagens universais por autores dos sécs. XVII e XVIII têm um precursor importante no final do período medieval, inclusive reconhecido como tal por vários desses autores, que é Raimundo Lúlio (1233-1315) com sua *Ars Magna*. Essa obra distingue-se essencialmente das gramáticas especulativas, das quais é praticamente contemporânea, por ir além delas na medida em que não pretende apenas buscar os princípios gerais subjacentes a todas as línguas, mas antes construir um método efetivo para o descobrimento de verdades a partir de um *cálculo* que engendrasse mecanicamente novas definições segundo certos princípios gerais e definições básicas pressupostos. A proposta de Lúlio será explicitamente retomada por Leibniz e constitui não só uma tentativa de construção de uma linguagem artificial, mas principalmente de uma combinatória, ou seja, de um método efetivo e construtivo de produzir verdades, e portanto conhecimento, a partir de certas classificações de conceitos gerais, combinando estes conceitos entre si, produzindo assim novos conceitos, de acordo com determinados princípios. O método inclusive não se limita a uma formulação teórica, mas envolve a construção concreta de uma série de discos, que girando em torno de um centro comum, produziam as combinações necessárias. Trata-se, assim, de uma espécie de ancestral remoto da máquina de calcular e, porque não, do próprio computador.

O primeiro tipo de proposta de construção de uma linguagem universal no séc. XVII, entretanto, corresponde a um género bastante diferente de motivação. Com a progressiva substituição do latim como língua franca da Europa Ocidental pelos diferentes vernáculos regionais e nacionais, cria-se uma dificuldade de compreensão e de comunicação entre comerciantes e, em menor escala, entre cientistas e sábios, que dominavam ainda em sua maioria o latim. Isso é devido, ao menos em parte, à associação do latim com a Igreja Católica e à progressiva secularização da sociedade. Surgem então os *dicionários poligráficos* ou universais, por meio dos quais procura-se estabelecer uma correspondência entre os termos de diferentes línguas que tivessem um significado comum, designando-os frequentemente por uma notação numérica. Assim, por exemplo, poderia ser atribuído ao termo "homem" o número 500, a partir daí listando-se o seu equivalente em outras línguas. A primeira parte do dicionário relacionaria então termos em uma língua base, frequentemente ainda o latim, com números determinados. A segunda parte, organizada numericamente, conteria os termos equivalentes em outras línguas. A ordenação numérica permitiria classificar termos e conceitos em grupos afins, facilitando sua localização. Incluem-se dentre esses desde a *Polygraphia* de Johann de Tritheim, ainda no séc. XVI, até trabalhos como *The Universal Character* do inglês Cave Beck (1657), o *Character pro notitia linguarum universali* do alemão Joachim Becher de 1661, e a *Polygraphia nova et universalis* de Athanasius Kircher (1663), para citar apenas alguns.

Aparentemente esse projeto ter-se-ia inspirado na generalização do chinês escrito como língua franca no Extremo Oriente neste período. Por ter uma escrita ideográfica o chinês seria compreensível enquanto escrito mesmo por falantes de diferentes dialetos e línguas do Oriente que não poderiam se comunicar entre si através da linguagem falada. Há

uma referência neste sentido no *The Advancement of Learning* (1605) de Francis Bacon (livro II, séc. XIV, 1).

O passo seguinte em direção à formulação de uma linguagem supostamente universal se dá também neste período através da abstração do simbolismo, geralmente uma notação numérica, utilizado para a elaboração do dicionário poligráfico. Cria-se então uma linguagem artificial que substituiria assim as diferentes linguagens naturais. Como exemplo, temos do inglês Francis Lodwick (1647) *A Common Writing*. Tratam-se essencialmente de sistemas de cifrar um número conhecido de línguas em uma estrutura unificada.

É exactamente esse tipo de projeto, representado pelo dicionário poligráfico e pela linguagem universal artificial construída simbolicamente, que é criticado por Descartes numa carta dirigida a 20 de novembro de 1629 ao Pe. Mersenne, que viria ele próprio a construir uma linguagem deste tipo entre 1636-1637. Nessa carta, comentando a proposta, hoje perdida, de construção de uma dessas linguagens por um certo Hardy, Descartes mostra-se profundamente cético. Para ele, trata-se, no fundo, de apenas uma linguagem cifrada, uma espécie de taquigrafia, sem nenhuma grande vantagem sobre as línguas que deveria substituir, na medida em que reproduziria suas imperfeições lógicas e não contribuiria para um melhor conhecimento do real. Estas linguagens viriam então simplesmente se sobrepor às línguas já existentes, sem de fato acrescentar nada exceto uma certa economia na expressão e rapidez no aprendizado, sendo que até isso desapareceria caso estas linguagens se generalizassem e se tornassem mais complexas.

Para Descartes, a construção de uma linguagem desse tipo, que tivesse realmente relevância filosófica, ou seja, que fosse um aperfeiçoamento lógico da linguagem natural, só seria possível a partir da existência acabada da *verdadeira filosofia*, para que aí então se determinassem os conceitos básicos do entendimento humano, segundo os quais a linguagem seria construída. A elaboração de uma linguagem universal dependeria assim então da existência efetiva de um conhecimento universal que esta linguagem expressaria. A possibilidade de unificação da linguagem pressuporia assim a unificação da ciência. Este argumento reflete a concepção cartesiana, mais próxima da clássica, de que a linguagem não tem nenhuma função relevante no processo de construção do conhecimento, que dar-se-ia apenas através da relação entre o pensamento e o real, isto é, da representação correta da realidade por meio da idéia. A linguagem apenas expressaria um conhecimento previamente constituído, como dissemos acima (cf. pág. 279). A filosofia de Descartes é sobretudo uma *filosofia da consciência* e, enquanto tal, não admite papel importante à linguagem.

A proposta de construção de linguagens universais artificiais evolui, a partir desse momento, no sentido da busca da construção do que se considerava uma "linguagem filosófica". Essa linguagem deveria determinar critérios de classificação de certos conceitos básicos que formariam seu núcleo e a partir dos quais os demais seriam derivados através de certas regras. Tudo isso deveria ser formulado em uma notação simbólica que explicitasse estas relações e derivações de maneira clara e rigorosa. Há assim uma diferença importante entre essas propostas e as dos dicionários poligráficos, de que são praticamente contemporâneos, já que correspondem até certo ponto a objetivos diversos. Dentre as propostas desse tipo encontramos a do tcheco Comenius (1651), a do inglês Seth Ward (1654), professor de Astronomia em Oxford, desenvolvida pelo seu seguidor George Dalgarno (*Ars Signorum*, 1661), e a mais completa e famosa, a do inglês John Wilkins, intitulada *An essay towards a real character and a philosophical language*, escrita por encomenda da Royal Society, da qual o autor era secretário, e publicada em 1668.

É Leibniz, conhecedor da obra de Wilkins, que, no entanto, tratará a questão de forma decisiva, sendo que seu projeto influenciará mais tarde as propostas de Frege e Russell. Retomando as objeções de Descartes ao Pe. Mersenne na carta citada, Leibniz considera, contra Descartes, que embora seja preciso supor a *verdadeira filosofia* para a elaboração de uma linguagem deste tipo, não é necessário, entretanto, supor que esta filosofia esteja acabada ou completada, isto é, que tenhamos uma classificação total dos conceitos e definições básicos para construir esta linguagem. Ao contrário, para Leibniz, uma linguagem construída nesses moldes serviria precisamente de instrumento para se chegar a este conhecimento perfeito, a esta ciência universal unificada. Inverte-se assim a questão. Ao invés de a linguagem universal pressupor a ciência universal, é esta que supõe a linguagem universal para sua realização. Encontramos aí então toda uma linha de pensamento desenvolvida depois por Condillac já no séc. XVIII e que influenciará tanto as propostas de construção de linguagens lógicas dos sécs. XIX e XX, quanto as considerações sobre a importância da lógica para as teorias científicas, como, p. ex., o projeto de membros do Círculo de Viena de elaboração duma Enciclopédia da Ciência Unificada e de aperfeiçoamento de uma linguagem lógica para a ciência.

Essa *Characteristica universalis*, segundo Leibniz, teria três funções primordiais:

1) Constituir-se em uma *língua internacional*, permitindo a comunicação entre indivíduos cujas línguas maternas fossem diferentes. Seria, portanto, uma língua artificial no sentido das mais recentes ou contemporâneas como o Esperanto, o Volapuk, a Interlíngua, a Interglossa, dentre outras.

2) Formular um *sistema simbólico* simplificado que permitisse expressar de maneira exata e rigorosa o conhecimento efetivo e possível, promovendo assim a possibilidade de unificação da ciência.

3) Ter uma *função heurística*, servindo como instrumento de descoberta e de demonstração, exibindo as implicações do que já conhecemos e possibilitando o desenvolvimento deste conhecimento.

Encontra-se desta forma em Leibniz uma espécie de síntese dos diferentes projetos desse período, acrescentando-se como sua contribuição específica e fundamental a retomada da ideia de Lúlio, mencionada acima (pág. 280), de uma *ars combinatoria*, ou seja, de um cálculo de conceitos que permita então, recorrendo-se a um modelo algébrico de cálculo, desenvolver o conhecimento. É esta última ideia, sobretudo, que terá influência marcante no desenvolvimento da lógica e das filosofias da ciência e da linguagem, e que levará em última análise à concepção de que a ciência pressupõe uma linguagem bem construída, que se encontra formulada já em Condillac (*La langue des calculs*, 1798), chegando até à concepção de cálculo racional dos lógicos do séc. XIX, bem como, mais tarde, às propostas de Frege.

Por outro lado, é, segundo autores como Cohen (1954), a incompatibilidade entre essas três tarefas exigidas deste tipo de linguagem que levará ao fracasso e consequente abandono dessa proposta ainda no contexto dos sécs. XVIII e início do XIX. Uma linguagem deste tipo não poderia constituir-se como língua internacional, ao mesmo tempo mantendo seu papel de unificação da ciência e sua função heurística. Além disso, a necessidade de classificação de conceitos supõe a consideração de seus conteúdos, o que inviabiliza o processo de formalização pressuposto pela construção do cálculo, que se reduz à formulação de uma notação simbólica. Só quando estas tarefas se separam é que poder-se-á então chegar à construção, com Frege, de uma lógica formal simbólica, seguindo a utilização da álgebra como modelo desenvolvida por Boole e por Morgan no séc. XIX. Esta proposta se encontra, em germe, já no próprio Descartes que defende a matemática como

a verdadeira linguagem da ciência, sem no entanto dar-se conta das implicações disto para as propostas de construção de *linguagens filosóficas* correntes em sua época.

Bibliografia

AARSLEFF, H., (1982) "John Wilkins", in *From Locke to Saussure*, Univ. of Minnesota Press, Minneapolis, págs. 239–277.

BURSILL-HALL, G.L. (1971) *Speculative grammars of the middle ages*, Mouton, Haia.

COHEN, L.J. (1954) *"On the project of a universal character", Mind*, LXIII, 249.

CONDILLAC, E. (1975) *Obras*, Col. Os pensadores, ed. Abril, S. Paulo.

DESCARTES, R. (1964–1975) *Oeuvres*, ed. C. Adam e P. Tannery, Paris.

LAND, S. (1974) *From signs to propositions: the concept of form in eighteenth century semantic theory*, Longman, Londres.

LEIBNIZ, G.W. (1975) *Obras*, Col. Os pensadores, ed. Abril, S. Paulo.

ROBINS, R.H. (1979) *A short history of linguistics*, Longman, Londres, 2ª ed.

SALMON, V. (1972) *The works of Francis Lodwick*, Longman, Londres.

5

A CIÊNCIA PERANTE A COMPLEXIDADE E A INCERTEZA

SCIENCE FACING COMPLEXITY AND UNCERTAINTY

Abraham A. Moles

Inst. Psychologie Sociale des Communications
Strasbourg (France)

Y a-t-il des Sciences de l'Imprécis?

Résumé

Nous vivons au milieu de phenomènes vagues, de choses imprécises, de situations perpétuellement variables dans lesquelles il faut nous décider, réagir ou agir, prendre position. Notre propos est de montrer que le mot science *signifie connaissance des formes regulières qui se présentent à l'esprit; il ne signifie que cela, mais il signifie tout cela, et donc il ne se confine pas aux sciences "conventionnelles" parce qu'exactes. A côté des sciences exactes, il y a des sciences de l'inexact, de l'imprécis, du flou, des corrélations faibles qui approchent, plus que les sciences de la nature, la connaissance du reél tel que nous avons à nous confronter avec lui, et pour lesquelles nous devons construire une epistémologie, une métrologie et une methodologie.*

Resumo

Vivemos no meio de fenómenos vagos, de coisas imprecisas, de situações perpetuamente variáveis no âmbito das quais temos de decidir, de reagir ou agir, de tomar posição. O nosso propósito é mostrar que o termo ciência *significa conhecimento das formas regulares que se nos apresentam ao espírito; não significa senão isto, mas significa tudo isto, logo, não se confina às ciências "convencionais" porque exactas. Ao lado das ciências exactas há ciências do inexacto, do impreciso, do esfumado, das correlações fracas e que se aproximam, mais do que as ciências da natureza, do conhecimento do real tal como temos de o enfrentar, para as quais devemos construir uma epistemologia, uma metrologia e uma metodologia.*

> Nous ne vivons que dans un des mondes: le nôtre. Ceux qui en sont, s'y meuvent aisément, puisque partout où ils vont, dans la lune aussi bien que sur la terre, ils sauront faire de la géométrie, compter les minutes et discuter sur les causes. C'est bien simple, ils tirent partout leur montre et mesurent les lieux.
>
> B. Groethuysen, préface à Kafka

Nous vivons au milieu de phénomènes vagues, de choses imprécises, de situations perpétuellement variables dans lesquelles il faut nous décider, réagir ou agir, prendre position. Si vagues soient-elles pourtant, toutes ces choses apparaissent à notre conscience comme des objets conceptuels, nous leur donnons des noms, et nous faisons sur elles des opérations, mentales d'abord, pratiques ensuite, à nos risques et périls. Vivre, c'est se confronter avec des choses vagues. Le monde n'est pas un laboratoire où les phénomènes sont épurés, isolés, contrôlés, au gré et au loisir de l'expérimentateur qui joue avec eux, pour découvrir une vérité transcendante, incontestable, car épurée sous la forme de corrélations fortes entre des variables évidentes. Nous parlons de température, et nous

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 287/302
1992, APDC & SPF, Lisboa

jugeons de bien-être, nous parlons de justice, et nous jugeons de nos intérêts, nous parlons de Bien et de Mal, et nous pensons investissements.

Les êtres et les valeurs qui nous guident dans notre vie quotidienne, qui s'imposent dans notre flux de conscience, ne sont pas, ou guère, de nature proprement "scientifiques" au sens conventionnel que notre culture a donné à ce terme. Cependant, c'est avec eux qu'il faut vivre et agir; ce n'est que dans des cas bien privilégiés, dans des situations spéciales, que nous nous trouvons véritablement confrontés avec des variables exactes dont la définition est claire et sans ambiguité. Le héros de Kafka était arpenteur, par son métier particulier il rencontrait les lieux et tirait son mètre pour leur donner cette exactitude physique sur laquelle nous sommes tous d'accord: le vendeur comme l'acheteur. Mais cette connaissance exacte qu'il en tirait, ne lui servait de rien dans son destin personnel.

Le destin scientifique de l'Occident

Et certes, notre monde a pris l'option de devenir de plus en plus "scientifique": la science est le seul véritable triomphe de la pensée, le triomphe de l'intelligible et de la cohérence universalle. Pourtant, la science n'est pas née toute armée du cerveau du physicien, elle est un processus avant d'être un achèvement, elle est un pénible effort pour recommencer perpétuellement à penser juste. En fait, dans le budget-temps de la vie de tous les hommes, la pensée scientifique "exacte" n'occupe qu'une bien faible proportion. Et ce que nous cherchons dans notre vie, c'est d'appréhender ces choses vagues qui nous entourent d'une façon un peu moins arbitraire que nous l'avons fait autrefois: c'est peut-être cela que nous appelons être rationnel: une démarche et non un état. Qu'avons-nous à la disposition de notre esprit pour nous aider à penser, à prévoir, à faire? Pas grand chose. La science telle que nous la connaissons ne nous parle guère de ce qui est imprécis, de ce qui est fluctuant, de ce qui change et ne se répète qu'approximativement. Elle préfère, par la voix de ses représentants attitrés, les corrélations fortes entre les variables aux corrélations faibles de la vie. Les phénomènes vagues, les relations imprécises, ce sont d'abord des systèmes où les corrélations entre les grandeurs sont faibles (0, 2–0, 4–0, 6) et donc les relations incertaines. Mais une corrélation faible est quand même autre chose que pas de corrélation du tout: elle est bien objet de science et un objet important car c'est le tissu de la vie réelle. Pourtant, ce sont bien là des *phénomènes*, des choses qui apparaissent à notre conscience avec des contours constants, avec un début et une fin, dont nous reconnaissons l'identité ou la similarité, avant d'examiner la différence: ce qui les fait varier. La forme existe avant ses variations, elle transcende ces variations, c'est ce que nous enseigne la Gestalt.

Et donc, ces formes dans notre conscience doivent être objets de science, tout comme le sont les longueurs et les courants; on les nomme, on doit pouvoir les catégoriser, les dominer, les énoncer, et peut-être en trouver des lois. Pourquoi le faisons-nous si peu, pourquoi nous reposer sur l'essai et l'erreur, sur les idées déjà faites, sur l'impulsion de l'instant ou "l'intuition", quand nous avons à traiter de toutes ces grandeurs imprécises: l'amour de notre maîtresse, la dureté de l'inspecteur des impôts, la séduction de l'argument publicitaire, la sagesse du chef, la fragilité du social. Parce que personne ne les a étudiées? Certes. Mais pourquoi donc, encore? La "catégorie" des phénomènes imprécis n'a pas – ou guère – de "science" ni méthodes qui lui soient propres. Ce domaine n'aurait-il pas de règles? C'est la réponse facile que nous suggèrent les sciences exactes: il est inexact donc sans règles et sans lois, il est anomique, il est rejeté du royaume de la Science.

En fait, c'est se détourner des idées vagues, des concepts flous, en les confondant,

allègrement, avec les idées fausses, et en abandonnant "tout cela" à une famille de disciplines mal séparées encore de la philosophie-mère qui les a engendrées qu'on range sous le nom - tout aussi imprécis - de "sciences de l'homme" ou "sciences sociales". Notons au passage, nous y reviendrons, que cette attitude imposée à travers le développement occidental de la connaissance rationelle, n'est pas la seule qui se présente dans le champ épistémologique. L'Orient en particulier a toujours pratiqué d'autres méthodes de conviction et de découverte. Même en Occident la "pensée en surface" par la figure clé, par le schéma démonstratif, par la coupe anatomique, par le rapprochement heuristique des tableaux numériques a proposé d'autres voies d'accès à l'évidence et à la prédictivité, avant de se rallier (provisoirement?) à la Raison Linéaire occidentale.

Notre propos est de montrer - sinon démontrer - que le mot "Science" signifie connaissance des formes régulières qui se présentent à l'esprit: il ne signifie que cela, mais il signifie tout cela, et donc qu'il ne se confine pas aux sciences "conventionnelles" parce qu'exactes, qu'il y a là une déviation, certes profitable à un certain moment de la pensée, mais déviation tout de même, de l'esprit humain qui s'est détourné d'un grand nombre de phénomènes parce qu'ils étaient vagues, et donc difficiles à maîtriser avec les outils dont il disposait. Il était plus simple, plus rentable, plus confortable, de concentrer son effort sur les sciences de la Nature, dont les objets sont peu dépendants de l'homme, ou, en tout cas, sont apparus comme tels jusqu'à la naissance de la micro-physique dans laquelle "la notion d'une réalité indépendante de l'observateur est apparue comme dépourvue de sens" (D'Espagnat).

Ainsi, à côté des sciences exactes(?), il y a des sciences de l'inexact, de l'imprécis, du flou, des corrélations faibles qui approchent, plus que les sciences de la Nature, la connaissance du réel tel que nous avons à nous confronter avec lui, et pour lesquelles nous devons construire une épistémologie (règles pour parvenir à la vérité), une métrologie (sciences et techniques de la mesure de l'imprécis) et une méthodologie (connaissance des procédés qui permettent à l'homme d'agir sur les choses vagues).

Catégories de l'imprécis

Il y a trois grands types de phénomènes vagues selon les causes de leur imprécision:

1) Il y a des phénomènes vagues parce que l'erreur probable sur leur détermination est grande ou très grande et que le phénomène en tant que forme a des contours vagues, ou variables changeant d'une occurence à l'autre. Nous avons certes à connaitre les causes de la variabilité de ces contours, mais en attendant nous avons à nommer et traiter ces phénomènes. Exemple: le micro-climat local d'une petite ville (pleuvra-t-il ou non demain?), est un "fait" qui, ici où je suis, a une très grande importance et qui échappe à notre appréciation "métrologique".

2) Un grand nombre de phénomènes nous restent vagues parce que nous ne disposons pas de techniques de *mesure* adéquates. Notre tâche indiquée est alors certes d'effectuer un effort *conceptuel* pour les enserrer dans des mesures possibles. Exemple: qu'est ce que la "grandeur d'une action?". Mais nous ne pouvons attendre d'y être parvenu et nous devons les traiter comme ils sont. Entrent dans cette catégorie un grand nombre de phénomènes sur lesquels il nous est impossible *d'expérimenter* à proprement parler - même si nous pouvons concevoir une expérimentation.

3) Il y a enfin des phénomènes *vagues par essence*, c'est-à-dire dont les concepts qui servent à les énoncer sont eux mêmes vagues, peut-être inadéquats mais ce sont les seuls dont nous disposions. Notre thèse sera qu'il est nécessaire de les étudier en se servant de

ce que nous avons sans vouloir nous efforcer de pousser les concepts vers une précision qui en détruit le sens, en développant les relations qui existent entre ces concepts d'une façon "rigoureuse", c'est-à-dire conforme aux lois de la raison.

L'idée de "cause possible" donnant lieu à un effet visible (phénomène) établit l'élément de base de la chaîne causale A + B: la propagation de causalité de A vers B (ou de B vers A?). Elle est concrétisée mathématiquement dans la notion de *corrélation*, et par un "diagramme de correlation" quand on sait mesurer, ou apprécier cause et effet comme des *grandeurs*. Si les grandeurs sont mal définies, c'est là déjà une bonne raison pour que la correlation soit faible. Mais là encore c'est mieux que rien.

En bref, quand la précision, la mesure, ou la conceptualisation s'avèrent insuffisantes, plutôt que de nous détourner des phénomènes par paresse de la conscience, acceptons le conseil de Henri Michaux à l'épistémologue: "Ne désespérez jamais, faites infuser davantage".

De la mesure et de la précision

L'Imprécis est le contraire du précis. La pensée humaine, surtout en Occident, s'est acharnée à conquérir la précision, c'est-à-dire à imposer sur le monde un système de mesures traduites par des grandeurs et des chiffres, et à enserrer la totalité du monde, dans tous ses aspects, à l'intérieur de ce réseau de mesures que le mathématicien appellerait "dimensions". Avons-nous eu raison de voir dans la précision le critère général de ce qui a constitué la Science?

La *mesure* nous vient de l'origine elle-même du chiffre et de l'idée de comptage, c'est-à-dire, comme nous l'enseignent les livres d'arithmétique élémentaire, de la comparaison entre deux séries:

- l'une est la série des nombres, commencée dans la perception 1 doigt, 2 doigts, 3 doigts... (digits), et achevée dans l'intellect: la série numérique 1, 2, 3, 4, 5..., est elle-même enracinée dans l'acte de lever des doigts (digits) pour compter.
- l'autre est un ensemble d'objets qui ont un caractère plus ou moins commun: - un doigt, une pomme, - deux doigts, deux pommes, - trois doigts, trois pommes, etc...

Après 5, il y a un petit problème (comment matérialiser?: prenons les deux mains), et après 10, il y a un autre problème (où trouver des éléments de comparaison incorporés à celui qui compte?: prenons nos pieds). Les Aztèques ont su compter dès l'abord jusqu'à 20 puis sont allés très loin dans la correspondance entre une série d'objets incorporés et des objets externes à l'être, ceux du monde qu'il faut décompter. Alors l'esprit invente le "chiffre". Babylone et les civilisations méso-américaines apprennent à le maitriser, les Mayas puis les Arabes inventent le zéro et la virgule, et tout ceci est déjà tardif dans l'Histoire humaine.

Mais, parallèlement, et indépendamment, émerge la Mesure qui elle aussi s'établit sur la correspondance entre une *conception* interne à l'homme et un *stimulus* qui lui est externe, ce que le philosophe va appeler la "grandeur", dont le psychologue montre qu'elle fait partie des jugements immédiats avec lesquels l'être traite le spectacle du monde: dans la longueur, il y a le pas, le coude (cubit) ou le pied (foot), voire le pouce (inch), dans le volume, il y a la poignée, il y a la brassée, etc... Mesurer, c'est donc comparer un étalon empirique commun (à peu près) à tous les hommes, à une chose du monde extérieur; c'est ramener par la pensée les choses du monde à l'homme, c'est ramener le monde à l'homme, et il n'y a plus qu'à franchir les étapes de l'infini du monde qui donnaient le vertige à Pascal. Apparaissent les techniques de mesure diverses qui s'appliquent à toutes les

''grandeurs'' que l'on est capable d'imaginer et, parcourant un chemin laborieusement tracé, vont poursuivre ce chemin par des prolongements analytiques successifs et illimités. A travers les grands chiffres et leurs logarithmes, elles vont maitriser par la pensée l'infinité du monde (le diamètre de l'Univers), pour l'inclure dans le fini (''ça se mesure''...), et plonger par les décimales dans l' ''infiniment'' petit, aussi jusqu'à ses limites subatomiques.

Tout ceci est très evident; c'est ce que nous enseigne l'Histoire de la Mesure comme un fragment de l'Histoire des cultures. Mais dans cette conquête illimitée du monde par la métrologie est apparu un souci de *normalisation* – qui est un souci *d'universalité* –: le *système* métrique, et un détachement progressif des références qui étaient incorporées à l'être (built in), pour passer aux unités les plus abstraites (la révolution du Mètre, le multiple de la longueur d'onde de la raie rouge du cadmium). Et, surtout un *vertige de la précision*, très sensible à la fin du XIXe siècle, s'empare de cette fraction de l'humanité qui est lancée dans la conquête métrologique du monde: ''les savants''. Apparaît alors le danger de *confondre la mesure et la chose* et de croire qu'il suffit de mesurer pour maîtriser, de connaître pour faire, d'expliquer pour comprendre, en subissant dans cette conquête d'une connaissance – qui est incontestable – le mirage de la précision.

S'il en était ainsi, la ''jauge'' de la culture humaine serait la précision même avec laquelle l'homme enserre le monde dans des séries de dimensions toujours plus nombreuses: c'est la domination de la mesure qui entraîne ce vertige de la précision. C'est à ce vertige que les premiers conquérants en ce domaine, ceux qu'on appelle maintenant les Physiciens, une caste noble et remarquable du monde scientifique, ont, semble-t-il, succombé: la Mesure est Bonne et la Précision est meilleure, tel est l'axiome qu'ils ont subtilment introduit dans nos esprits. Leur succès, qui justifiait pragmatiquement leur méthode, a été tel que ces idées se sont transformés en une Idéologie. C'est contre cette idéologie que nous proposons un concept: les Sciences de l'IMPRÉCIS.

De l'idéologie de la précision

Elle s'est introduite par des voies subtiles: si ce qui est précis est bon, très bon, excellent, le meilleur, *par conséquent*, ce qui est imprécis est grossier, mauvais, très mauvais. Une équivalence qu'aucune épistémologie ne justifie, sous-tend le monde de la connaissance, l'imprécis n'est plus simplement le contraire du précis, l'imprécis c'est le Mal car le précis c'est le Bien, et donc tout ce qui est imprécis n'est pas digne de la pensée. C'est, plus simplement, une connotation qu'effectue la part irrationnelle de notre esprit en l'appliquant à la partie qui se veut rationnelle, et cette connotation est en soi source d'erreur. Le difficile usage de notre rationalité, toute neuve à l'échelle de l'évolution – 2000 ans sur des millions d'années –, se trouve en lui-même corrompu par une irrationalité qui s'insinue subrepticement dans notre raison, par le canal de notre science.

Nous passons d'une constatation: ''la précision, c'est bien'', à une idéologie: ''n'est bien que ce qui est précis''. Le terme même d' ''imprécision'' est une dérivation *négative* de l'idée de précision, une dérivation qui apparaît nécessairement comme fachêuse. Or il se trouve que si la pensée de la Physique (''Natural Philosophy'' de Lord Rayleigh), nous a imposé une réalisation achevée relative au monde observable qui est autour de nous, ce monde de la Physique, comme Science de la Nature, n'est pas le seul monde qui occupe notre esprit, et pour lequel nous devions nous efforcer de faire usage de notre scientificité. Rappelons-nous que les ''théories'' sont des échafaudages qu'on refait ou qu'on perfectionne pour construire l'édifice scientifique à partir d'une *intuition de structure* et que quand celui-ci est terminé dans son *évidence*, les échafaudages sont enlevés et

disparaissent dans le rayonnement de l'édifice achevé.

Dans notre vie quotidienne, dans notre environnement, se situent des objets qui n'appartiennent guère au monde de la Nature: ce sont des formes et des couleurs, ce sont d'autres hommes, ce sont des impressions dans notre champ de conscience et toutes celles-ci se répètent ou se renouvellent selon des lois que nous avons longtemps ignorées mais dont aucun d'entre nous, - y compris physiciens, astronomes et biologistes - ne peut refuser l'évidence ni le caractère général.

Science prédictive et marges d'incertitude

La Science, c'est ce que l'on connaît, ce que l'on sait de façon telle que l'on puisse en prédire l'ensemble en fonction des parties, le devenir en fonction de l'état. Or le concept de "prédiction", bien élucidé par les statisticiens, implique essentiellement le concept d' "erreur" relative, c'est-à-dire une imperfection, toujours provisoire, dans la prédiction. La science est la prédiction d'une configuration, d'un arrangement des choses du monde. Quand elle parle d'autres "choses" que d'êtres de l'esprit (mathématiques) la science ne se construit que sur la base de la reconnaissance à la fois d'une *prédictivité* des êtres et des formes dans leur devenir: l'idée de Loi, et en même temps, des *limites* de cette énonciation: ce qu'on appellera *l'erreur statistique*, que celle-ci s'applique à l'observation de ce qui est ou à la prédiction de ce qui devra être.

La Physique, au sens le plus général du terme, celui d'Aristote qui a perduré jusqu'au XVIIIe siècle: de "science de la Nature", reconnaît, dans son essence même, l'existence de limites à ce qu'elle peut affirmer. C'est précisément sa grande force: comme nous le disait Poincaré au siècle dernier, la Valeur de la Science c'est sa capacité de ne jamais avoir à revenir sur une affirmation ou une loi établie dans les limites des hypothèses qu'elle s'est fixée, dans les limites des observations qu'elle peut faire, et dans les limites à la prédiction qu'elle veut faire. C'est un système cumulatif et progressif, c'est l'exercice d'une pensée qui connaît ses limites et qui incorpore celles-ci dans tous ses énoncés.

Et certes, cette affirmation d'erreur ou d'écart dans l'observation et dans l'expérience, cette affirmation de tolérance nécessaire ou d'imprécision minimale, qui sont fondamentalement conjointes à l'affirmation de la chose scientifique elle-même: la loi, la propriété, la forme, cette affirmation est souvent passée sous silence, tellement elle est axiomatique.

Le grand public entre autres, dans sa vie quotidienne en tout cas, manifeste une imperméabilité si totale à l'essence de la pensée scientifique, qu'il est apte à s'abuser sur la notion même de vérité scientifique avec sa relativité essentielle, et à la dichotomiser - voire la scotomiser - en un Vrai "absolu" et un Faux tout aussi absolu, méconnaissant par là le fondement de la raison. En fait, nous restons près des anciens âges, et nous regrettons toujours le beau rêve d'une magie opérationnelle des recettes qui réussissent toujours, alors que ce qui nous est proposé, ce sont les recettes qui, dans les limites de l'hypothèse et de la précision, réussiront avec un haut degré de probabilité.

En bref, nous oublions souvent les restrictions au moment des affirmations. Ce "Nous" implique tout le monde, non seulement ceux qu'on appelle "le grand public", mais aussi, plus subtilement, ceux mêmes qui fabriquent cette science: les ouvriers de la vérité, et qui Maîtres d'eux-mêmes et de leur pensée dans *leur* domaine propre, retombent dans le reste de leur vie, à l'état, moins glorieux, d'êtres à la raison vacillante et aux affirmations péremptoires car il est tellement plus commode de vivre dans des certitudes fausses que dans des incertitudes mesurées.

De la rationalité de l'être

L'être humain n'est pas (pas encore, ou jamais?) un être rationnel, et la raison ne suffit pas à rendre compte de la totalité des faits et des actes de notre vie. Le comportement de l'homme est un mélange en proportions variées, de ce que nous appelons la *pensée rationnelle* (les Psychologues l'appelleraient volontiers pensée sémantique, dénotative, déductive, logique), et ces *impulsions irrationnelles*, ou plutôt en dehors du rationnel, qui sont des mouvements de l'esprit et de la sensibilité, induits par des facteurs subtils relevant de la circonstance – souvent un de ces facteurs est la brièveté du temps imparti à l'exercice de la pensée. Pour penser, il faut du temps – suivant la situation, le décor, les valeurs, or nous devons, souvent, penser vite, et nous sacrifions sur la qualité de notre pensée. A ce moment on pourrait penser que le domaine de l'irrationalité serait simplement et toujours provisoirement celui de l'inexpliqué. De Palma en cite des exemples en économie: stratégies mixtes, shopping expérimental probablement passeront bien plus tard à un stade meilleur de rationalisation, mais ici et maintenant nous n'en savons rien et continuons à les pratiquer.

Il n'y a pas à cet égard de différence fondamentale entre l'homme dit "de science" et l'homme tout court; peut-être des différences de degré, et surtout des différences de situation. Tout orgueil péremptoire de celui qui se qualifie (ou que ceux qui l'entourent qualifient) d'"homme de science" – sous-entendu des "sciences de la Nature": le physicien, le chimiste, etc. –, orgueil par lequel ce dernier se prétendrait plus "rationnel" ou plus "sérieux" que les autres, n'est qu'une suffisance, facilement acceptée, il faut le dire, par un grand nombre de gens, y compris par ceux qui sont victimes de cette suffisance.

Une image utile: un cerveau à deux étages

Les plus récentes découvertes neuro-physiologiques suggèrent fortement que notre cerveau – dont la comparaison avec un ordinateur n'est plus à faire (elle a déjà suffisamment été faite) – se présente comme un assemblage de deux grandes parties dont chacune se trouve en charge de deux aspects différents de la réactivité de l'être aux stimulis du monde extérieur.

Il y a ce que des spécialistes ont appelé le *"paléo-encéphale"* (Gastaud, Osgood, Leroi Gourhan), ce cerveau profond, héritage d'une longue marche de l'évolution, qui se trouve assumer la tâche de traiter avec una grande rapidité les stimulis du monde extérieur pour les transformer en réactions, afin de *faire face* rapidement, quasi-instantanément à notre échelle, aux situations nouvelles d'une façon adéquate. Cette part du cerveau est chargée de la perception immédiate, des réactions brèves, et largement conditionnées, c'est-à-dire préparées par l'expérience passée, elle a donc la vertu d'être rapide, une vertu qui est compensée par sa limitation à des schémas réactionnels pauvres, routiniers, sommaires; et qui surtout ne ressortissent guère de ce que nous appelons la Raison. Pourrait-on la dénommer "cerveau de l'irrationnel", du sensible, du vécu? Il parait vraisemblable que dans sa continuité avec le système nerveux corporel, ce cerveau ait été l'outil adéquat pour la réaction des ancètres de l'homme aux situations que lui proposait la nature: ce qu'on appelle maintenant le paysage d'action. Mais ce cerveau est lui-même très peu capable de *déduction*, car la déduction implique syllogismes en chaînes, chaîne de pensées séquentielle et elle exige du temps et de l'attention pour parcourir cette chaîne sans se tromper en propageant la causalité au cours des propositions. Cette part de la

"mentalité" (Ashby) est donc très mal armée dans les environnements complexes où les liaisons entre les éléments du paysage d'actions sont à la fois subtiles et multiples et où chaque détail doit propager son effet sur l'ensemble pour entraîner une conclusion.

A cette part du cerveau s'adjoint chez l'homme une autre partie, connectée, mais distincte, de la précédente, qui constitue, nous disent encore les physiologistes, la partie externe, superficielle, un volume relativement faible de ce que nos anatomistes appellent le "cerveau" en général, partie récente dans l'évolution génétique, partie que l'on peut avec une certaine ironie qualifier de "cerveau de luxe", elle se trouve responsable largement de ces interconnexions particulières entre éléments de pensée qui respectent des règles inflexibles, des principes de non contradiction, précisément tout ce que nous appelons la "rationalité", elle est responsable entre autres des classifications analytiques que nous faisons des êtres de la perception, des critères d'attribution de valeurs, en bref, de la connaissance scientifique, au sens d'un système universel de non-contradiction. C'est bien la pensée rationnelle au sens conventionnel du terme. Et ce qui caractériserait cette région de l'ordinateur de l'esprit, ce serait, conformément à des règles que l'informatique connait déjà, un *délai de traitement* bien plus considérable, en tout cas par rapport aux réflexes et au conditionnement tout préparé du cerveau précédent.

Tard venu dans l'évolution, ce "cerveau de luxe" se présente comme l'outil qui permet de pénétrer et de subsister dans ce monde *complexe* de la science que nous avons *construit*, mais dont la richesse dépasse chacun de nous, en faisant passer progressivement sensations et perceptions au crible de l'analyse des valeurs, de l'accumulation mémorielle, de la résolution des dilemmes, et de la terminologie (un mot pour chaque chose). A bien des égards, il n'est pas abusif alors, de considérer que la Logique elle-même est un *luxe de la pensée*, un luxe couteux en bits/sec., un luxe que nous ne pouvons nous permettre que si nous disposons de temps pour "réflechir", c'est-à-dire pour revenir sur nos impressions fugitives en vue d'en dégager les aspects communs.

Quelle que soit la validité scientifique de cette image, elle nous aide à penser, ne serait-ce que pour la contredire; c'est ce qui justifie de l'esquisser ici. En fait, des gens aussi divers que les biologistes, les psychologues, les informaticiens, et les publicitaires – ces redoutables expérimentateurs de ce qu'on peut appeler la communication brève et l'art de la conviction – concourent pour confirmer cette image, un peu schématique, de deux parts de la machine humaine à traiter l'information: l'une, à faible débit mais à grande vitesse, qui traite – grossièrement mais rapidement – les aspects les plus rudimentaires de l'information sensorielle, l'autre, dérivée et branchée sur la précédente, qui traite avec soin, avec logique mais avec des délais substantiels, les mêmes données en les épurant, les élaborant, les rationalisant, c'est-à-dire en éliminant leurs contradictions internes vis-à-vis d'une "Logique Universelle" dont les bases seraient de l'ordre de ce qu'on appelle, d'une autre façon, la pensée mathématique.

Vers un statut des sciences de l'Incertain?

Une Science de l'Imprécis n'a du point de vue de l'épistémologie aucun statut défavorable: ce n'est rien de plus qu'une autre branche de la Science globale.

(1) Une très grande part de l'activité mentale de l'être se trouve devoir traiter, de gré ou de force, d'aspects imprécis du monde, d'objets fluctuants et variables, de valeurs comportant un très fort degré d'erreurs;

(2) Ces éléments nécessaires à la pensée, possèdent des régularités approximatives, une prédictibilité vague, mais sans pourtant et pour autant être jamais totalement aléatoires;

donc qu'ils sont de jure *objets de science*. Il n'y a de science que du général et la science a pour objet des régularités observables du monde extérieur;

(3) Une bonne part des aspects des sciences de l'imprécis – mais pas tous – ressortissent de ce qu'on a appelé les *sciences humaines*, ces sciences dont l'homme est l'objet, une formule critiquée par G. Gusdorf ("Ce ne sont pas des sciences et l'Homme n'est pas un objet"), mais que nous reprendrons à notre compte, laissant au lecteur le soin de nous contredire;

(4) Que peut être une épistémologie des sciences de l'imprécis, c'est-à-dire quelles sont les lois auxquelles elle obéissent et les prédictions auxquelles leurs objets peuvent donner lieu;

(5) Une "méthodologie", un recueil et une tactique, des méthodes permettant de les traiter. Beaucoup de ces méthodes ne sont rien d'autre que nous ne connaissions déjà. Certaines nous sont connues parce qu'elles sont, de fait, largement appliquées dans les sciences sociales, pas toujours, à notre avis, dans l'esprit qui convient à des sciences de l'imprécis, mais plutôt dans l'esprit de vouloir singer les sciences du précis pour des raisons de conformisme intellectuel.

Quelques unes de ces méthodes (échelles ordinales, équations symboliques, approche phénomenologique) apparaissent bien comme des attitudes universelles de la pensée dans son traitement d'objets vagues. Il ne paraît pas outrecuidant d'estimer que les sciences du vague ne sont qu'une étape génétique de la pensée scientifique tout court: il n'y a qu'une seule Science, ou plutôt un seul esprit scientifique dont les modalités d'exercice changent avec les caractères épistémologiques propres des objets qu'elles étudient.

Ces méthodes seraient ici le rendu de la dette qu'ont contracté les sciences sociales, en imitant (peut-être inconsidérément, mais c'était plus facile), les sciences de la Nature, lors de leur constitution au début du siècle. Car enfin le *monde des phénomènes* (ce qui "apparaît" à l'esprit prolongé par des outils d'observation) excède infiniment *l'univers du précis* qui a facilité l'observation de choses constantes et stables, la chute des corps, les lois de l'électricité. Et on pourrait se demander si dans leur extravagant succès les sciences de la Nature dure n'auraient pas de quelque façon trahi la Science elle-même en se cantonnant aux domaines du précis et en négligeant l'Imprécis par des motivations de confort et de conformisme intellectuel. Devrait-on soupçonner la majestueuse Physique d'avoir négligé tout un pan de l'activité de connaissance en se servant de l'excuse souvent entendue, rarement écrite: "tout ça n'est pas précis, ça n'existe pas."

Un portrait partiel et partial de la réalité fait par les Sciences de la Nature?

En fait, notre science contemporaine dominée par les sciences bio-physico-chimiques et les techniques qu'elles ont suscitées, apparait soumise à la dictature de la métrologie dans une interprétation un peu totalitaire de la célèbre phrase de Lord Kelvin: "c'est une bien pauvre science que celle qui ne peut s'exprimer par des mesures".

L'idée même de mesure, de confrontation entre le concept de grandeur tel que l'homme le ressent et tel qu'il le rencontre dans l'objet qu'il mesure, est certes un des éléments essentiels d'une pensée scientifique: si pour le physicien il est d'abord la mise en place objectivable d'une grandeur, et par là-même de tolérance dans l'observation ou la domination de l'objet qui a cette grandeur, il est, pour celui qui rencontre le phénomène in statu nascendi – à l'état naissant – et qui va créer l'idée nouvelle, *d'abord* une manière de se distancer de ce qu'il rencontre, d'en neutraliser l'aspect affectif, et par là d'en retrouver l'étrangeté créatrice.

C'est bien souvent cette fonction qu'accomplit la mesure: c'est le cas du jeune chercheur en vacance d'inspiration, le jeune chien du laboratoire qui ne sait par où prendre son problème, auquel, si souvent, le patron suggère, pour "voir venir", de "mesurer" tous les aspects mesurables du phénomène avec lequel il se confronte, avec l'espoir vague – mais souvent réalisé –, qu'au bout du compte émergera de cette familiarisation distancée une *idée* permettant de prendre à parti le phénomène de manière opérationnelle.

C'est aussi la première et la plus importante des fonctions de la mesure dans les sciences du vague – fréquemment les sciences humaines –, toutes entachées de subjectivisme, d'a priori, et d'arbitraire; *l'hygiène* de la mesure et de la recherche de méthodes de mesure, dérive l'esprit du chercheur à la fois vers ce qu'on peut appeler les aspects sensibles d'un phénomène (ce qui se mesure le plus aisément), et vers une remise en perspective de ce dernier dans un tableau qui neutralise les composantes affectives, qui situe l' "étrangeté" ("comment cela est-il possible..."), cette étrangeté dont nous savons qu'elle est le ferment de l'esprit d'innovation. On retrouve ici de façon sous-jacente cette distanciation forcée qui au lieu de "parler" de la chose en termes généraux, se complait à en décrire les aspects par le jeu des "grandeurs".

Mais la dictature méthodologique de la Mesure, qu'il n'est pas possible de contester dans une science expérimentale, entraîne, nous venons de le voir, de redoutables séquelles: de la mesure comme méthode, on passe à la mesure comme frénésie, et de la frénésie de la mesure, à la frénésie de la précision (la "frénésie du rationnel" n'est pas en soi une rationalité). Par là, on sort, disions-nous, de la Science pour passer à l'Idéologie et aux jugements de valeur sur les connaissances du monde extérieur que nous pouvons avoir.

Un historique des sciences de l'Imprécis?

Il serait trop facile de remarquer que *toutes* les sciences, ont commencé dans l'Imprécis, qu'elles ont commencé par la conscience vague de régularités dans l'observation du monde extérieur et plus particulièrement de la Nature qui se présente a priori comme l'objet neutre et impassible auquel notre esprit se confronte. Faire un historique de l'imprécis, ne peut avoir de sens que dans la mesure même où la Science ne fait pas seulement un effort héroique et désespéré pour tirer sa connaissance des marécages du vague, ce qui est son mouvement le plus général, mais où elle accepte que la connaissance possède des territoires de marécages, de flou et de nébuleux, que ceux-ci sont importants en soi à la connaissance générale de l'univers, et qu'il faut bien les prendre en compte.

A cet égard, les premières analyses de la création intellectuelle, ont donné les bases de ce qu'on appelle maintenant Heuristique ou Science de la découverte, vue comme une branche d'une psychologie appliquée à la production d'innovation. Elles ont du, dès l'abord, reconnaître que l'esprit humain est flou dans son fonctionnement, ambigu dans ses concepts, vague dans ses définitions. Le Roy rappelait, il n'y a pas si longtemps, que "l'invention s'accomplit dans le nuageux, dans l'obscur, l'inintelligible, presque dans le contradictoire. C'est dans ces régions de crépuscule et de rêve que naît la certitude. Un souci malencontreux de rigueur et de précision stérilise plus surement que n'importe quel manque de méthode" *(Logique de l'invention)*. Et le fait est que la lutte pénible de l'esprit pour réduire ce nuage, éclairant comme avec un projecteur le concept vague pour en délimiter nettement les contours, est une des démarches les plus immédiates et les plus essentielles de la pensée qui se veut rationnelle. C'est une espèce de quête de l'intelligible qui constitue l'invention, et spécialement à l'époque de la constitution de la Physique ou de la Chimie dont nous disions plus haut qu'elles ont eu tendance à constituer les canons

de notre pensée scientifique, proposant un modèle impératif de la rationalité.

Nous sommes devenus plus modestes: c'est le calcul des probabilités, issu de Bernoulli, de Pascal, et de Laplace, trois excellents representants de ce que les français appellent le cartésianisme, qui est venu peu à peu saper, puis détruire, la croyance dogmatique à une perfectibilité illimitée, synonyme d'une précision également illimitée. Reconnaître tout d'abord que les formes de notre esprit et de notre observation ne sont jamais que des moyennes statistiques, des Gestalts plus ou moins indépendantes des éléments qui les constituent, des constances à l'intérieur des variations, reconnaître en d'autres termes de l'aléatoire de l'univers, dans sa structure "atomique" au sens étymologique a été un effort considérable: l'idée, bien popularisée par Eddington, que ma table ou une brique se présenteraient, avec un microscope qui serait suffisament puissant, comme des essaims de moucherons de formes rectangulaires où chaque atome se meut dans un mouvement brownien autour d'une vague position de quasi–équilibre. C'est cette structure probabiliste du monde matériel qui est venue occulter – voire détruire – la confiance dans l'idée que les formes de notre entendement ne sont que des "rectifications" par la pensée des formes de notre observation.

A grande échelle, le monde est incertain, non pas à cause des incertitudes de principe sub–atomiques des particules élémentaires, mais à cause de nos incapacités multiples de le cerner. Ce sont ces dernières qui nous concernent sous le nom d'"Imprécision", notre incapacité toujours provisoire et toujours renouvelée d'assurer les conditions d'une expérimentation correcte au sens des sciences "exactes".

Ainsi au titre d'une Science de l'Imprécis ici revendiqué, ce plongeon dans l'infiniment petit – qui est certes un des grands achèvements intellectuels du siècle –, n'a guère apporté. C'est dans la recherche de la précision qu'on a rencontré l'imprécision, elle était un mal nécessaire, et s'est transformée en une nécessité du mal (Dieu joue–t–il aux dés ou non?) avant de devenir une condition épistémologique qui se restreignant au monde bien particulier de la Microphysique, n'atteignait guère le monde du quotidien. Ce dernier monde n'a guère changé et ce qui le change, c'est plutôt l'émergence laborieuse des sciences humaines ou des sciences de l'esprit, suggérées par Wolff, un disciple de Leibniz, qui a fourni matière à la reconnaissance des régularités à grande échelle, à notre échelle, celle de l'Homme mesure de toutes choses, celle de la Psychologie et de la Sociologie, dont l'idée même ne date que du milieu du XIXe siècle.

Une adolescence nécessaire de chaque discipline scientifique?

On peut penser – au titre d'hypothèse d'une génétique de la science –, que les sciences de la Nature ont effectué leur gestation dans ces siècles d'erreurs fécondes et de contradictions répétées qui ont précédé les Encyclopédistes et Lavoisier, qu'il y aurait une sorte d'état d'adolescence nécessaire pour qu'une science trouve ses méthodes et ses concepts à partir du moment où l'idée même, qu'un domaine scientifique particulier émerge comme possible, et se traduit dans ses premiers résultats. Ainsi chaque discipline aurait sa période de latence, d'essais et erreurs. Cette période serait passée pour les sciences de la nature. Elle serait en cours pour les sciences de l'esprit. Cette idée recouvre de façon plus précise celle de sciences "jeunes" opposées aux sciences "vieilles": la Physique serait plus ancienne que les sciences de l'esprit. En fait la conception même qu'il puisse exister des sciences de l'Homme, ne remonte guère, pour des raisons qui nous échappent (provisoirement), au–delà des idées de Wolff et du laborieux travail de Wundt, de Helmholz à Iéna, de Galton à Londres. En conséquence, serait justifiée la remarque,

souvent faite, que l'épistémologie des sciences humaines n'existe pas encore, qu'il faut d'abord que celles-ci se constituent en disciplines vraiment autonomes, et non pas tributaires de celles qui les ont précédées, avant que leur contribution à l'épistémologie générale puisse correctement s'établir. Le terme même de sciences de l'Imprécis serait-il ici un élément de cette genèse?

Sur le plan conceptuel, c'est le calcul des probabilités d'une part, comme prise en compte rigoureuse des répétitions et de l'émergence de la stabilité statistique à travers des variations des composants, c'est, d'autre part, la théorie de la Gestalt, pur produit de la pensée dialectique allemande, qui affirme que le tout est différent de la somme de ses parties et qu'il est indépendant de la nature même de ces parties, qu'il est donc justiciable d'une connaîssance en soi-même, qui constituent les contributions notables à ce que l'ont peut appeler le domaine des sciences de l'imprécis. Ils fournissent les éléments d'une réconciliation entre la logique et la probabilité, entre l'analyse des complexions et l'existence de leurs éléments.

Une contribution importante à ce que nous appelons Sciences de l'Imprécis, est l'émergence de la notion de "logique du probable" de Hans Reichenbach. Celui-ci dans un travail maintenant classique, a souligné que la valeur "Vérité" prise généralement par les logiciens comme binaire (V/F), peut très bien être remplacée dans le jugement des propositions par une valeur intermédiaire entre 0 et 1 qui est en fait, ce qu'on pourrait appeler la probabilité subjective d'une "vérité", et entre autres, sa capacité de se propager du début d'une chaîne de syllogismes à sa fin. Dans ces conditions, les chaînes de syllogismes, chacune affectée d'une valeur V/F, ne peuvent se suivre les unes les autres que dans la mesure où les probabilités composées en chaque maillon de la chaîne, dépassent nettement une valeur (par exemple 1/2), en deça de cette valeur, la chaîne du raisonnement se rompt d'elle-même car elle perd tout intérêt prédictif: il y a plus de chances que le "raisonnement" basé sur la chaîne soit faux, que n'en aurait une prédiction au hasard faite relativement à son dernier maillon. La logique de probabilité de Reichenbach apporte un certain type d'instrument pour manipuler des concepts dont la vérité est loin d'être catégorique, soit qu'ils sont mal établis sur le plan expérimental, ou que les définitions des mots eux-mêmes sont trop floues pour entrainer l'adhésion de la raison.

Science de l'Incertain, du vague, ou du flou: l'émergence d'un domaine

Nous avons introduit le terme de "Sciences de l'Imprécis" dans notre ouvrage *La création scientifique* paru en 1957, en même temps que le concept *d'Infra-logique* . Peu de temps après, O. Zadeh, un autre chercheur en Physique mathématique, alors à l'Université de Columbia, a défini dans un article de 1965 souvent repris, le terme de "fuzzy concepts" (concepts flous), dans une acception très proche de l'idée de concepts imprécis telle que nous l'avions définie: des ensembles (fuzzy sets) qui, tout en restant parfaitement opérationnels au niveau de la pensée et de la création, possèdent des définitions assez vagues, et qu'il n'est guère utile de préciser abusivement, car une définition trop étroite en évacue la valeur heuristique, et en fait, les vide de leur contenu.

Nous citions à l'époque le concept de "métal", si clair aux yeux du technicien et du chimiste, voire du quincailler ou de l'apothicaire, mais que le progrès de la physique a conduit à définir ultérieurement à partir de certaines types de répartitions d'électrons sur les différentes couches externes autour du noyau. Cette idée simple et forte introduisait alors des cas limites et des exceptions si nombreux et si divers, que le terme même de métal y

perd sa valeur opérationnelle au niveau de la pratique expérimentale. Ainsi dans cette opération de l'esprit basée sur la recherche d'une définition rigoureuse, la pensée scientifique globale a certainement gagné, mais en même temps, les faciles critères de propriété qui étaient attribués à l'opposition métal/métalloide, se "désagrègent", l'image s'affaiblit dans l'esprit du chercheur: pour certains c'est un bénéfice théorique, pour d'autres, peut-être les plus nombreux, il y a perte de prégnance et d'aptitude combinatoire de la pensée créatrice.

Vers 1973, un ouvrage d'Arnold Kaufmann sur les sous ensembles flous a eu un grand retentissement dans les milieux de l'heuristique et de la créativité appliquée. L'idée de "fuzzy information" et ses différents aspects a été reprise avec Zadeh dans un ouvrage de 1983.

Le problème de l'émergence de formes à partir d'un agregat d'éléments qui sont les items d'une possible classification, et de la stratégie de l'esprit vis à vis de ces agregats (concepts d'attracteurs et de nuées dynamiques) a été l'objet de plusieurs analyses en raison de son importance dans la théorie des classifications par exemple des photographies d'êtres ou d'objets appartenant à un monde "incomprehensible": formes géographiques, vues de satellites d'observation, ou vues microscopiques. Comment y faire jaillir des gestalts, des formes pregnantes pour l'esprit auxquels on donnera un nom et qui seront de futurs universaux pour ce monde sans balises?

Le concept de similarité est un exemple de grandeur imprécise et prégnante. Nous l'avons étudié en effectuant des matrices de similarité vers 1967. Il a été repris et développé brillamment par les chercheurs des Laboratoires Bell, en particulier Sheppard et Kruskal qui en ont fait une méthode pratique d'emergence de phénomènes flous (1970).

Certes, le mouvement de la science évolue vers une terminologie aussi liée que possible à des critères précis et sans ambiguité, mais il n'est pas discutable non plus que le processus même de l'invention se fasse par une lutte permanente avec des concepts flous qu'il faut que l'esprit juxtapose et, pour ainsi dire, aménage. On ne s'étonnera donc pas que des concepts imprécis soient générateurs d'innovation dans une mécanique de la pensée qui est de nature conflictuelle: "un abus de rigueur stérilise plus surement la découverte que n'importe quel manque de méthode". Il n'est pas douteux que la "vertu épistémologique" soit liée fondamentalement à l'effort de traquer le vague et le flou du monde pour en faire une image dépourvue d'ambiguité, mais il est aussi sûr que la pensée créatrice, dans ses mécanismes fondamentaux qui commencent à être maintenant bien étudiés sous le nom d'"Heuristique", fonctione selon des règles bien différentes de celles de la preuve formelle: nous le savons bien – les mathématiciens nous le disent eux mêmes à l'envi –, on trouve d'abord, on démontre ensuite, et certes, la démonstration peut mettre en évidence le fait que ce qu'on a "trouvé" était faux, auquel cas on recommence, sinon ad infinitum, au moins jusqu'à parvenir à la conviction d'erreur, et ce n'est qu'alors qu'on abandonne en recueillant les sous – produits concrets d'une errance infructueuse. Mais de toute façon, la démarche essentielle de la découverte ne coincide guère – ou pas du tout – avec les caractères de la chose trouvée.

On doit ici se rappeler la discussion contemporaine sur le rôle de l'ordinateur dans la création, régie par le célèbre théorème de Gödel, et la distinction très pragmatique entre l'invention que nous avons appelé ailleurs *"variationnelle"*, dans laquelle l'intelligence artificielle de l'ordinateur se substitue plus ou moins aisément à l'esprit humain pour explorer le champ des combinaisons possibles de concepts et de combinaisons suggérées par le psychologue ou le méthodologiste, et l'invention *absolue* (transcendentale) – celle qui précisément ne peut se résumer á une simple combinatoire d'éléments ou de

syllogismes déjà trouvés. L'invention absolue, elle, procéderait d'un mécanisme d'oppositions et de contradictions, puis de résolution de ces contradictions par étapes, avec la naissance d'autres contradictions, etc... qui rappellent quelque peu les errances du rat dans le labyrinthe, en l'espèce, celles de l'esprit dans les champs de possibles limités par les barrières à la fois imaginaires et durement ressenties de ce qu'on peut appeler les "contraintes perçues".

Deux aspects de la "science": science établie et science en train de se faire

Pour situer ici le domaine de ce que nous appellerons les sciences de l'Imprécis, il n'est pas inutile d'accentuer le contraste que nous avions largement souligné dans des ouvrages antérieurs, entre deux acceptions du mot "science", indissolublement liées dans la culture:

– la science établie, à chaque instant de l'évolution culturelle, nous disposons d'un corpus de connaissances qui augmentent d'instant en instant, et qui est constitué de tout l'ensemble de toutes les publications scientifiques – c'est-à-dire contrôlées selon les règles de la logique et de la méthode expérimentale – virtuellement accumulées dans une sorte de "bibliothèque universelle", et qu'on pourrait comparer a une sorte de gigantesque mur des livres et des publications. Ce mur des livres aurait pour caractères qu'il possède une cohérence totale par rapport aux règles de la logique, que rien n'y est contradictoire à rien: les lois de la biologie sont des commentaires aux lois de la chimie et ne leur sont pas contradictoires, même si leur style est bien différent, et les lois de la chimie ne sont pas contradictoires à celles de la physique nucléaire, même si elles parlent d'autres particules. Ce mur des livres est totalement accessible à tous, nul ne peut s'en voir dénier l'accès, et plus, tout chercheur scientifique doit, en théorie, en dominer la totalité, même si sa capacité d'information est ici et maintenant provisoirement limitée;

– la science en train de se faire se situe alors en contraste avec le mur des livres de la science achevée (à l'instant t). C'est bien plutôt un champ des possibles tout découpé et circonscrit, à chaque instant, par toute une série de murs séparant le possible de l'impossible, le conceptuellement concevable ("vrai") du conceptuellement inconcevable ("faux"), le vrai et le faux n'y sont jamais éternels, ils sont l'illusion permanente de chaque chercheur, ils diffèrent largement tout en gardant des traits communs d'un chercheur à un autre, les murs de l'impossible mental changent avec chacun, et le chercheur se saisit lui-même comme un point – le rat dans le labyrinthe des possibles – dans une attitude d'exploration des canaux du possible. Il erre d'un lieu à un autre selon une trajectoire plus ou moins complexe, plus ou moins aventureuse, et parcourt des distances plus ou moins grandes depuis son point de départ qui est l'origine de sa réflexion. Ce chercheur possède une culture scientifique, c'est-à-dire une vision plus ou moins approximative, plus ou moins étendue de ce mur des livres de la science achevée, vision généralement située dans une zone de ce mur (spécialisation), et qu'il projette sur le plan où il circule mentalement sous la forme de ces contraintes logiques ou culturelles dont il connaît et ressent l'existence à l'intérieur du mouvement de son propre esprit. Le champ des possibles lui apparaît comme un labyrinthe qu'il cherche à parcourir, et par là, à comprendre. Sa vision est limitée, limitée en étendue, car sa capacité de traiter des connaissances est elle-même limitée, elle est aussi limitée en accuité, les contours des murs et des couloirs sont vagues, flous, et ne se voient mieux que quand il s'en rapproche, pour donner lieu à d'autres perspectives et d'autres contours. La découverte scientifique n'est autre qu'un trajet établi dans le champ des possibles, un trajet vérifié, avec des indicateurs et des balisages, un trajet transmissible à d'autres que lui-même par la voie de la

publication scientifique qui va s'insérer dans une étape ultérieure dans le mur des livres, le mur vertical de la science établie, en augmentant la densité et l'extension.

Ce qui nous intéresse donc ici, c'est essentiellement cet aspect flou et incertain, ce qui ne veut pas dire nécessairement faux, de ce labyrinthe des possibles où il n'est jamais certain que les cloisons qui limitent le trajet du chercheur soient rigides, non plus qu'il ne soit certain qu'il n'existe pas un trajet subtil et lointain qui permet de contourner les murs de l'impossible présent. Ici, comme nous l'ébauchions plus haut, le vague, l'imprécis, le nuageux, ne sont pas des obstacles rédhibitoires, mais des données du champ de conscience du chercheur, et leur "fausseté" reste toujours à montrer ou à démontrer.

L'opposition accusée entre ces deux concepts recouverte par le même mot de science, est une opposition de mentalités et d'attitudes, la circulation dans le champ de la science établie est toujours "sûre" (si l'on respecte les règles du jeu scientifique, et si l'on est capable de poursuivre la cohérence déductive jusqu'au bout), puisque celui-ci est régi par le balisage de la logique universelle, l'errance, au contraire, dans le champ des provisoirement possibles et provisoirement impossibles, est toujours risquée et conditionnée par l'image que chacun s'en fait, elle est une situation humaine par essence, alors que l'édifice majestueux et immense que celui-ci contemple vaguement et par fragments, est une image essentiellement inhumaine: l'image de la Raison.

Conclusion: inconnaissance pratique ou sciences de l'imprécis?

Il y a deux sortes d'invention très différentes: l'une est l'invention artistique ou scientifique, l'autre est celle du discours et de l'argumentation (Bacon).

1) La puissance de la méthode scientifique dans l'accès à une rationalité du monde, est parfaitement établie sans discussion. Elle permet entre autres l'élimination de ce qu'on peut appeler les "fausses connaissances" et les "fausses sciences". Mais elle a ses limites dans le champ pratique de ses applications.

2) Un grand nombre de domaines de ce qu'on appelle "la pensée", ou l'établissement d'un "système du monde", échappent à la connaissance scientifique, et ce ne sont pas nécessairement les seuls domaines de l'illusion et du rêve.

3) L'empire scientifique est limité en microphysique par des indéterminations de principe liées à l'absence d'un déterminisme rigoureux. Celui-ci, selon une analyse souvent répétée, n'interesse guère les aspects du monde à notre échelle, car il laisse à cette échelle la place, par la loi des grands nombres, à un déterminisme statistique: celui avec lequel nous vivons, où la connaissance des phénomènes est limitée d'abord par la précision de nos observations.

4) L'extraction d'un signal à partir du bruit, ou d'une forme d'un fond, donne lieu elle aussi à des principes d'incertitude "appliqués" très généraux, qui se superposent sur le plan théorique, mais souvent à une autre échelle, à celui de Heisenberg au niveau des sciences de l'observation.

5) Par ailleurs, dans toutes les disciplines où l'observé est *réactif* à la présence de l'observateur, il existe des limitations de principe à la connaissance que cet observateur peut avoir du monde, qui se traduisent par des "principes d'incertitudes" des sciences humaines: la photographie en fournit un bon exemple.

6) A ce niveau, l'empire de la connaissance scientifique est limité par un autre ordre de facteurs qui sont de nature humaine ou sociale:

a) Il y a le coût généralisé global de certaines expériences. Le coût est toujours le prélèvement effectué sur des ressources "discrétionnaires" que l'être ou cette société

affectent à la connaissance, selon l'axiome "Knowledge is power". Ces ressources discrétionnaires étant limitées, notre capacité de vérifier et de reproduire certaines expériences à un moment donné de l'évolution de la pensée se trouve aussi limitée. Le système scientifique laisse donc quelquefois la place à l'argument d'autorité selon lequel nous devons croire ce que nous disent les "hommes compétents" sans avoir la possibilité de le contrôler. Le "crédit" du chercheur remplace alors la capacité de vérification.

b) l'expérimentation elle-même est nécessairement limitée dans toutes les sciences du macroscopique par les moyens matériels qu'elle impliquerait: géographie, géologie, astronomie, en sont des exemples: l'expérimentation, interaction forte où l'on touche au monde est alors remplacée selectivement par l'observation qui est une interaction faible.

c) un très grand nombre d'affirmations avec lesquelles nous vivons couramment comme éléments de notre paysage d'action - et que nous désirerions voir fondées scientifiquement - n'ont aucune base rationnelle. Les expériences qu'elles exigeraient pour les vérifier, sont techniquement irréalisables, et spécialement en sciences humaines, car elles reposent sur l'examen et l'analyse d'ensemble d'objets ou d'êtres (échantillons) trop grands pendant des temps trop longs. Ces affirmations n'appartiennent pas au monde scientifique mais elles sont considérées commes telles par la société. Elles sont "pseudo-scientifiques".

d) le pouvoir de l'expérimentation est limité en sciences humaines non pas seulement, comme vu précédemment, par l'ampleur des moyens ou des échantillons à considérer, mais - et il le sera de plus en plus - par des considérations de type déontologique, qui sont d'un autre ordre que la recherche de la vérité par la méthode expérimentale.

Il existe donc des domaines dont il est concevable qu'on puisse les connaître rationnellement, et qu'ils soient importants, mais qui se trouvent *interdits à la recherche expérimentale*. Ces domaines ont des chances de se multiplier dans l'avenir immédiat.

7) dans la plupart des domaines qui sont interdits, de fait ou de droit, à l'expérimentation, la pensée scientifique possède comme autre ressource l'observation, et l'analyse d'un grand nombre d'observations qui ne représentent chacune qu'un niveau d'interactions faible - mais pas nul - entre l'observateur et l'observé. La construction de formes hypothétiques régulières à partir de la statistique et de l'analyse des facteurs de variance des observations est alors le domaine fondamental de la pensée rationnelle. Mais celle-ci ne peut se poursuivre comme un pur système mathématique sans avoir recours dans la création et la formulation de ces hypothèse, à ce que nous avons appelé les sciences du vague, les sciences de l'imprécis, c'est-à-dire à des concepts provisoires mais prégnantes, qui seront reformulés et éliminés ou précisés à une époque ultérieure de la connaissance, en perdant par là d'ailleurs leur valeur heuristique.

8) Vu la valeur pratique de la science dans ses applications, il paraît probable que c'est le développement de la science elle-même, dans sa volonté de recherche, d'expérimentation ou d'observation, qui se trouvera limité par la Société. Ceci se produira à cause des conflits de valeurs émergeant de la nature même de la pensée active, qu'elle soit artistique ou scientifique, quand elle est mise en contact avec un sentiment social global, resté latent ou diffus jusqu'à présent, mais qui commence à saisir les conséquences réelles de la méthode scientifique, ou même de la création artistique, comme premières ébauches (expérience pour voir, combinatoires originales, matrices de découverte) de la démarche scientifique.

Manuel José Lopes da Silva

Dep. Comunicação Social, Fac. Ciências Sociais e Humanas
Univ. Nova de Lisboa, Lisboa (Portugal)

Hierarquia e Complexidade

Resumo

O conceito de Hierarquia, fundamental na Teoria da Complexidade, é analisado a partir do estudo de vários sistemas complexos de natureza físico-matemática, social, política e psicológica.

Examina-se a relação entre os conceitos de Hierarquia e Informação, quer do ponto de vista físico, quer do ponto de vista organizacional.

São sublinhados alguns dos mais importantes aspectos de sistemas muito complexos (sociais) e hipercomplexos (homem).

Abstract

The concept of Hierarchy, which is of basic importance in the Theory of Complexity, is analysed by means of the study of several complex systems of mathematico-physical, social, political and psychological nature.

The relation between the concepts of Hierarchy and Information is examined, from the physical as well as the organizational points of view.

Some of the most important aspects of very complex (social) and of hyper-complex systems (man) are underlined.

A Noção de Complexidade

1. Os físicos atómicos, surgidos nos começos deste século (Max Born, Heisenberg, Weizsäcker), foram dos primeiros investigadores que dedicaram muito do seu tempo a reflectir sobre a complexidade da Natureza. Dispondo de um instrumento conceptual unificador, que consideravam extremamente potente (a teoria atómica...), consideravam que a sua utilização permitiria atingir uma visão unitária e totalizante dos objectos físicos. Como é natural porém, imediatamente identificaram as suas fronteiras do conhecimento, e o diálogo então iniciado com a Biologia ainda não terminou.

As reflexões de Weizsäcker acerca da estrutura dum cristal continuam a ser atraentes, pela tentativa interdisciplinar de relacionamento do seu tipo de estrutura atómica (e molecular), com as propriedades macroscópicas observadas.

Devido à extrema complexidade das estruturas atómicas, com um número excessivamente elevado de elementos e relações, os investigadores destes domínios foram levados a abandonar os métodos matemáticos tradicionais (a física matemática clássica), e tudo o que eles implicavam, ou seja, uma teoria geral da causalidade eficiente, e a adoptar métodos matemáticos de índole estatística, propondo conceitos globais (matrizes) e um

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 303/309
1992, APDC & SPF, Lisboa

princípio de Incerteza que constituem uma primeira aproximação de sistemas muito complexos.

De resto, os mesmos físicos atómicos adoptaram a noção de "escala" para a delimitação dos conjuntos de objectos diversificados que passaram a fazer parte dos modelos científicos.

A escala natural é a nossa escala, a humana, a dos objectos directamente acessíveis aos nossos sentidos; depois usamos as escalas molecular, atómica, nuclear e subnuclear. Em sentido ascendente temos a escala do Sistema Solar, a das Constelações, a Galáctica e a do Universo.

2. Embora os físicos continuem a propôr perspectivas unitárias para integração dos seus conhecimentos, que gradualmente se vão revelando cada vez mais intrigantes, na verdade a unidade físico-matemática tem sido difícil de conseguir, contrariamente ao que esperavam aqueles que tanto se entusiasmaram com a Teoria de Newton.

Recentemente surgiu a Teoria dos Fractais, instrumento extremamente apropriado para a modelização através dos computadores.

Não cabe aqui um desenvolvimento sobre esta teoria, bastando referir que se trata duma modelização de tipo geométrico com base num objecto designado por conjunto de Mandelbrot, considerado por alguns como um dos seres matemáticos mais complicados alguma vez existente. Estas figuras matemáticas, que se designam por fractais, têm a particularidade de se repetirem através de mudanças de escala, mantendo a mesma forma, duma série de iterações computacionais para a seguinte. Há portanto infinitos níveis geracionais, de escalas de amplitude crescente (ou decrescente), mantendo-se a forma dos objectos duns para os outros.

Uma geração contém em si, evidentemente, o programa das gerações seguintes. Poderiamos admitir a existência duma gradação hierárquica de gerações, se o termo "hierarquia" estivesse suficientemente esclarecido neste momento.

3. No entanto, o recente conceito de Complexidade surgiu sobretudo relacionado com as complicadas redes de comunicação dos nossos dias, com enorme número de elementos (nós) e de relações (ligações) entre eles. Estas redes (e também outros sistemas complexos como os programas de computadores) são estudados por "grafos", esquemas que têm a finalidade de mostrar visualmente oa elementos dos sistemas e as relações que os dinamizam.

Os grafos podem tornar-se em geral extremamente complexos, mas alguns deles podem estudar-se através de instrumentos de cálculo apropriados.

Trata-se dos grafos hierarquizáveis, isto é, dos grafos em que todas as setas entre nós têm o mesmo sentido, e não há portanto circuitos de retroacção. Nestas condições dizemos que há uma hierarquia, com níveis superiores e inferiores, definidos pelo sentido das setas.

Como refere Le Moigne, a ordenação de níveis pode ser estabelecida por precedência temporal, tal implicando, manifestamente, uma relação causa efeito de tipo "comando". No entanto o estudo dos fluxos de informação nas organizações mostra que pode haver outra explicação para os desníveis hierárquicos: a da existência de maior informação duns níveis, que é transferida para outros que dela têm necessidade, e que estão a um nível inferior.

Podemos portanto estabelecer, neste momento, que o conceito actual de hierarquia tem que ver com precedências de tempo/espaço e de informação/comando.

Redes de Comunicação

4. A complexidade que resulta para as redes de comunicação do facto de assegurarem comunicações bilaterais entre pontos dispersos, tem solicitado a atenção de inúmeros investigadores. A situação inicial de todos os elementos contactarem directamente com todos os outros, pode ser num primeiro passo ultrapassada através de uma solução simplificadora com comutadores intermédios, proposta pela primeira vez por Charles Clos, dos Bell Laboratories, nos anos 50.

A solução dos Clos permitiria, por exemplo, obter uma rede de 1188 nós perfeitamente equivalente a outra de 1296 nós.

A ideia de base consiste em substituir uma estrutura matricial unitária por uma estrutura subdividida em matrizes e submatrizes.

Tal solução corresponde ao agrupamento dos elementos de base em células, que por sua vez são ligadas a um nível superior, constituindo-se em nível intermédio. Podemos assim dizer que se cria uma hierarquia sistémica de índole espacial.

Todavia há outras possibilidades de criar uma estrutura hierarquizada nos sistemas de comunicações.

Elas foram desenvolvidas partindo da necessidade duma melhor utilização das infraestruturas físicas dos sistemas radioeléctricos existentes.

Trata-se do método conhecido por "multiplexagem temporal", que se baseia na transmissão numérica ou digital dos sinais.

Os vários canais de comunicação são inseridos nos intervalos de tempo uns dos outros. Há uma partilha do tempo de transmissão por todos os canais, até aos segmentos temporais mais reduzidos.

Isto exige que nas vias que encaminham a maior quantidade de comunicação (por exemplo as interurbanas) sejam utilizados os ritmos binários mais elevados, por exemplo, 565 megabits por segundo, e por isso se diz que constituem a 5ª hierarquia da rede.

As hierarquias inferiores utilizam ritmos binários (caudal de informação) mais reduzidos, ou sejam respectivamente 140, 34, 8,448 e 2,048 megabits por segundo.

É perfeitamente evidente que neste caso, à ordem hierárquica está associada a quantidade de informação processada em cada nível.

Redes em Organismos

5. Tem sido exaustivamente estudada a estrutura hierárquica das organizações, devido talvez à antiquíssima existência de instituições sociais que de certo modo têm sobrevalorizado este aspecto das respectivas estruturas.

Uma das mais aliciantes abordagem desta questão foi feita por Arthur Koestler, que se inspirou sobretudo nos organismos vivos.

Ele identifica os vários níveis sistémicos, estabelecendo princípios e leis específicas que se aplicam dentro de cada nível, e que também articulam entre si os diferentes níveis. Para designar as entidades agrupadas nos vários níveis, propõe o termo "holon", com a significação de "parte de um todo".

Os holons estão inseridos em redes de estrutura arborescentes, radicadas nos níveis superiores.

Invocando o apólogo dos dois relojoeiros Hora e Tempus, já proposto por Herbert Simon, ele explica que a única maneira dos escalões superiores tratarem da informação global do sistema, é através da delegação de tratamento nos níveis inferiores que têm de

actuar com uma determinada amplitude de autonomia.

A informação de comando tem o sentido descendente, mas Koestler identifica nos organismos canais de comunicação transversais, entre holons do mesmo nível, de árvores diferentes. Tal informação não é de comando, mas de qualquer modo é uma comunicação de estrutura, por ele designada de reticulação.

Existem também circuitos de retroacção que encaminham para os níveis superiores informações de controle, possibilitando o exercício duma função fundamental nos organismos vivos, a da regulação, tão importante como as funções metabólica, genética e relacional.

O comportamento dos holons deve obedecer a um princípio orgânico básico: harmonização da tensão para a integração determinada pelo nível superior, com a pulsão para a asserção autonómica do nível inferior.

Este comportamento bipolar é inerente ao conceito de ordem hierárquica, e é, de facto, uma caractística universal da vida (Koestler).

Hierarquias de Empresas

6. A estrutura básica duma empresa é também de tipo arborescente. De resto esta forma é estabelecida para corresponder ao princípio fundamental da unidade de comando e de direcção, decisivo para o bom funcionamento de qualquer grupo humano.

Os canais de comando constituem ligações descendentes, existindo também circuitos de rectroacção relativos ao controlo, que não violam o princípio da exclusividade da rede hierárquica de comando.

A delegação de funções, poderes e responsabilidades é feita com o objectivo de aligeirar o trabalho de tratamento da informação corrente e de decisão dos gestores, libertando-os para as análises globais e estratégicas.

A quantidade de informação tratada pelos gestores de topo é maior do que a que é acessível aos gestores intermédios, verificando-se também aqui que aos níveis hierárquicos superiores está associada a maior informação.

Há ainda nas empresas uma rede especial de canais de comunicação não claramente identificável, que veícula a informação informal do tipo convivial, mas que é de toda a conveniência seja considerada pela Alta Direcção.

7. A estrutura arborescente é, nas empresas de hoje, frequentemente completada com uma estrutura reticular do tipo matricial.

Trata-se da constituição de grupos de trabalho para levar a cabo projectos específicos, encabeçados por um chefe pertencente à hierarquia (ao nível dos directores de linha) mas constituídos por elementos dos vários departamentos especializados da estrutura. Cada elemento do grupo fica assim sujeito a dois fluxos de autoridade, um correspondente à autoridade de linha personalizada pelo director do departamento a que pertence, outra correspondente à autoridade de projecto personalizada pelo director do projecto global.

Pode manifestamente haver conflito entre autoridades, mas os inconvenientes daí resultantes são largamente superados pelo dinamismo e flexibilidade com que a empresa passa a responder à mudança, sendo de resto a vida de cada projecto normalmente limitada.

O mais importante numa Empresa é o claro estabelecimento das metas que todos devem procurar atingir. Elas por vezes não podem ser reduzidas a números ou enunciados muitos simples, antes constituem um quadro que há vantagem em ordenar com vista a uma melhor apreensão.

Hierarquia de Fins

8. O problema da hierarquização das finalidades dos sistemas, ou seja o do estabelecimento do seu quadro teleológico, não é diferente do problema da hierarquização dos Valores, proposta por Max Scheler, nem do da hierarquização das normas jurídicas proposta por Kelsen.

Tal como Kelsen remete para uma norma superior a justificação e aprofundamento duma qualquer norma do ordenamento jurídico, tendo por vezes da se chegar a uma norma fundamental, também os fins dos sistemas se podem sempre fundamentar noutros mais gerais situados num nível superior.

Tomando como exemplo o caso dos sistemas de comunicação social, em relação a cujas metas há por vezes alguma confusão, uma perspectiva hierarquizante leva a estabelecer três níveis, assim ordenados:

Nível A, da Missão: integração social harmonizada com a asserção pessoal.

Nível B, da Finalidade: transmissão da herança cultural e promoção do diálogo social.

Nível C, das Políticas: articular a cultura com a informação e o divertimento.

O exame do quadro teleológico anterior permite esboçar uma proposta de quadro de referência para a análise teleológica dos sistemas em geral.

Ao mais alto nível devem colocar-se as metas que têm que ver com a constituição dos próprios sistemas, uma vez que não estando a ser atingida, os sistemas tendem a extinguir-se como sistemas; designaremos este nível por constitutivo.

Depois vêm os fins que apontam para o processo de desenvolvimento dos sistemas, devendo promover o seu progresso e abertura a novos cenários, mas sem perder a sua identidade, e respeitando a experiência adquirida: trata-se do nível evolutivo.

Finalmente, temos os fins directamente relacionados com a actividade própria do sistema, situando-se já a um nível de gestão de meios, e que podemos designar por nível operativo.

Seria interessante, mas não o faremos aqui, verificar a correspondência entre estes níveis e os enunciados correspondentes, existentes nos vários instrumentos jurídicos de vários países.

Seria de esperar que os enunciados mais gerais existissem nas próprias Constituições nacionais, e os enunciados operativos nos Estatutos das empresas, mas tal pode não se verificar devido justamente à falta duma perspectiva sistémica.

Hierarquia da Cognição

9. O desenvolvimento das Ciências e Tecnologias da Cognição (CTC), tem conduzido ao desenvolvimento duma carta polar da cognição em que os níveis exteriores englobam os interiores, significando isso que os exteriores são mais gerais que os interiores.

O núcleo duro das CTC é o "cognitivismo", designação cómoda para uma orientação que considera a inteligência de tal modo próxima do que é intrinsecamente um computador, que a cognição pode ser definida pela computação de representações simbólicas. Ao cognitivismo está associada a ideia do tratamento em série de símbolos, segundo regras preestabelecidas.

A Inteligência Artificial (IA) é a projecção literal da hipótese cognitiva. Ainda que este domínio de investigação tenha começado a introduzir técnicas de tratamento paralelo, na realidade este tipo de processamento é sobretudo utilizado numa orientação da investigação mais rica que a anterior, e que designamos por "conexionismo".

Nesta orientação de estudo, o ponto de partida é o reconhecimento da emergência de estados globais numa rede de componentes simples. Não há necessidade duma Unidade Central de Processamento, antes se verifica poderem regras locais gerar operações individuais, e regras de mudança gerar ligações entre elementos. Assim, as propriedades emergentes (e a estrutura resultante) podem ser identificadas com uma faculdade cognitiva.

Mas a orientação da investigação mais completa (e mais difícil...), tem ambições que se radicam nas Teorias de Alternativa e no recurso ao senso comum. Designamo-la por "enaccionismo", e no seu âmbito a cognição é a acção produtiva: historial do acoplamento estrutural que enacciona (faz-emergir) um mundo.

Esta enacção recorre a uma rede de elementos inter-ligados, capaz de sofrer modificações estruturais ao longo dum historial não interrompido.

No âmbito da enacção, a inteligência não se define como a faculdade de "resolver" um problema, mas como a de "penetrar" um mundo partilhado.

O ponto de vista da enacção insiste sobre a existência duma codeterminação do sistema cognitivo por constrangimentos exteriores e por uma actividade gerada interiormente, o que o demarca de qualquer forma de construtivismo ou de neo-Kantismo biológico.

Caminhar dos níveis superiores para os inferiores (centripetamente) é passar da emergência ao simbólico, excluindo a base donde emergem os símbolos e trabalhando com o seu valor nominal; ou passar da enacção ao conexionismo pressupondo regularidades predefinidas no domínio de operação do Sistema.

Na direcção oposta, dos níveis inferiores para os superiores, exclui-se progressivamente o que parece estável e regular, passando a privilegiar o estudo das origens dessas regularidades, incluindo as dimensões preceptuais do nosso mundo humano.

Considerações Finais

10. Uma análise global leva-nos a associar aos níveis hierárquicos a ideia de quantidade de informação processada. Poderíamos até dizer que a cada nó de junção de dois elementos dum nível inferior, corresponde a poupança de tratamento da informação de 1 bit para nível superior. Mas será talvez forçar excessivamente a analogia com a teoria da informação, o tentar aplicar tal ideia aos Sistemas em geral.

Na realidade, a orientação actual da Teoria da Empresa por exemplo, parece ser no sentido de suprimir níveis hierárquicos, como transmissores da informação de linha, uma vez que os sistemas informáticos o permitem fazer melhor e mais rapidamente.

A empresa do futuro seria assim mais um sistema de estrutura celular (no sentido que referimos anteriormente), do que de estrutura arborescente.

Em outros sistemas sociais surgem situações semelhantes: parece que o princípio da unidade de comando e direcção por todo o lado é posto em causa.

A analogia com a teoria da informação confirma-se todavia nos sistemas físicos. O desenvolvimento dos novos sistemas de comunicação exige a instalação de canais de interligação de grande capacidade, portanto de hierarquias de alto nível. E essa é, sem dúvida, também uma tendência dos nossos dias.

E assim como no parágrafo anterior verificávamos o homem a organizar-se em pequenas células fechadas sobre si próprias (quase autónomas), perdendo por assim dizer o sentido de co-responsabilização no âmbito geral da empresa, também aqui verificamos que ele disporá de cada vez maior quantidade de informação, permitindo-lhe uma actuação cada vez mais diversificada.

É uma situação que nos parece de grande ambiguidade, e qualquer analogia com as

antigas formas de produção artesanal é certamente forçada e gratuita.

11. Reflexões recentes levam a uma nova revalorização epistemológica do conceito de hierarquização como operação fundamental do conhecimento, associada às operações de disjunções/união, centralização/descentralização (Edgar Morin).

A perspectiva sistémica revaloriza o aspecto auto-eco-organizacional dos sistemas muito complexos, levando mesmo à suspeita de que a sua capacidade auto referencial nos sistemas de nível superior possa produzir uma misteriosa qualidade designada por "consciência de si".

Não se trata, no entanto, de identificar esta qualidade com o sujeito filosófico, tal como emerge, por exemplo, de Kierkegaard.

De facto, uma reflexão sobre as operações do cérebro humano permite identificar fenómenos verdadeiramente espantosos de muito alta complexidade, e a adoptar uma nova noção capital para considerar o problema humano: a hipercomplexidade.

12. Paralelamente, reconheceu-se um aspecto patológico do pensamento moderno, que introduz uma hipersimplificação no conhecimento tornando-o cego à complexidade do real.

Devem-se evitar os doutrinismos e os dogmatismos, que fecham as teorias sobre si próprias e as petrificam.

As propostas oriundas duma teoria da complexidade generalizada (envolvendo princípios como o de "Unitas Multiplex") não constituem um novo paradigma, mas contribuem certamente para uma revisão dos actualmente existentes.

Carlos Fiolhais

Dep. de Física, Fac. de Ciências
Univ. de Coimbra, Coimbra (Portugal)

Física, Computação e a "Inteligência" da Natureza

Resumo

Discutem-se as relações da Física com a Computação, sublinhando o facto de as leis físicas poderem ser vistas como "software" e a matéria como "hardware" de um gigantesco computador. Comparam-se as perspectivas do teórico e do experimentalista sobre os processos de cálculo que ocorrem na Natureza. É abordado o problema do carácter contínuo ou descontínuo do espaço-tempo e são enunciados os limites físicos fundamentais da actividade de computação.

Abstract

The relationships between Physics and Computation are discussed, emphazising that the physical laws may be seen as software and matter as hardware of a giant computer. The theoretical and experimental viewpoints of the computational processes which occur in Nature are compared. The problem of the space-time being continuous or discontinuous is referred to and the fundamental physical limits of computation are enumerated.

As diferentes épocas históricas costumam criar as suas metáforas e a nossa não está imune a esse fenómeno de retórica sazonal. É indubitável que a metáfora informática está hoje muito em voga. Neste caso, porém, talvez se trate mais do que um simples exercício de estilo, que apenas serve para caracterizar o discurso. Uma metáfora útil torna-se, quando repetida, uma verdade profunda. Os cientistas invocam hoje muitas vezes o computador como analogia, nos mais variados domínios. Até os físicos, que não são excessivamente dados a liberdades de linguagem, falam do computador de uma maneira paradigmática. Assim, já houve um físico, Heinz Pagels, que chamou às leis da física o "código cósmico". Do seu ponto de vista, as leis da física não seriam mais do que o conjunto de "software" que anima o "hardware" (partículas, campos) do universo. Pagels disse mesmo que: "O cérebro, o tempo atmosférico, o sistema solar, mesmo as partículas quânticas são todos eles computadores. Claro que não parecem computadores mas estão de facto a computar as consequências das leis da natureza. Na perspectiva computacional, as leis da natureza são algoritmos que governam a evolução do sistema no tempo, tal como fazem os programas de computadores. Por exemplo, os planetas, ao moverem-se em torno do Sol, estão a efectuar computações análogicas das leis de Newton".

O mundo é pois comparável a um grande computador analógico, a uma enorme máquina de cálculo. Pode ser visto como um imenso autómato celular, um tabuleiro de células cujos estados são transformados sucessivamente noutros, tal como no "Jogo da Vida" do matemático inglês J. Conway.

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 310/316
1992, APDC & SPF, Lisboa

Portanto, mais do que pôr a ênfase na inteligência artificial, que ele porventura ajudou a criar com as tecnologias do silício, ao físico interessa realçar a inteligência natural da natureza. O autómato pode ser (parece que é) uma boa metáfora para o funcionamento do cérebro humano, mas é de certeza uma metáfora ainda melhor para o funcionamento do mundo. Não temos tanto de nos admirar com o facto de sermos inteligentes, mas mais de admirar, como seres inteligentes, a inteligência de uma natureza regida por um código uniforme e universal.

O homem começou por compreender esse código prevendo certos fenómenos naturais com a ajuda de algoritmos matemáticos. Verificou-se que o "programa" que descreve o comportamento de uma maçã e da Lua é o mesmo, e que tanto maçã como Lua estão a efectuar o mesmo "cálculo", exemplificam o mesmo algoritmo. O planeta Neptuno foi descoberto como aplicação do algoritmo (lei) da gravitação universal; foi tudo realizado no papel pois não existiam ao tempo dispositivos de cálculo automático adequados. Neptuno estava também a realizar o mesmo "cálculo" que a maçã e que a Lua. Um planeta, Neptuno no caso, olha para os outros, "calcula" o inverso do quadrado das distâncias e ajusta a sua aceleração em conformidade... Um físico olha para os outros planetas, calcula o inverso do quadrado das distâncias e diz onde está Neptuno. A natureza é pois inteligível, tem uma "inteligência" que a nossa só a espaços e parcialmente consegue vislumbrar. Conseguiu-o Newton, quando fundou a física, conseguiu-o Einstein, quando reformulou a doutrina da gravitação universal.

O físico pode portanto ser denominado de "utente" do computador da natureza. Confronta-se no entanto com objectivos completamente diferentes dos de um vulgar utente de um computador artificial. Um utilizador informático dispõe de uma máquina criada com um certo objectivo, uma máquina congeminada para desempenhar um certo conjunto de tarefas. O interior dessa máquina é, em princípio, bem conhecido. O físico está, porém, diante de uma "máquina" que ele não criou, cujo objectivo e/ou sentido lhe escapam, e cujo interior está longe de ser conhecido. Os resultados perante certas entradas são muitas vezes completamente inesperados: algumas das grandes descobertas deste século, como a mecânica quântica, revelaram-se uma necessidade experimental - uma imposição da natureza - não sendo o resultado de nenhum esquema lógico profundo e preliminar nem de nenhum senso comum porventura equidistribuído. Pelo contrário: há quem teime que a mecânica quântica é inconciliável quer com a lógica quer com o senso comum. Como disse Bohr: "Quem não achar a mecânica quântica estranha não a compreendeu verdadeiramente".

O utente normal da informática, que se senta por exemplo diante de um microcomputador, procura comparar os dados entrados e saídos; conhece a entrada e pretende conhecer a saída. Não lhe interessa o interior da máquina. Em contraste, o físico analisa a entrada e a saída de um processo para perceber o "programa interno", para compreender a mensagem ROM ("Read Only Memory"), que a natureza traz incorporada. Trata-se de ROM e não de RAM, para utilizar a terminologia moderna, porque ele não pode alterar "o código memorizado" e fazer, por exemplo, o electrão ter outra carga ou as estrelas andarem de outra maneira, que não seguindo a regra da razão inversa do quadrado da distância como a maçã, a Lua e Neptuno.

A diferença entre um físico e um utilizador de computador é portanto mais uma diferença de intenção do que uma diferença de procedimento. Diria até que, no que respeita aos procedimentos, não há diferença absolutamente nenhuma.

Hoje em dia, as experiencias físicas são em geral extraordinariamente complicadas. Não estamos mais no tempo em que Galileu observava o isocronismo das pequenas oscilações do candeeiro da catedral de Pisa para deduzir as leis do pêndulo. Nem no tempo

em que para encontrar uma nova estrela e ficar famoso bastava olhar para o céu com uma luneta rudimentar. As grandes experiências modernas exigem equipamento sofisticado e são, total ou parcialmente, computadorizadas. Em física de altas energias, por exemplo, a análise das colisões de partículas nucleares em aceleradores gigantes seria impossível sem os poderosos meios de processamento de informação actualmente disponíveis. Assim, o bosão Z, descoberto no CERN em 1983, não passa de um pico num gráfico de computador. A natureza é ligada a um sistema de processamento de informação para revelar um dos seus segredos, a bola do jogo da interacção fraca. Nas ciências planetárias, para dar outro exemplo, utilizam-se sondas artificiais, que são enviadas para os espaços. Mas não basta a sonda olhar para os outros planetas e decidir, em conformidade, qual deve ser o seu caminho: são os computadores da NASA que escolhem ou permitem a escolha do melhor caminho e fazem desencadear as acções necessárias a esse desiderato.

Mesmo os dispositivos experimentais que não utilizam computadores, no sentido usual do termo, podem ser designados por computadores porque, no fundo, fazem corresponder a uma certa entrada uma certa saída: trata-se portanto de um processo de cálculo, embora um processo de cálculo com um fim muito específico e particular. São computadores analógicos e não computadores digitais.

E os físicos teóricos? Cada vez mais, a actividade destes se prende com a formulação e utilização de modelos computacionais, que descrevem com maior ou menor precisão a realidade. Podem simular, por exemplo, o processo de colisão de um protão com um antiprotão sem ter necessidade do acelerador dispendioso. Usando algoritmos apropriados (i. e. semelhantes ou iguais aos algoritmos da natureza), procuram obter o pico que é o bilhete de identidade de uma partícula Z. Simulam – outro exemplo – a aproximação de uma nave de um planeta exterior no sistema solar, podendo, se a nave cair por engano, voltar atrás e recomeçar tudo de novo. Podem, por conseguinte, ser mais levianos que os seus colegas experimentais. Dizia P. Kapitza, físico soviético experimental: "perguntem a um teórico e façam o contrário". Os físicos experimentais podem não levar demasiadamente a sério as locubrações dos seus parceiros teóricos, porque elas são repetíveis e modificáveis. Modificam-se à medida que se aguça a arte e o engenho de programação de um computador digital. Os físicos experimentais têm de levar dramaticamente a sério as suas montagens laboratoriais, porque elas são computadores analógicos particulares e onerosos. Os teóricos dispõem de muitos mundos para experimentar, mundos de imaginação de entre os quais um – o que afinal interessa – está mais próximo do mundo exterior, enquanto que o experimentalista só dispõe de um, o "difícil", "miserável" mas único mundo que habitamos.

Os experimentalistas pretendem saber quais são os constituintes e fenómenos fundamentais, por interrogação directa, por vezes violenta, da natureza. Querem saber como é o mundo e "apertam-no" até ele se render e confessar a verdade, toda a verdade e nada mais que a verdade. Os físicos teóricos, por sua vez, averiguam as consequências de determinados processos, suaves, de modelação para saber se a natureza é como é por alguma razão especial de entre as várias alternativas possíveis. Para usar uma imagem teológica, que não deve ser interpretada literalmente, pretendem saber se Deus teve alguma escolha na criação do mundo, se por exemplo o espaço-tempo em vez de 4 podia ter um outro número qualquer de dimensões (como 3,99).

Tem contudo de se dizer que a dicotomia entre teóricos e experimentais, que se deu neste século e que foi fruto da procura de um nicho ecológico numa sociedade científica extremamente competitiva, não faz hoje, na era informática, demasiado sentido. Quer uns quer outros utilizam autómatos para produzir conhecimentos, significados e comunicações

sobre o mundo. Ambos utilizam "tecnologias do artificial" para compreender o natural. O computador é um instrumento que está a permitir plataformas de interdisciplinaridade, anteriormente insuspeitadas, e pontes de entendimento, outrora julgadas impossíveis, entre teóricos e experimentais. Usando a linguagem das ciências de computação, não é difícil ver que estão bastante próximos uns dos outros. Uns usam o "Grande Computador Natural" directamente, carregando muitas vezes nas teclas às apalpadelas. Os outros utilizam, num processo de tentativa e erro, computadores artificiais, para tentar imitar o "Grande" ou descobrir o que ele porventura tem de inimitável. Os primeiros, para poupar tempo e dinheiro, perguntam aos segundos. Os segundos, para poupar dinheiro e tempo, recorrem aos primeiros. A sua sinergia traz vantagens mútuas, nomeadamente quando a experimentação é difícil ou impossível ou quando a teoria é mais "wishful thinking" do que conjunto de relações causais bem estabelecidas.

Mas os físicos não estão isolados. Existem químicos, biológos, engenheiros, etc.. Os físicos têm não apenas de se entender entre si mas também de se entender com os outros. Num tempo dominado pelos instrumentos de cálculo, os intercâmbios horizontais estão a substituir as antigas hierarquias verticais. As simulações computacionais, que hoje são comuns tanto em física como em química, biologia, engenharia, não são teoria tradicional nem experiência tradicional, mas um casamento em comunhão de bens, tanto herdados como adquiridos, de uma e outra.

Quais são hoje os bens adquiridos? Podemos sumariá-los recorrendo à designação, boa à falta de melhor, de "ciências da complexidade". Uma característica do "miserável" e "difícil" mundo que habitamos é o extremo grau de complexidade da maioria dos seus fenómenos. O grande número de constituintes elementares (ou supostamente elementares) da natureza e a sobreposição dos seus vários modos de interacção dá origem a processos extremamente complicados de discernir e interpretar. Ora esta complexidade pode residir mais na saída do que na entrada ou no programa interno. Os avanços das tecnologias de informação permitiram-nos verificar directamente que algoritmos simples podem dar origem a resultados extremamente complexos. Essa diferença qualitativa deve-se à não-linearidade do algoritmo. A saída é, nesses casos, extremamente dependente da entrada, pelo que falamos de sensibilidade às condições iniciais. Muitos fenómenos da natureza evidenciam complexidade, apesar do programa ser eventualmente simples.

Actualmente, está-se a descobrir que o sistema mais determinista e periódico de todos – o sistema solar – é caótico e imprevisível (a longo prazo, evidentemente), embora a lei da gravitação seja bastante simples e não seja difícil especificar com alguma precisão as presentes efemérides astronómicas. Os físicos – esses profetas do comportamento da natureza – arriscam-se portanto a ser falsos profetas, se persistem em efectuar profecias a longo prazo. Falham estrondosamente na missão para a qual se julgavam mais vocacionados – saber o que vai acontecer a seguir – porque não dispõem de meios seguros de previsão à distância. A qualidade das suas profecias tem a ver com as possibilidades e limites do cálculo.

Ao falar de imprevisibilidade, é oportuno referir, ainda que só de passagem, o problema da imprevisibilidade do pensamento e acção humanos. Uma máquina que passe o teste de Turing é aquela que consegue imitar na perfeição o comportamento de um ser humano. Não existe, mas não há qualquer razão de fundo para que não exista. Julgo que a previsão da impossibilidade de uma máquina destas pode falhar e o melhor é esperar e ver. Não há qualquer razão para que não se tentem construir aproximações melhores ou piores a máquinas que imitem a inteligência e a acção humanas. Se o computador substitui partes da natureza, não há qualquer motivo a priori para não poder substituir essa parte da

natureza que é o homem. O programa da chamada "inteligência artificial" é decerto bastante ambicioso, mas não é utópico. Ao contrário de outros físicos, não desvalorizo por isso a "inteligência artificial" como projecto, embora esse domínio do saber tenha prometido mais do que aquilo que, até à data, foi capaz de mostrar (o fenómeno não é novo pois já aconteceu nos anos sessenta com a energia nuclear). Julgo que o teste decisivo a uma máquina inteligente será colocá-la debaixo de uma macieira e ver se ela é capaz de fazer como Newton, i.e. criar a física, compreender a "inteligência" da natureza a partir da observação da maçã e da Lua. Sei porém que esse teste não se vai realizar nos tempos mais próximos. No longo processo para se conseguir essa máquina, passar-se-á provavelmente pela abolição progressiva da distinção entre artificial e natural, com a substituição do silício por componentes biológicas (mas os pormenores das técnicas não é garantido que funcionem satisfatoriamente!).

Existem no entanto diferenças, pelo menos aparentes, entre a natureza e um computador digital. Vejamos a mais importante.

Um computador digital trabalha com quantidades discretas. A lógica do computador assenta numa base sim-não, ligado-desligado. A natureza, pelo contrário, é normalmente considerada pelos físicos como contínua. A lógica quantitativa parece à primeira vista mais rica do que a simples lógica disjuntiva do tudo ou nada. Mas trata-se de uma questão de escola: achamos preferível sempre a lógica na qual fomos ensinados. Assim, tanto a mecânica clássica como a quântica traduzem-se matematicamente por equações diferenciais num contínuo espacio-temporal. Para resolver essas equações num computador digital tem de se discretizar o problema, i.e. assumir que o espaço e o tempo não são contínuos mas sim constituídos por um conjunto de pontos que formam uma grelha. Mas será que o espaço e o tempo são mesmo contínuos? Ninguém sabe. Nunca ninguém viu um infinitésimo (a não ser obviamente os matemáticos, que acreditam em tudo, ou pelo menos em tudo o que inventam). Hoje sabe-se da teoria quântica que existe uma distância mínima e um tempo mínimo observáveis (esse limiar microscópico do espaço-tempo chama-se "barreira de Planck"). O princípio da incerteza de Heisenberg impede que se possa saber algo de concreto para além desse espaço e desse tempo. Termina aí a nossa física.

Pensa-se que os problemas dos infinitos em teorias de campo (um dos maiores quebra-cabeças em física) possam ter a ver com a estrutura contínua do espaço-tempo. Feynman, profundo conhecedor dessas teorias , acha muito estranho que seja possível um conhecimento infinito numa região espaço-tempo. O autómato universal que é o mundo executaria uma infinidade de cálculos num espaço limitado.

Talvez as leis físicas actuais devam por isso ser substituídas por esquemas recursivos discretos e o contínuo espacio-temporal que consideramos hoje seja uma abstracção, uma abstracção útil, quando muito, mas apenas uma abstracção. Alguns físicos, mais ousados, propuseram já leis da física discretas e, dada a natureza aproximada, para não dizer grosseira, das nossas medidas actuais, não podemos ainda confrontar as leis contínuas com as discretas e decidir por umas ou por outras. Talvez até nunca seja possível uma decisão desse tipo. Existem argumentos estéticos para um e para outro lado, mas nem sempre verdade e estética estão do mesmo lado. Existem argumentos filosóficos para um e para outro lado, mas boa filosofia nem sempre é boa física. A boa física é conhecedora das suas limitações.

As ciências da computação já estão a mudar a metodologia, para não dizer o conteúdo, da matemática, colocando a ênfase nas estruturas discretas e nos procedimentos recursivos. A física continua tradicionalmente atada ao velho cálculo e às quantidades infinitesimais, que têm algo do ideal platónico. Não é já tempo, agora que as simulações computacionais se estão a democratizar, de pôr a física mais de acordo com essas práticas correntes? Ou

será antes aconselhável pôr as ciências da computação mais de acordo com a física? No primeiro caso, as leis físicas seriam algoritmos de diferenças finitas, enquanto no segundo caso as ciências da informação deveriam incidir mais no contínuo e no analógico.

Seja o espaço-tempo contínuo ou discreto, pode-se dizer que até aqui os computadores têm ajudado a física mais do que a física tem ajudado os computadores (ajudou, bem entendido, com a invenção do transistor, das tecnologias VLSI, dos supercondutores, etc. : tudo isso são sistemas físicos). Mas é chegada talvez a hora de a física contribuir decisivamente para as ciências de computação. Um dos domínios onde se pode dar esse contributo reside no estudo dos limites da computação e dos computadores que a materializam.

Os cientistas de computação falam de autómatos universais e de processos de computação abstractos. Mas, em última análise, qualquer processo de computação exige um suporte material (quer seja um ábaco, uma calculadora de bolso, um supercomputador, uma rede de computadores, um cérebro humano). Como o suporte material obedece inexoravelmente às leis da física, convém averiguar quais são as consequências para as ciências da computação de alguns limites fundamentais da física. Essas restrições da física exprimem-se por desigualdades e são essencialmente três:

1) A mais antiga surgiu na física macroscópica. Dá pelo nome de segunda lei da termodinâmica. Esta foi para Eddington e é ainda para grande número de cientistas a lei mais importante de toda a física e o obstáculo mais intransponível à observação humana. A partir da análise quantitativa do rendimento de máquinas, descobriu-se que existe uma limite à nossa informação sobre o mundo. Num sistema isolado a entropia (medida da falta de informação) não pode diminuir. Expressa de uma outra maneira, a segunda lei diz que é impossivel o chamado demónio de Maxwell, uma hipotética criatura que conseguiria, manobrando o movimento de moléculas, colocar em contacto um sistema quente e um sistema frio, ficando o quente mais quente e o frio mais frio! Num sistema que contivesse um demónio de Maxwell, a informação aumentaria, apesar de o sistema estar aparentemente isolado: ficariamos a saber, de uma maneira cada vez mais nítida, onde era frio e onde era quente. Mas essas criaturas terríveis nem existem nem podem existir, proibidas como estão pela famosa segunda lei! É curioso notar que a segunda lei da termodinâmica é hoje compreendida recorrendo às ciências de computação. Os físicos falam de informação e da falta dela. Dizem que é fatal a existência de fluxos de calor sempre que se apaga informação num registo. Por isso é que tal demónio é impossível: ele teria de registar o estado das moléculas e actualizar esses registos, o que tem custos termodinâmicos. A informação nova só se obtem com o auxílio de proventos energéticos da vizinhança. Neste contexto, assumem particular relevância os trabalhos de L. Brillouin, autor de uma obra sobre "Física e Teoria da Informação", C. Shannon, autor da moderna teoria da informação e, mais recentemente, C. Bennett e R. Landauer, dos laboratórios IBM, que estudaram a relação da termodinâmica com as ciências da computação. Quem diria que a termodinâmica desenvolvida com a máquina a vapor iria progredir com as máquinas de silício?

Vejamos um breve resumo de alguns dos avanços conseguidos desse esforço de cooperação. É sabido que a computação é um processo normal e naturalmente irreversível. Por exemplo, a partir do resultado de uma soma não se pode reconstruir o valor das parcelas (estas, em circunstâncias usuais, são apagadas da memória da calculadora). Landauer mostrou porém que se podiam conceber e construir máquinas lógicas reversíveis. Essas máquinas não produzem perdas de calor, desde que operadas a velocidade suficientemente pequena. A possibilidade de uma computação reversível faz com que os computadores artificiais se aproximem do grande computador natural, que é em princípio reversível (embora haja o problema da violação da reversabilidade do tempo em alguns

processos atómicos, pensa-se que a assimetria que se constata à nossa volta é essencialmente o resultado de condições iniciais assimétricas e não de leis assimétricas).

Podemos resumir a restrição termodinâmica dizendo que a computação exige energia. Quando não exige muita, é preciso muito tempo e portanto muita paciência.

2) A segunda limitação provém do estudo dos fenómenos electromagnéticos e da teoria da relatividade. De acordo com a teoria da relatividade restrita, existe uma velocidade máxima para a propagação de sinais, que é precisamente a velocidade da luz. As consequências para a engenharia de computadores são evidentes: se temos pressa, convém construir máquinas muito pequenas, tão pequenas quanto possível, para a informação demorar o mínimo de tempo entre as várias componentes. A máquina da natureza é grande (grande, digamos, quando comparada com um átomo) e por isso nela coexistem vários tempos e vários processos de cálculo. Mais do que um autómato com um único processador central, podemos dizer que o universo é uma rede de autómatos, ligados à velocidade da luz.

Em suma, a existência de um limite para a propagação de informação impede os processos de cálculo instantâneos. A computação exige tempo.

3) Para diminuir o tempo de cálculo, é conveniente que os autómatos sejam o mais pequenos possíveis. A terceira restrição tem a ver com o princípio da incerteza de Heisenberg, que reina no mundo microscópico: os fenómenos à escala atómica são imprevisíveis. As nossas máquinas de calcular, embora pequenas, não são tão pequenas como isso. Os constituintes da memória não atingem as proporções atómicas. Feynman conjecturou que seria possível colocar informação em átomos e pode-se especular sobre a possibilidade de pôr informação nos núcleos atómicos. É óbvio porém que estamos limitados a um tamanho mínimo (o das partículas elementares) para suporte material da informação e, à medida que descemos a esse tamanho mínimo, cresce a insegurança sobre a informação armazenada.

Concluindo: não existem máquinas infinitamente pequenas, nem máquinas seguras muito pequenas. A computação exige espaço.

Existem portanto limites teóricos bem definidos para as máquinas de calcular mas deve dizer-se também que estamos ainda muito longe dos limites teóricos. A tecnologia de computadores não atingiu ainda nem atingirá tão cedo os limites da física. "Quem não admirará os progressos do próximo século!"

Permitam-me que faça, para terminar, um exercício de prospectiva. Deve ser uma boa profecia dizer que, no próximo século e nos seguintes, o universo vai continuar sujeito às leis da física, evoluindo tal como um autómato celular. A vida em geral e a vida inteligente em particular são uma parte do universo (uma parte desse processo de calculo). Julgo ser também uma boa previsão afirmar que a inteligência natural do homem vai continuar a decifrar a inteligência natural da natureza, procurando entender o "código cósmico" de que falei no início. Mas parece-me evidente que o nosso "hardware", que ao longo dos tempos tem sofrido um processo de evolução, vai continuar sujeito a modificações. Não está provado que aquilo que chamamos de inteligência humana necessite do suporte biológico actual. Há até (o físico F. Dyson, de Princeton) quem faça especulações, com algum fundamento, sobre o futuro da vida a muito, muito longo prazo num universo aberto. Talvez nessa altura a vida inteligente, i.e. o processamento de informação de acordo com certas exigências, se realize materializada em poeira interestelar ou revista uma outra forma exótica qualquer.

Fica a questão, suponho que inteligente, de saber se o universo, que chamei de inteligente, existirá sem inteligências para o compreender e admirar.

Basilio Catania

CSELT (Centro Studi e Laboratori Telecomunicazioni),
Torino (Italy)

The Physics of the Boolean Observer

Abstract

Interaction of human beings, as well as of other entities in the biosphere and of most modern artifacts, with the environment, is performed through organs (sensors or actuators) made of bistable devices, hence is based on boolean algebra, i.e. on the game of "yes or no".

However, in addition to this "analytic mode" of behaviour, human beings seem also to possess a less understood, but very powerful, "synthetic mode", not based on distinguishability and verbalization (in other words not on numbers and words) but on a sort of union with the Chaos. From this back door of our mind seem to come artistic inspirations, as well as scientific intuitions or other unconventional modes of "information transfer under sensory shielding", such as telepathy and precognition.

A mathematical expression of the wave function of quantum physics with complex exponent, containing a pair of quantities representing "information" and "action" respectively, allows to grasp an understanding of the real and the imaginary world of the observer.

Resumo

A interacção dos seres humanos, bem como de outras entidades da biosfera e dos artefactos mais modernos, com o seu meio ambiente é efectivada mediante orgãos (sensores ou agentes) formados segundo dispositivos de quadro duplo, uma vez que se baseia na algebra booleana, i.e. no jogo de "sim ou não".

No entanto, adicionalmente a este "modo analítico" de comportamento, os seres humanos parecem possuir um "modo sintético", menos compreendido mas muito potente, que se baseia não na diferenciabilidade e na verbalização (noutros termos, não em números e palavras) mas sim numa como que união com o Caos. Por esta porta trazeira do nosso espírito parecem chegar as inspirações artisticas, assim como as intuições científicas ou outros modos não convencionais de "transferência de informação sob cobertura sensível", tais a telepatia e a pré-cognição.

Uma expressão matemática da função ondulatória da física quântica com expoente complexo, contendo um par de grandezas que representem respectivamente "informação" e "acção", permite-nos alcançar uma compreensão dos mundos real e imaginário do observador.

1. Introduction

There are two doors through which we acquire knowledge: one is from the outer world, the other is from the inner world. This latter door, however, is harder to penetrate and requires particular attitude as well as long training, which we seldom do, to listen to the "inner voices".

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 317/326
1992, APDC & SPF, Lisboa

The door towards the outside world is made of our (five) senses: therefore it works on distinguishability, i.e. on whether a bistable device in any of our senses is either excited or at rest. Think, for instance, to the retinal molecule in the human eye, that switches from its "11-cis" to the "trans" configuration, when it is hit by a photon. In other words, the acquisition of knowledge through the door of senses rests on the pillars of boolean algebra, i.e. on the game of "yes or no" devices, not too diferent from those used in many types of digital machines that human beings have devised to be helped in their daily activities.

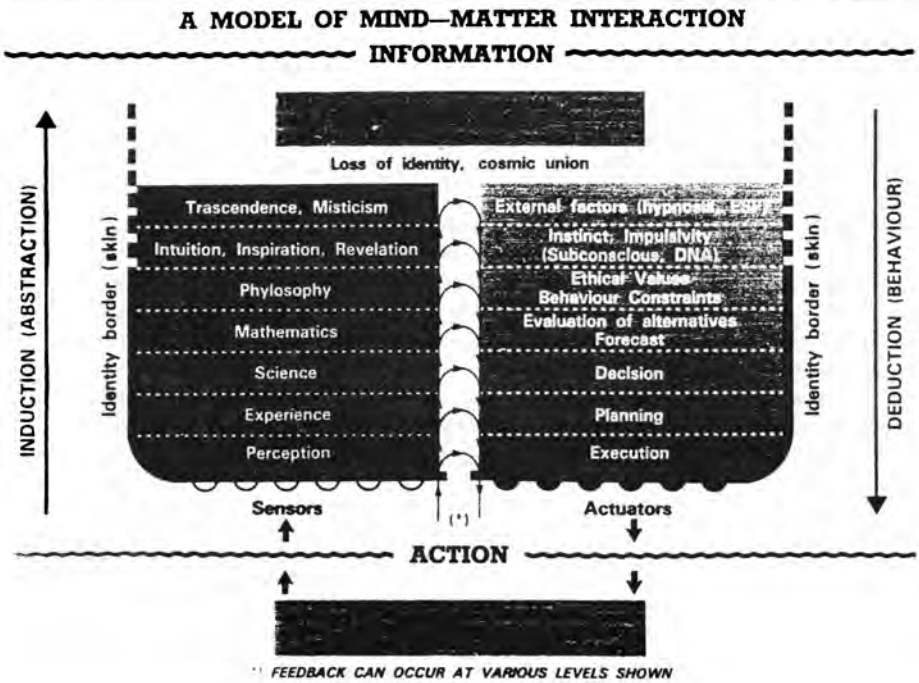


Fig. 1.

The acquisition of knowledge through this outside door proceeds as follows (see left side of fig. 1, lower five levels): signals received from the outer world are "perceived", i.e. they induce change of state in some sensing device and this change is amplified up to a level suitable to excite a pattern of neurons in the brain. Important factors of this *perception* process, are "attention" and "verbalization", where "verbalization" is the assignment of a category to the perceived pattern, stemming from the "a-priori knowledge" of the observer, i.e. from his/her past learning of "words", mostly done during infancy and adolescence, whereas "attention" is the assignment of a suitable memory space, the size of which depends on the depth or accuracy with which we want that observation to be made.

Subsequent to verbalized perception is the construction of *experience*, which requires some extraction of repetitive patterns from several perceived events, to be utilized for prediction of future events of similar nature.

One step further to experience is the construction of *science* which, apart from a higger level of generalization (i.e. of extraction of common patterns from a wider class of events), aims at creating models as independent of observers as possible, as well as at extending their

validity possibly to the whole spacetime domain.

Verbal communication among human beings appears to be increasingly decisive along the three steps considered above. However, the language of *mathematics* lends itself, better than the language of words, to express general models of science in a more compact and more universal form. As the Pythagorean scientist Phylolaos (as well as, much later on, Lord Kelvin) maintained, there is no "true" knowledge if it cannot be expressed with numbers, thus implying that knowledge only expressable in words is just the beginning of knowledge, not yet "true" knowledge.

This view is, however, to be taken with some caution, mainly because *philosophy*, in the quest of that unique principle, if any – which is believed to govern all events in the visible universe, and therefore to stay atop of all sciences and physical laws – usually utilizes again the language of words, though trying to obey to precise logic patterns and rules of reasoning as those accompanying the mathematical expressions of science. On the other hand, according to more recent scientific believes, anything which is expressable in words can be dealt with by knowledge processing machines, therefore again can be represented by streams of ones and zeroes, i.e. by binary numbers.

The whole process described above – from perception to philosophy – of acquisition and construction of ever higher ranks of knowledge can be termed as "induction" (from latin "in–ductio" which means "taking in") and leads from mere perception and experience of everyday life to ever more abstract but powerful models, where details become blurred, as when we look at the surface of earth from higher and higher elevations. We may then say that the constructions of mind lead from the analytic mode of the process of observation, based on boolean distinguishability, namely on the "principle of two", to the synthetic mode of the olistic conception of the universe, i.e. to the "principle of one".

Conversely (see rightside of fig. 1, lower five levels), the active behaviour of human beings towards the outside world follows the opposite path, that of deduction[1] (from latin "de–ductio" which means "taking–out"), which transforms general principles, as those of ethics or behavioural constraints, of philosophical nature, into actions performed through boolean actuators (muscles), through successive steps such as *simulation* and *evaluation of alternatives* (implying prediction of their effects), *decision* on which alternative to implement, and finaly *planning* and *execution* of actions, to induce observable changes in the outside world. Thus, this process can be deemed as that converting abstract principles (high–rank knowledge, or information) into physical actions. However, feedback from the induction to the deduction process (and viceversa) can occur at any level shown, whereby actions/events are converted into information and therefrom information is converted back into actions/events. Of course, the capability to effectively interact with the environment is as great as greater is the knowledge acquired in the past, therefore, it generally increases with age and education, apart from innate qualities, which we will touch upon in the following.

It is also to be remarked that especially superior animals (e.g. mammals), but also, in various degrees, all entities in the biosphere, possess at least one or two of the lower levels shown in fig. 1. The same can be said of some artifacts.

In fact, modern AI techniques such as neural networks, autolearning, person/machine communication with natural language, problem solving and robotics (which perhaps will utilize most of other AI techniques) aim to conferring to artifacts all five induction/deduction levels shown in the lower part of fig. 1 (perhaps with the exception of philosophy) and this will be progressively achieved as long as the logic path which takes from the outside world to superior ranks of knowledge/behaviour will be more and more

clearly understood. One decisive step in this direction will be the achievement of full self learning capability, through which a "baby machine" may become a fullfledged "experienced machine" capable of successfully challenging the rational behaviour of the most "expert" human being. The conceptual jump in the design of such superintelligent artifacts will be that of switching from today's "stored algorithms" mode of operation to that of "stored paradigms" i.e. to the storage of scopes and constraints, leaving the execution of steps to strategies permitted by the built-in structure (e.g. neural networks) and by the subsequent machine learning process.

2. The boolean and the ecstatic observer

Despite the marvels of the above sketched rational behaviour (which has been attributed by many scholars to the human brain's left hemisphere), according to Einstein, "there is no logic path that takes from observation of the outside world to the discovery of fundamental laws of nature". Einstein maintained (but anyone of us could easily agree with him) that his greatest ideas came from intuition, vague at first, then progressivelly refined with logic reasoning and with confrontation with experience. Therefore the growth of a bright idea is certainly indebted to the rational behaviour, but the birth of it has to be attributed to a sort of uncontrolled emergence from the ocean of the uncertainty, hence to irrational behaviour; according to the abovesaid scholars, this is the domain of the brain's right hemisphere, which works on a synthetic instead of an analythic mode, and whose acquisition of knowledge is generally accompanied by that sentiment which we call "emotion". This is the "back door" which we were speaking of at the beginning of the paper.

To better understand this "synthetic mode" of acquiring knowledge, think to a summer sunset, as observed by either a "boolean observer" (100% rational) or an "ecstatic observer" (100% emotional). The former may behave like a sophisticated robot, measuring the intensity and wavelenght of light from all directions, the chemical composition, temperature and velocity of the surrounding air, the various patterns of objects in the visual field, all acoustic sound patterns and their probable sources, and so on. The final result will be myriads of bits, though well organized, classified and properly labeled. The boolean observer, with its 100% analythic mode, will break the reality in an infinite number of bits and pieces, as deep as deeper (i.e. more detailed) knowledge acquisition is demanded.

As for the ecstatic observer, he/she must first forget his/her own identity and become mentally part of the observed scenary, trying not to distinguish (even not the self from the environment), not to "reason", not to analyse, but to *live* that experience as a whole. Strangely enough, he/she has not to shut down his/her senses; on the contrary, these should be all fully alert to receive all sensations from the environment[2]: colours, shapes, perfumes, breezes, sounds, etc., though possibly in a non-verbalized perception mode, aiming to sense more the harmony of all those sensations, as constituents of a unified experience. Yoga and Zen techniques, as well as any form of transcendence or misticism, lead to that "ecstatic" ("ecstasis", in Greek, means "union") or synthetic mode of acquiring knowledge (but perhaps "knowledge" may not be the best word to use here) and, as known, they cannot be "taught" (it would require to know the "science" of a nonscientific behaviour) but must be lived or experienced through intense and longlasting exercises, though with the aid of a senior experimenter.

Of course, human behaviour follows from a mixture of the two modes outlined above,

also because the two brain's hemispheres are connected through a (non too high) number of fibers of the "corpus callosum".

A first understanding of the human limits to follow either mode of behaviour may be gained through a review of the Heisenberg's uncertainty principle.

3. The boolean observer and the uncertainty principle

The uncertainty principle is usually stated in terms of the limit to the product of accuracies attainable on the simultaneous measurements of two "conjugate" quantities, such as energy and time or momentum and position.

However, it can also be formulated in terms of "action" (a physical quantity given by the product of the two abovesaid conjugated quantities), which has the advantage of being a relativistic invariant, i.e. independent of the observers. In this case, the uncertainty principle says that "any event, involving an action "A" less than $\hbar$ cannot be observed"[3]. This means that any stimulus with duration t_s and energy E_s such that $A=E_s\, t_s \leq \hbar$, is not able to fire a bistable device, i.e. to induce an elementary change of state, which would correspond to the acquisition of one bit of information by a boolean observer. Thus, the maximum information I which can be gained by said observer, whenever an action $A>\hbar$ is involved, is:

$$I=A/\hbar \tag{1}$$

or, the minimum action A required to acquire I bits of information is:

$$A=\hbar I \tag{2}$$

Relation (1) says that, in the ideal case of maximum (boolean) observation efficiency, the number of action quanta equals the number of information bits. This can be visualized as in fig. 2, where the construction of knowledge by a boolean observer is shown as a pile of information bits, corresponding, one-to-one, to the number of action quanta received by his/her/its sensors. Of course, knowledge is given by events happened in the "past" of the observer and the "herenow" sensation is that coming from any new (additional) event, superimposed to those already interiorized in the memory structure of the observer.

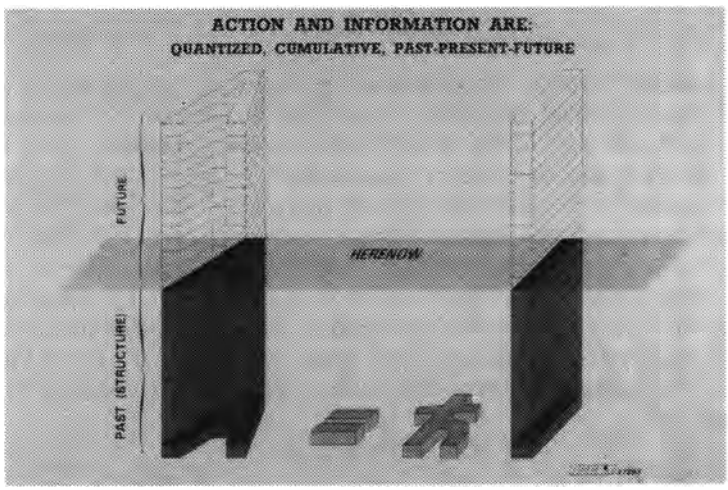


Fig. 2

The construction of higher ranks of knowledge, then, proceeds from all elementary perceptions along the scheme outlined in fig. 1, with a continuous interaction with past (or a-priori) knowledge and with an increasing degree of abstraction, which latter, somehow, destroys the actual event perceived, by substituting, in its place, a sort of convolution of present sensations with past experiences of the individual observer. Communication among many observers, then, tries to construct a sort of "objective truth" as the "invariant" component drawn from the varied experiences of the individuals.

The same as certainty is constructed with the unit of information called "bit", uncertainty may be deemed as being constructed with units of doubt, which may be called "ambiguity" or "negative bit", in that it represents the elementary uncertainty between two equiprobable states. To better understand this unit of doubt, the reader may refer to any of the three examples shown in fig. 3, by verifying that the observer's mind is incapable of deciding which one of two categories (names) to assign to the perceived visual patterns.

AMBIGUITY OR "NEGATIVE BIT"
(Uncertainty between two equiprobable states)

Q.	Concave or Convex?	Head or Vase?	Stairs or Capital?
A.	Concave and Convex	Head and Vase	Stairs and Capital

Fig. 3

More precisely (the reader can experiment by staring at each figure for, at least, 10 seconds), the observer's mind "switches" from one to the other of two "meanings" every few seconds, so that, at the end, the answer to the (boolean) question: "A *or* B?" is "A *and* B", because the observed pattern contains *both* A and B. This suggests a more pleasurable interpretation of uncertainty in that it appears to possess more "plenitude" that certainty, also because certainty implies "choice", i.e. destruction of all other alternatives, whereas uncertainty leaves all the alternatives "alive", i.e. with equal chance of existence.

Therefore, the construction of large ensembles of information, as well as that of large ensembles of doubt, may be deemed as stemming from binary ladders made of either one building blocks (fig. 4), which lead to either a choice or a non-choice (i.e. a "plenitude") of one among N equiprobable states. This can be expressed, in logarithmic form, by the quantity M, given by:

$$M = \log_2 N, \tag{3}$$

representing either (binary) information I or (informational) entropy S, given by the number of either (binary) choices or non-choices needed to arrive at one member of the

whole roster of alternatives N considered[4]. The above justifies the name of "neg-entropy" given by L. Brillouin to information, as well as the name of "neg-information" which may be given, conversely, to entropy.

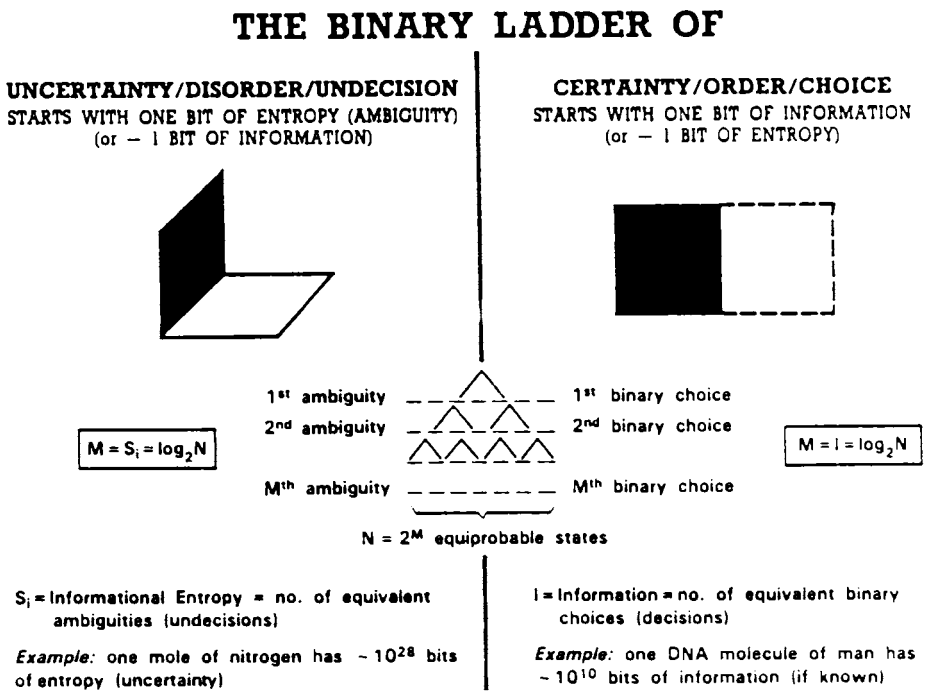


Fig. 4

It should be pointed out that certainty or uncertainty about objects or events – hence their measures: information or entropy – are quantitatively related to the amount of knowledge obtainable on them by a human observer. Therefore, when we say, e.g., that a mole of nitrogen has 10^{28} bits of entropy, it means that we lack 10^{28} bits of information *to know* all the states of that volume of gas. On the other hand, when we say that we have 10^{10} bits of information on some object or event (e.g. a DNA molecule), it means that *we know* all binary microstates that describe that object or event.

Fig.4 also suggests that, beyond the Heisenberg's barrier to the acquisition of (analythic) knowledge by both human mind and artifacts, the huge ocean of Chaos extends with about the same structure of Order, if it were not for the "and" glue that holds together the bits of Chaos, instead of the "or" acid that breaks the bits of Order apart from Chaos.

4. Real and Imaginary World

From what preceeds, it appears that the process of observation – more generally, the process of interaction of the observer with the outside world – results in the construction of an inner, progressively immaterial world, made of "subjectivized" records coming from

the abstraction[5] of words, sounds, visual patterns, signs, physical events, etc. perceived in the past: the older they are, (i.e. the lower they are located in the pile of fig. 2), the less easily distinguishable (recallable) they become at "present", as if they had rejoined the ocean of chaos and uncertainty.

To an alert mind, the outside world, perceived through senses, represents the "real world", that where actions are performed and events (i.e. observable changes) happen; it also represents the "future", as the observer clearly sees the possibility of physically getting in touch with the objects perceived, even if the information received about those objects refers on how they were in the near or far past, depending on their spacial distance from the observer and on the nature of the information carrying messengers. Again, very distant objects become less easily distinguishable, up to the complete black-out occurring at an approx. 15 billion light-years distance, corresponding to the emersion of the universe from the Chaos (Big Bang).

Therefore, far past and far future events seem to merge into the ocean of Chaos, being therefore precluded to the analytic mode of observation by human mind.

Around the "herenow", the near past and the near future seem to touch in a sort of a moving borderline, that of the observer's conscious life, which acts as a two-way transformer of either past into future (actions) or future into past (observations/ information).

Quantum physics allows us to grasp an understanding of such processes (fig. 5). In fact, either any observation or any action performed, implies the occurrence of a *change* of state of some entity. This is described in mathematical terms by the change of the entity's wave function $\bar{\Psi}$ from its initial state $\bar{\Psi}_{in}$ into its final state $\bar{\Psi}_{out}$, related one to the other by:

$$\bar{\psi}_{out} = e^{-j\frac{Ht}{\hbar}} \bar{\psi}_{in} \tag{4}$$

whereby the change of $\bar{\Psi}_{in}$ into $\bar{\Psi}_{out}$ is attributed to a *cause* (stimulus), expressed by the so-called Hamiltonian H, which represents the energy of the stimulus, or, better, by the Hamiltonian action Ht performed by the Hamiltonian, applied along the time interval t. The effect of the applied stimulus on $\bar{\Psi}_{in}$ is, as shown by (4), to only change *the phase* of $\bar{\Psi}_{in}$, not its magnitude; more precisely, the number of radians of the induced phase shift on $\bar{\Psi}_{in}$ equals the number of action quanta ($Ht/\hbar$) of the applied stimulus. Therefore, the magnitude of $\bar{\Psi}_{out}$ equals the magnitude of $\bar{\Psi}_{in}$.

It is know from quantum physics that the square $|\Psi|^2$ of the magnitude $|\Psi|$ of any wave function, represents a (scalar) probability. Therefore, if we write the wave function in a full exponential form, i.e.:

$$\bar{\psi} = |\psi| \, e^{j\alpha} = e^{\beta + j\alpha} \quad , \tag{5}$$

we may see that the real part ß of the complex argument of $\bar{\Psi}$ represents a (negative) information quantity, as:

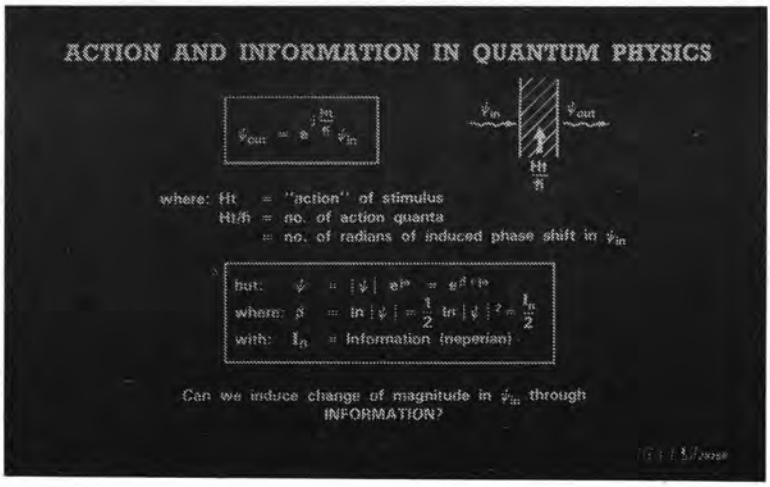


Fig. 5

$$\beta = \ln |\psi| = \frac{1}{2} \ln |\psi|^2 = -\frac{I_n}{2} \quad , \tag{6}$$

given the fact that information is defined as the negative logarithm of probability. As the logarithm is with base "e", the information I shown in (6) is neperian, and is therefore measured in nats, instead of$_n$ bits, the two measures differing merely by a constant of proportionality (ln 2).

In conclusion, we see from (5), that, in force of (6), the description of "reality" in quantum physics rests on two basic (real) quantities ß and α, the former representing *information*, the latter representing *action*, and that these two quantities seem to pertain to two somewhat separated and independent worlds, though both codetermining the observed "reality". In fact, it is a firm result of quantum mechanics that the description of entities and events in the microcosm cannot be obtained without using complex quantities, such as the wave function $\bar{\Psi}$ or its complex argument ß + jα, as in (5). Any dissection of this "complex" reality (made of a "real" and an "imaginary" part) would lead to a monocular understanding of nature, whereas complex numbers, as was pointed out by Leibnitz since year 1690, possess the "amphibian property" of taking into account the influence of both the "real" and the "imaginary", whatever meaning is attributed to these two adjectives, apart from the strictly mathematical one.

However, as it was shown in the foregoing, the world of "pure" information is that inside the observer, that of "pure" action is outside the observer. One is immaterial (the world of mind), the other one is material (the world of energy/matter) and this latter is the so-called physical world, in which observable changes either happen or are intentionally provoked by observers.

Now, we have learned, along the evolution of the biosphere, how to induce such changes in the physical world, through physical actions, as by relation (4). Will we learn how to manipulate pure information (i.e. use the "force of mind") to also induce changes in the physical world, e.g. by changing the magnitude, instead of the phase, of $\bar{\Psi}$? Or, by utilizing the intrinsic capabilities of entities to transform information into action, by taking control of their inner world through entering their "back door" through our mind, as done, e. g., with hypnosis?

Answers to these questions are being increasingly sought, along the last few decades, by several prestigious researchers in various Universities, with enough scientific evidence being acquired on telepathic and precognitive phenomena (technically labeled as "information transfer under sensory shielding") as well as on "low-level psychokinesis" and "out-of-the-body" experiences, none of them being explicable through any known (subluminal) information or action or carrying messenger.

It is a firm belief of the writer that future progresses in this direction could be better achieved if we become fully aware of the existence of the abovesaid two worlds and if we will succeed in taking control of the innermost "imaginary" world of entities in the universe.

Notes

(1) In this context, the term "deduction" is used in a somewhat broader sense, with respect to that adopted e. g. in AI or in psychology.

(2) The reader may recall the "ganzfeld" conditions requested in some experiments of parapsychology, as also performed by the Nobel Prize on Physics, R. P. Feynman.

(3) h is given by the Planck's constant h, divided by 2π. Moreover, the above statement is to be taken statistically, as known.

(4) M also represents the memory size, i. e. the number of physical bistable devices, needed to store all N alternatives, in that it can assume $N=2^M$ possible configurations.

(5) The word "abstraction" (from latin "abs-tractio", meaning "drawing from") not only has the usual meaning of rendering "abstract" (i. e. intangible) what is physically perceived, but also of extracting what is fundamental to that perception, i.e. what may be common (or "invariant") to a wider class of similar observations.

II

TEORIA DOS USOS

THEORY OF USES

1

CONCEPÇÃO & ORGANIZAÇÃO

DESIGN & ORGANIZATION

Ronald M. Lee

Dep. of Management Science and Information Systems,
Univ. of Texas, Austin (USA)

Limits to Technological Rationalization: the Sociology of Semantics

"We believe that managers live in a constant state of transition. Perplexity is always within the manager's mind, and this will not change. The manager will continue to act without fully understanding and will not consider this to be a problem; while attempting to increase his understanding, he never expects to arrive at a full understanding."[1]

Abstract

Rapid improvements in computer technology and the widespread implementation of computer networks offer substantial economic incentives to rationalize administration, not only within the organization, but also between organizations in such forms as electronic banking, electronic purchasing and other forms of contract negotiation and performance monitoring. Broader forms of social rationalization will no doubt follow in areas such as tax collection, import/export regulations, insurance and welfare administration. But there are limits to such technological rationalization, due to the social mechanisms of semantic change. In this paper, the functions of an information system are characterized abstractly in terms of predicate logic deductions. The domain-dependent semantics of a predicate logic application are specified as the denotations of individual names and predicate constants. The sociology of determining these denotations in organizational and social settings is examined.

Resumo

O rápido avanço na tecnologia de computadores e a implementação cada vez mais generalizada de "redes" de computadores oferecem incentivos económicos substanciais para racionalizar a administração, não só no interior de organizações, mas também de inter-organizações, sob a forma de banco electrónico, compra electrónica e outras formas de negociação de contratos e avaliação de performance. Formas mais amplas de racionalização social seguir-se-ão sem dúvida em áreas tais como colecta de impostos, regulamentação de importação/exportação, seguros e administração da segurança social. Há, no entanto, limites para tal racionalização tecnológica, devido aos mecanismos sociais de modificação semântica. Neste trabalho, caracterizam-se abstractamente, em termos de deduções de lógica predicativa, as funções de um sistema de informação. Especifica-se a semântica (dependente do domínio) de uma aplicação de lógica predicativa em termos de denotações de nomes individuais e constantes predicativas. Examina-se a sociologia da determinação destas denotações nos contextos organizacionais e sociais.

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 331/344
1992, APDC & SPF, Lisboa

1.Trends Toward Technological Rationalization

During the past several decades, the use of information technology within organization has evolved into a major factor in the design of administrative systems. More recently, with complementary advances in communication technology, information technology is having an impact on the conduct of business between organizations and on social activities more broadly. Examples include electronic funds transfer systems (EFTS); various airline, car rental and hotel reservation systems; credit card systems; and electronic shopping systems such as the video disk terminals seen at some airports. Financial markets, most notably the London Stock Exchange and the Over The Counter market in the US, are becoming automated, allowing access over communication networks for worldwide, around the clock trading (Clemons and Adams, 1988). Most significant, perhaps, is the development of Electronic Data Interchange (EDI) standards for electronic communications of structured business documents (e.g. purchase orders, invoices). This standardization is being conducted on a worldwide basis by the International Standards Organization (ISO), which coordinates national standards groups such as the X12 Commitee of the American National Standards Institute (ANSI) in the US (e.g. Hinge, 1988).

The impact of standardizing and automating business communication is far reaching. An immediate effect is that transmission becomes (effectively) instantaneous, rather than relying on postal systems. More significant, however, is that the processing of inter-firm transactions can be automated. Consider purchasing for a manufacturing firm: the firm's inventory program detects an upcoming shortage of item X. It sends out a request for quotation, citing quantities and delivery requirements, to each of the vendors supplying X. Their computers receive these quotation requests, and after examining their own inventory, pricing and delivery schedule files, respond to the requesting firm's computer with a quotation. This computer then selects a quotation and transmits a purchase order to the chosen vendor's computer. The entire process is automatic! This vision not only is technologically feasible, but also provides such tremendous economic advantages that its continual development seem assured.

One may imagine this process extending to other types of business and social activities that are conducted through structured communications. Of particular interest are activities that are subject to complicated regulations and/or that involve intricate contractual terms, resulting in large amounts of formal documents or 'red tape'. Examples include import/export procedures, immigration and naturalization procedures, insurance claim procedures, and welfare applications procedures. By automating these communications, the processes could be made not only more efficient, but also more just. Such systems could inform the clients of the rules applicable to their situation, any particular rights or privileges to which they may be entitled, as well as the current status of their request (Lee, 1988b).

Taken together, these technological developments have a broader effect on the design of organizations and modes of doing business. Most organizations, as they learn from experience, tend to implement preferred ways of doing things in the form of explicit rules and procedures. This process is known as organizational rationalization. Organizational rationalization also aids in the coordination of the firm in that personnel can be relied upon to perform duties in prescribed ways, eliminating the need for communication to coordinate their activities.

However, as the number of rules and procedures accumulates, their complexity itself becomes a problem. 'Red tape', the shuffling of standardized paper forms, begins to proliferate. Personnel in the organization may not thoroughly understand the procedures, especially how to handle exceptional problems, leading to their deferring the responsability

to someone else, i.e. 'passing the buck'. Thus, the negative side of organizational rationalization is bureaucracy.

The new technology has the power not only to electronically inter-connect members of the organization, but also -- where the communications are standardized -- to make automated deductions about them. Thus, more and more, the rules and procedures of the organization are being controlled and monitored by its computer. The principal effect, gradually being realized, is to relieve many of the complexity barriers to organizational rationalization, enabling the organization to more effectively 'program' its policies in the form of automated rules and procedures.

As we have indicated, this trend applies not only within organizations, but also to transactions between firms. We refer to this trend as *technological rationalization*. Its impact, we believe, has only just began, but is a potentially far-reaching as that of the telephone in its extent, and much more profound in its effect on organizational and social administration.[2]

But, how far can such technological rationalization continue? This paper argues that technological rationalization, while useful, is also limited. This limitation arises from the dynamics of organizations -- the socio-political processes of organizational change. These processes, we will argue, also bring about subtle and pervasive changes in the semantics of the organization's language. Insofar as this language is also embedded in the information system, the correctness of the system's inferencing will depend on its being able to accomodate this linguistic change.

We approach this problem at a very fundamental level, examining the logical and semantic underpinnings of information systems. Our focus is on the mechanisms by which semantic change is introduced and conveyed throughout the social context of the information system. The role of semantics in computer applications is to validate computational inferences. For instance, if the computer has a rule

$$\forall x \; \mathrm{FRUIT}(x) \rightarrow \mathrm{LEMON}(x)$$

the inference (modus ponens) that any particular lemon is also a fruit is true if and only if the denotation of the term **LEMON** is a subset of the denotation of the term **FRUIT**. If, for reasons of organizational or social interest, the semantics of these changes so that the subset relationship no longer holds, the computational inference will no longer be valid.

This lead us to examine the philosophical aspects of *reference*, how a symbol used in the system is linked to its denotation. This pertains both to the naming of individuals and the naming of classes. A perspective is given in the work of such philosophers as Kripke (1971, 1972) and Putnam (1970, 1978). Schwartz (1977), a collected edition on this subject, dubs their collective proposals 'the new theory of reference'.

The discussion to follow focuses initially on individuals and the epistemic aspects of proper names. Building on that, we consider the recognition and naming of natural kinds, processes dependent on social conventions that, particularly for economic goods, change over time. Social movements and economic innovation are reflected in linguistic changes. Organizational and social adaptation likewise depend on linguistic evolvability. The reliance of structured information systems on a fixed, stable semantics constrains this evolution, thus limiting technological rationalization.

2. Logic Modeling Perspective

In order to regard information systems more abstractly, we characterize their functions as deductions within predicate logic. For some time, predicate logic has been regarded as a useful representational foundation for theoretical research in database management

systems (e.g. Gallaire and Minker, 1978). It has also been proposed as the reference paradigm for artificial intelligence representations (e.g. Genesereth and Nilsson, 1987). Further, predicate logic is the basis for the class of programming languages known as logic programming, the most common dialect being Prolog (e.g. Sterling and Shapiro, 1986).

A difficulty with predicate logic is that it is computationally undecidable. Prolog avoids this problem by restricting its rules to be Horn clauses, having the general form:

$$\Phi_0 \leftarrow \Phi_1 \ \& \ \Phi_2 \ \& \ \ldots \ \& \ \Phi_n$$

where the Φ_i are predicate expressions. All unbound variables are presumed to be universally quantified at the left of the rule. A basic fact its expressed as a single predicate expression of the form:

$$\Phi_0$$

For example,

LIKES(TOM, MARY).
LIKES(DICK, MARY).
LIKES(HARRY, ALICE).
RIVALS(x, y) ← LIKES(x, z) & LIKES(y, z).

Obtaining information from a logic program amounts to proving a specified goal theorem, instantiating variables in the process. For example,

?- RIVALS(x, y).
x = TOM Y = DICK

There are of course many further subtleties to logic programming beyond this trivial example[3]. As a computer language, Prolog has both virtues and faults. Its principal virtues include its modeling flexibility and the elegance and economy of this notation. Its principal faults are that, even with the Horn clause restriction, logic programs tend to be inefficient, necessitating non-logical devices to avoid fruitless searching. While the practicality of logic programming for large-scale information systems may be debated, it does nonetheless serve well as a theoretical tool.

At the Graduate School of Business at the University of Texas we have been using logic programming (Prolog) for several years to develop operational, predicate logic based representation languages for modeling various aspects of business, including bureaucratic procedures, accounting control systems, and electronic contracting systems. The goal of this work is to develop formal languages for business communication. For example, we hope to represent a substantial class of contracts in such a formal language. The functions of the system include analyzing the contract, looking for loopholes, recommending legal controls, and -- when used on a communication network -- monitoring contract performance. A special emphasis in this work is to model deontic inferencing, that is, the dynamics of commitments, including obligations, permissions, prohibitions and waivers.

Like other forms of information systems analysis, this work has the objective of formalizing business operations. In our case, however, we view this formalization from the perspective of predicate logic, what we call logic modeling. A specific advantage of this approach is that it uses denotational (model theoretic) semantics to give sharp focus to the semantics of the representation language. The semantic rules for the logical connectives and quantification in predicate logic are well developed (e.g. van Fraasen, 1971; Dowty, Wall and Peters, 1981). What remains for a particular predicate logic application is to specify the semantics for its individual and predicate constants[4]. Normally, these constants are simply stipulated in the meta-language, e.g. English. But, when used as the basis for information systems used by many people in multiple organizational contexts and over extended periods of time, we must consider this stipulation of meaning with respect to these broader

social processes.

In the discussion following, we examine the practical limitations of semantics specification at two levels: identifying (1) the individuals denoted by logical names, and (2) the classes denoted by predicate constants.

3. Identification of Individuals

Identification of individuals requires being able to recognize the denotation of an individual name. To use a name (in a formal language or a database) presupposes that all of the users of the language/database agree on the object designated by that name. Thus, the problem of identification is related to the problem of sameness. If A and B are names, A = B asserts that these name the same object[5]. This has a static aspect, recognizing the individual at a point in time, and a dynamic aspect, recognizing that individual as it undergoes change. Useful for these issues of identification is the ontological study by Strawson (1959). He claims that our basic criterion for recognizing individuals is the object's location in a spatio-temporal framework. Hence, the objects easiest to identify with reasonable consensus are e.g. physical objects and events somehow related to physical objects. However, even for these individuals, difficulties may arise. Kent (1978) cites numerous examples arising in data processing aplications. For example, consider the recognition of an individual street:

> What is one street? Sometimes the name changes; that is, different segments along the same straight path have different names. Based on a comparison of address, we would probably surmise that people on those various segments lived on different streets. On the other hand, different streets in the same town may have the same name. Now what does an address comparison imply?
>
> Sometimes a street is made up of descontinuous segments, perhaps because intervening sections just haven't been built yet. They may not even be on a straight line, because the ultimate street on somebody's master plan curves and wiggles all around. And sometimes I can make a right turn, then after some distance make a left turn and be back on a street with the same name as the first. Is that one street with a jog? When do we start thinking of these as different streets having the same name?

Abstract objects, not locatable in space and time, pose even more difficult identification problems. Consider for example Beethoven's 5th Symphony. What is the designation of this term? Is it the event of Beethoven's composing this piece? Is it the paper it was originally written on? Is it the collection of all paper reproductions of this original? Is it a musical performance of this piece? Is it all musical performances, past and future, of this piece? We would like to say that it is none of these. Beethoven's 5th Symphony is an idea, and these examples are all mere conveyances of this idea.

A more modern example, of growing economic importance, is the identification of computer programs. Like a symphony, computer programs have static representations on disk, in core, etc. as well as performances -- in the execution of the program. Yet, again we would like to claim that the computer program is actually an idea.

Consider the problem of software theft. Typically when something is stolen, the owner suffers its loss. In the case of stolen software, however, the owner often can't even detect the loss. Theft is probably the wrong word here -- it is actually more like plagiarism. However, the point is the same: if we want to talk about a particular symphony or computer program, a spatio-temporal framework seems inadequate for consensual reference of these individuals; some other conventions need to be established[6].

Another aspect of the problem of identifying individuals relates to parts decompositions. For instance, a car consists of a body, motor, tires, etc. The motor in turn consists of engine block, crankshaft, pistons, carburetor, etc. The carburetor in turn consists of valves, etc. In this case, each of the parts is detachable and replaceable in the whole. But what, then, is the whole? How can we recognize the individual if any aspect of its composition may change? A classic illustration from philosophy is the so called boat of Theseus. Imagine a wooden boat. We replace a plank of the boat, setting the old plank aside. Is the same boat? We replace other planks, one by one, until all the planks have been replaced. Is it still the same boat? If not, which replacement caused it to be a different boat? But now, we take the planks we have set aside and build a new boat. Is this not the original boat since it is composed of all and only those parts in the original? (To aggravate the argument, we can iterate this process to create an entire navy of apparently identical boats.)

Many other examples arise in business data processing, e.g. replacement of parts, phases of manufacture, remodeling of buildings, reorganizations, mergers of corporations. In many of these cases, there seems to be no essential criterion that determines the temporal extent of the entity(7). Fortunately, the problem for databases does not depend on an absolute answer, only on a consensual one. The important issue is the correspondence between the name we have for the individual as it undergoes transformations throughout time . In particular, if we have an individual which today we name X, how will we recognize that individual tomorrow?

In the case of (middle-sized) physical objects, the problem is often resolved by imprinting or tagging the individual with an identification number, e.g. serial numbers on vehicles; inventory codes on office equipment; and numbers on rooms, streets and buildings.

In the case of persons, the continuity of naming is generally maintained in the person's own memory. For example, a baby is named 'John Doe' by his parents. In his baby years he learns that name as his designator. As an adult, when I first meet him, he tells me, "My name is John Doe". When he phones me a month later he says, "Hello, this is John Doe", and so on.

People can change their names. They typically provide the continuity of identification by telling you -- e.g. "My name is Mary Doe -- I used to be Mary Adams before I was married."

There are, generally speaking, social incentives for a person to consistently report his name through time. It is principally through this convention that the person is known to the various social institutions. Also, because of society's dependence on this convention, criminals can sometimes 'change their identity' by altering their physical appearance, but most importantly, by reporting a different name when asked.

To avoid the problems of consistent spelling and possible duplication of ordinary names, organizations and government agencies often provide their employees, clients, etc. with identifying codes -- e.g. an employee number, customer number, social security number. Normally, these identifying codes are connected to the individual in a way similar to ordinary names, i.e., the person consistently reports the same code.

In other cases, where there might be incentives for mis-reporting, a further device is required to perpetuate the association of the code with the individual. A common example currently is credit cards. Here the perpetuating device is the physical possession of the card. A macabre example from the past was in concentration camps. Prisioners were physically tattooed with their identification number. Not only did this prevent them from misrepresenting their identity, but also facilitated identification after death.

In the philosophical literature, these social mechanisms by which a name continues to be associated with the thing it designates are called 'causal chains.' These have been used to explain philosophical puzzles in such sentences as:

'Mark Twain is Samuel Clemens'

'The Morning Star is the Evening Star.'

A more practical example, typical in data processing, is the following:

CUSTOMER #12 = SUPPLIER #57

Normally if two names uniquely designate a common object, they should be interchangeable in any context. If that were true, the above sentences would be tautologies. The fact that indeed convey information is a result of their having different causal chains associating each name with its referent.

Our acquaintance with a particular individual is typically not continuous across time. We see a friend one day, again a week later, etc. Further, the sense data we have of that person is often incomplete and highly varied -- we see the person with different dress, in different lights and from different angles and distances. That is, our sense data of that person amounts to samplings of different aspects of his/her physical appearance, voice, manners, etc. However, we need very little of this data to infer the continuity of that person through time. Lacking anything else, the consistent reporting of a name is often sufficient evidence for us especially if 'we' are an organization or institution. A college story goes that a fraternity enrolled their mascot dog in the university, putting his name on exams, etc., until he was finally graduated with a bachelor's degree. The story may be false, but it makes the point: an institution's acquaintance with individuals is based heavily on the reporting of names.

4. Identification of Classes

Of particular interest here are the categories we use for our everyday objects, so-called 'natural kinds', such as lemon, bottle, chair. When examined through the perspective of denotational semantics, the existence of such kinds is puzzling. This puzzlement arises largely as a result of adopting set theory as the principle device for explaining how real world objects are organized.

The problem is essentially that any collection of objects can constitute a set -- e.g. the set consisting of my toothbrush, the Eiffel Tower, and the planet Saturn. Given all the possible sets of things, why are some, e.g. lemons, chairs, given a special status and assigned a name? The problem becomes all the more complicated if the dimensions of time and possible worlds are added. (E.g. an arbitrary set might then include hypothetical individuals such as Abraham Lincoln's automobile, the present King of France, and the city of Atlantis.)

An early reply to this problem was that natural kinds are sets defined intensionally. That is, there are certain 'critical properties' that select the members of these sets.

An obvious problem with this view is that, from the standpoint of denotational semantics, this involves a circular argument: one cannot explain properties by their set denotations and then turn around and explain the sets denoted by their intensional properties. However, this circularity merely casts doubt on the denotational approach to semantics, suggesting that perhaps intensions should be taken as primary after all.

However, the criterial properties approach also quickly runs into difficulties. Consider the concept of a chair. What are its criterial properties?

a. that it has four legs? No, there are chairs without four legs.

b. that it has a horizontal surface and a vertical back? No, for instance a bean bag chair has neither of these.
c. that it is something to sit on? We can sit on many things that aren't chairs.

This view is conjunctive -- it requires that each element of the intended set satisfy the several criterial properties simultaneously.

An alternative view is disjunctive: that there are no single properties that hold throughout the set of things we call chairs, but it is rather the disjunction of several properties that defines this set. For example, a chair is:

four-legged **or**
has vertical and horizontal surfaces, **or**
is used to sit on, etc.

The problem with the disjunctive view is that it includes too many things -- e.g. tables would count as chairs by this definition. Wittgenstein (1953/1958) presented a classic philosophical discussion of the shortcomings of the criterial properties (or essentialist) view of semantics.

More recently, some have attempted to get around this problem by saying that the denotation of natural kind predicates is a fuzzy set. Without debating the adequacy of fuzzy set theory, we observe merely that this misses the basic point. The problem is not whether the boundaries of these sets are sharp or fuzzy, but rather why these sets are selected and named in the first place.

The mechanism of causal chains used to explain the naming of individuals, also pertains to natural kinds. Consider that a set of objects can alternatively be viewed as another (collective) individual[8]. For example, we often regard football teams, departments, or forests as individuals composed of other individuals. Now consider the case with water. Normally we do not deal with water individuals at the molecular level, but rather with water individuated in collective units or containers, e.g. water droplets, a cup of water, a puddle of water, a lake, an ocean. Mentally, we can easily conceive of emptying smaller containers into larger ones to form a larger water individual. If we consider that most of the oceans and seas are interconnected, we can fairly easily imagine the world as a very large water container. Further, our concept of the water it contains is a fairly permanent one. We think of the water in the world as going through various transformations (snow, ice, vapor, steam), but its sum total on the planet is basically fixed through time (ignoring molecular transformations). Is there then a sharp difference between a water individual and the natural kind, water? One might object that the kind, water, would also include water on other planets. However, that objection only requires us to imagine a larger water container.

As noted earlier, we come to know the features of an individual (person, etc.) through a series of occasional glimpses, each conveying certain aspects of that individual. However, where our own sense impressions are not sufficient for us to formulate a (spatially and temporally) cohesive image, we rely heavily on the social conventions (causal chains) by which proper names are conveyed as the basis for our knowledge of the individual. The new theory of reference argues that our knowledge of natural kinds has a similar basis. We encounter individuals of a natural kind, e.g. lemons, as glimpses or aspects of the entire kind. However, the knowledge we obtain by this direct experience would not, in general, suffice for us to know the absolute extent of the kind and/or to distinguish its criterial properties.

For example, how many people can distinguish a lemon from a yellow lime? Contrariwise, there are some green lemons growing in Brazil. Unless told otherwise, most people, even after detailed inspection, would probably mis-identify these green lemons as

limes that simply have a somewhat different taste.

When I go to the store to buy lemons, I rely heavily on their being labeled as such. On the other hand, even when the lemons aren't labeled, I usually get it right since they are the only small yellow objects in the fruit section (since yellow limes are rarely sold). Here I am using one of the characteristic (but not criterial) properties of lemons to select it from a limited range of alternatives. Further, the limitation to a few alternatives has been socially determined by the institution of supermarket, and the fruit section. I might not be so successful in the open jungle.

Putnam suggests that semantics, ultimately, depends on socio-linguistic considerations. In particular, the references of natural kind terms are seldom completely understood by people individually, but rather as a cooperative effort. He proposes a

> *Hypothesis of the Universality of the Division of Linguistic Labor:* Every linguistic community exemplifies the sort of division of linguistic labor just described; that is, it possesses at least some terms whose associated 'criteria' are known only to a subset of the speakers who acquire the terms, and whose use by the other speakers depends upon a structured cooperation between them and the speakers in the relevant subsets. ... We may summarize this discussion by pointing out that there are two sorts of tools in the world: ... there are tools like a hammer or screwdriver which can be used by one person; and there are tools like a steamship which require the cooperative activity of a number of persons to use. Words have been thought of too much on the model of the first sort of tool. [Putnam, 1970/1977:126–127]

5. Natural vs Artificial Kinds in Organizational Vocabulary

In the philosophical literature, the term 'natural kinds' is used to indicate (the referents of) a wide range of natural language substantives, e.g. water, lemons, chair, house. Quine (1969) pointed out that for *some* natural kinds, e.g. water, lemons, there exists a scientifically accepted procedure of identification. For instance, chemistry defines water as the molecular compound H_2O, botany (I think) has a criterial definition for lemons or at least for lemon trees. It is generally recognized that scientific explanation is ultimately a matter of social convention that changes as new theories are proposed (Kuhn 1962). We no longer accept the 'ether' as the basic substance of the universe. Likewise, it's conceivable (though not likely) that the scientific conception of water might change with further discoveries in particle physics.

But scientific explanation is a unique type of social convention in that it is *authoritative*. What science accepts, the world accepts. H_2O is accepted as *the* definition of water because chemistry says so. Our informal conception of water includes water plus other impurities, though, if disagreements arise, we generally accept the chemical explanation as the criterial definition of 'pure' water.

This is not to say that all people who use the term 'water' throughly understand its chemistry. No doubt, only a few do. Rather the semantics of this term rely on a social cooperation that leads ultimately to certain scientifically qualified individuals. A similar semantic cooperation exists in the common understanding of 'lemon,' which leads backwards from consumer, to supermarket, to farmer, to botanist.

However, not all scientist paradigms are international in scope. Physics is, economics insn't. Further, even within a given society, the various scientific disciplines have differing epistemological status. For example, physics and chemistry seem to have more social credibility than psychology and sociology. This has important linguistic consequences, for

without this credibility our informal usage of terminology can have a different denotation than the scientific usage. For example, we accept as (pure) water exactly that which a chemist analyzes as H_2O. However, we do not, for instance, accept the meaning of 'anxiety' to be that which psychometrics measures using Galvanic skin response.

The difficulty (Thom 1975, Berlinski 1976) is in the structural stability that can be assumed of the phenomena under study. We are comfortable with the assumption that physical phenomena are time/space invariant. Water is water whether on earth or on moon, in the eons past or those to come. Biological sciences have to consider evolutionary factors: fruit flies vary from one continent to another, bacterial diseases can evolve in a matter of months. The social sciences have an ever weaker claim to structural stability.

The importance of these observations to the current theme is in their linguistic consequences. The semantics of a given term is clearly a matter of social convention. However, that observation doesn't take us very far unless we understand the relevant socio-linguistic mechanisms. One of these mechanisms is the authority granted to scientific theories as defining the references to certain of our terms.

Thus the semantic problems of database translation and the verification of inference are likely to be least difficult for predicate terms that are derived from the natural (physical, biological) sciences. More semantic instability is to be expected for terms signifying social artifacts, what we might call 'artificial kinds.' Consider again the term 'chair'. It seems doubtful that there can *ever* be a scientific definition of this concept. Indeed, there are furniture design companies whose marketing strategy is to change our current conception of this term.

The vocabulary originating in these scientific disciplines enjoys a fairly wide semantic consensus and stability. Its rate of linguistic change is likely to be slow relative to the time frame of its users. By contrast, consider the terminology relating to technological applications. In these areas the rate of linguistic change is much more rapid. It is, nonetheless, a fairly organized evolution. For example, trade journals and industry wide meetings and exhibitions help to standardize usage. To enable compatibility between products and processes, industry standards are eventually developed. This too helps to standardize usage. However, a great deal of the vocabulary in organizational databases relates to socially defined phenomena. The relevant factor here is the social scope of the organization's interactions. An example is the relationship 'marriage.' In most cases, this can be accepted as a stable concept, relative to the interests of the organization. In other cases, e.g. the census bureau or cross-cultural organizations, linguistic variations have to be considered.

A more fundamental point is that the organization itself defines a social context and creates its own social artifacts. Prominent among these is its product offering, which, to be successful, is intentionally differentiated from related products in the marketplace. This product offering is furthermore dynamic, the effect of product development and marketing efforts. While the attendant linguistic change is managed within the organization, serious difficulties arise for e.g. regulatory, taxation, and consumer protection agencies.

The structure of the organization is itself a social artifact, which includes the identification of organizational substructures (divisions, departments, committees), organizational roles (manager, clerk), procedures, rules, standard documents, etc. These are described in a rich, locally defined, organizational vocabulary. As the organizational structure and processes evolve in response to changes in the environment, its vocabulary must correspondingly evolve. If this vocabulary is used only for informal communications within the organization, this evolution continues naturally. However, as more and more of

this vocabulary becomes embedded in the organization's structured information system, this linguistic evolution, hence the organization's adaptability, is constrained.

The difficulty follows from the semantics presumed for formal languages. The design of a formal language depends on Frege's 'Principle of Compositionality'. That is, the semantics of a compound expression are constructed from the semantics of its syntactic constituents. *If the assumption of semantic stability is removed, the deductions provided by the information system can no longer be guaranteed to hold.*

6. The Politics of Semantic Change

Semantics is the mapping from a symbolic representation to its referents. The concerns of the past section reflect not only the relationship between individual terms and their denotations (e.g. **LEMON** and all lemons), but the relationships between these individual terms to others, e.g. the use of lemons in cooking, their medicinal aspects, the marketing and distribution channels for lemons, the growing of lemons, the genetic aspects of lemon varieties. It is this range of aspects that requires a whole social network for complete understanding. However, the knowledge of a concept at any particular node could well be formalized and subject to mechanical inference. Indeed the successes of computers to date reflect this possibility.

The problems considered here are the mechanisms by which semantic change is introduced and conveyed throughout these social networks, and the consequences of this for computer aids to these cognitive processes. To repeat an earlier observation, the role of semantics in computer applications is to validate computational inferences. If the computer has a rule

$$\forall x \ \mathrm{FRUIT}(x) \rightarrow \mathrm{LEMON}(x)$$

the inference (modus ponens) that any particular lemon is also a fruit is true if and only if the denotation of the term **LEMON** is a subset of the denotation of the term **FRUIT**. If, for reasons of organizational or social interest, the semantics of these terms changes so that the subset relationship no longer holds, the computational inference may no longer be valid.

We, as people, make words mean what we want them to mean. Language is not an abstract entity, existing independently. It is a social artifact, a behavioral convention, that has been found to be pragmatically useful.

At the micro-level are so-called 'baptism' events. I can *name* my cat George or Ludwig or whatever. When people ask my cat's name, I tell them the name I chose. They accept that name as a matter of pragmatic convenience: it's hardly worth our time arguing over it. Further, since I own my pets, I have a social right to do the naming. Similarly, when I have a telephone installed, the phone company *tells* me the name (telephone number) for it. They, in this case, have the baptism right. (A computer might actually select the phone number in question based on an algorithm, but the system of phone number assignments is nonetheless the social right of the phone company.)

Now consider generic terminology. When a scientist makes a new discovery, invents a new process, etc., he/she or the research group *assigns* a name to it . This name is promulgated via the patent registration, research publications, academic conferences, etc. Likewise, in the economic domain, when a company markets a new product, it invents a name for it. If the product succeeds and is widely sold, it becomes a 'household word' and is incorporated into the social vocabulary. Sometimes, if the product dominates competitive products, the brand name is used generically, e.g. Coca-cola, Xerox machines, Kleenex,

IBM cards.

Some people, then, invent new vocabulary. Other people choose whether or not to accept it. Not all novel research is accepted for publication; many marketing innovations fall flat. Acceptance of linguistic innovations is a matter of pragmatic consensus.

The point is underscored by Wittgenstein's (1953/1958) distinction between *observing* and *deciding* the meanings of words. For much of our vocabulary, we are passive observers of the semantics. We accept the dominant conventions established by traditional usage. In some situations however, especially where we want to refer to some phenomenon or idea for which there is no term with the designation we want, we need to change the language. We make this change either by creating a new term or by modifying the designation of an existing term. Thus, to use one of Wittgenstein's examples, "What if the diviner tells us that when he holds the rod he *feels* that the water is five feet under the ground? or that he *feels* that a mixture of copper and gold is five feet under the ground?" (1958/1965:9). The term chosen for such new and unusual cases is typically one with a related meaning. When the altered usage is temporary, limited to a specific context, we call it metaphorical. But methaphors often catch on and become part of the popular vocabulary.

Thus, in deciding the semantics of terms, there are two steps: *proposing* and *accepting* the new meaning. The essential argument is that the mechanisms of semantic change rely on social pragmatism. Words change usage and new terms appear when people want to talk about new things or ideas. Language evolves to meet their changing interests and needs. Correspondingly, new uses of language are accepted by others depending on *their* interests and needs. New meanings are negotiated, like new products on the marketplace.

The point is that we humans have the social *privilege* to buy and sell new meanings. In contrast, when my dog growls at automobiles, I don't for a minute consider allowing that usage into my vocabulary. Dogs don't count as linguistic innovators. However, if a great social plilosopher should growl in the same way at automobiles (say, perhaps, accompanied by a convincing argument that they are the root of all evil), I might reconsider the usage and perhaps adopt it myself. Computers, like dogs, don't count as linguistic innovators. When a computer program bombs and starts printing garbage, I regard its output as a mistake, not as new vocabulary.

Therein lie the limits to technological rationalization: while it is conceivable that computers might eventually be able to induce changes in the meanings of words based on usage, that is, *observe* new meanings, it is highly unlikely that they will be allowed linguistic innovation, to *decide* new meanings. Thus, human input and control are intrinsic factors in an information system in which an organization's constantly evolving language is embedded. Accommodating linguistic change is a peculiarly human activity that involves political processes, that is, the negotiation of interests, values and preferences.

Acknowledgments

Many thanks to Sandra Dewitz for her useful comments and criticisms.

Notes

(1) Discussion group no. 4, at the IFIP/IIASA Task Force Meeting on Processes and Tools for Decision Support, Schloss Laxenburg, Austria. Proceedings as Fick and Sprague (1980). Members of the group were P. Boxer, F. Flores, R. Hackathorn, P. Hearson, G. Kochetkov, S. Pearson, C. Stabell, B. Trippett

(2) A popular concern is the effect these developments might have on the centralization of power, enabling an Orwellian type of bureaucratic state. Contrary scenarios are McLuhan's vision of a 'global village' or Toffler's 'third wave' where the technology encourages greater decentralization plus greater inter–dependency. These organizational and political issues certainly need further study, but are not considered here.

(3) Even here there is a subtlety: note that RIVALS (DICK, TOM) is also a theorem, since RIVALS is symmetric. To the user, this is redundant information.

(4) For our purposes here, we omit discussion of function constants.

(5) Two sameness problems are sometimes distinguished: sameness of individuals vs sameness of kind. For example, "John and Bill drive the same car" may mean they drive the same particular car, or that they each drive distinct cars that are of the same type (satisfy the same predicates). Insofar as the extension of a predicate is a set of individuals, sameness of kind also relies on individuation.

(6) The following is apt:
"Though the concept of a work covered by copyright is embodied in international conventions, there may be differences in national interpretations of the concept. ... in accordance with German tradition, protection is provided ... for an abstract concept of the work, determined by the needs for protection of literary and musical works. The French tradition, on the other hand, is more oriented towards the physical manifestations of the work determined by the greater interest taken in works of fine arts." (from *Information, Computer Communication Policy 8: An Exploration of Legal Issues in Information and Communication Technologies*. Paris: Organization for Economic Co–operation and Development, 1983).

(7) Note: living things seem to be the one major exception here. We take the ongoing process of life to mark the continuity of the individual, even though, like the boat, all of its cells may eventually be replaced. However, even here difficulties are beginning to arise, particularly surrounding the morality of abortion. Does the human individual begin at conception or at birth?

(8) As for instance in Goodman's (1951/77) 'calculus of individuals'.

Bibliography

Asimov, I.1978. *I, Robot.* New York: Fawcett.

Berlinsky, D. 1976. *On Systems Analysis: An Essay Concerning the Limitations of Some Mathematical Methods in the Social, Political and Biological Sciences.* Cambridge: MIT Press.

Boulding, K. E. 1956. *The Image.* Ann Arbor: The University of Michigan Press.

Boulding, K. E. 1978. *Stable Peace.* Austin: University of Texas Press.

Clemons, Eric K. and Adams, Jennifer T. 1988. *"International Opportunities for the Use of Information Systems in Securities Trading Created by Dereglation of the London Stock Exchange," Proceedings of the International Conference on Information Systems,* pp.25–32.

Cyert, R.M. and March, J. G,, 1963. *A Behavioral Theory of the Firm.* Englewood Cliffs, New Jersey: Prentice–Hall.

Deal, T. E. and Kennedy A. A. 1982. *Corporate Cultures.* Reading, Massachusetts: Addison–Wesley.

Dowty, D. R., Wall, R. W. and Peters, S. 1981. *Introduction to Montague Semantics.* Boston: Reidel.

Fick, G. and Sprague, R. H. Jr., eds. 1980. *Decision Support Systems: Issues and Challenges.* Oxford: Pergamon Press.

Gallaire, H., and Minker, J., eds. 1978. *Logic and Data Bases*. New York and London; Plenum Press.

Genesereth, M. R. and Nilsson, N. J. 1987. *Logical Foundations of Artificial Intelligence.* Morgan Kaufmann Publishers.

Goodman, N. 1951/1977. *The Structure of Appearance.* Third Edition. Dordrecht: D. Reidel.

Hinge, Kathleen C. 1988. *Electronic Data Interchange -- From Understanding to Implementation,* American Management Association Publications.

Kent, W. 1978. *Data and Reality.* Amesterdam: North-Holland.

Kripke, S. 1971. *"Identity and Necessity,"* in M. K. Munitz, ed., *Identity and Individuation.* pp. 135–164. New York University Press. Also reprinted in: S. P. Schwartz, ed. 1977. *Naming, Necessity, and Natural Kinds*, London: Cornell University Press.

Kripke, S. A. 1972. *"Naming and Necessity,"* In Davidson and Harman, eds., *Semantics of Natural Language,* pp. 253–355.

Kuhn, T. S. 1962. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Lee, R. M. 1984. *"Bureaucracie, Bureaucrats and Information Technology,"* European Journal of Operational Research, 18, pp. 293–303.

Lee, R. M. 1984. *"Automating Red Tape: The Performative vs Informative Roles of Bureaucratic Documents,"* Offices -- Technology and People, Vol. 2, pp. 187–194.

Lee, R. M. 1985. *"Bureaucracy as Artificial Intelligence,"* in Humphreys, P. (ed.) *Knowledge Representation for Decision Support,* Proceedings of IFIP WG 8.3 Working Conference (Durham, England; July, 1984), North-Holland.

Lee, R. M. 1988a. *"A Logic Model for Electronic Contracting,"* Decision Support Systems, 4, pp. 27–44.

Lee, R. M. 1988b. *"Bureaucracies as Deontic Systems",* ACM Transactions on Office Information Systems Vol. 6, No. 2, pp. 87–108.

Lee, R. M., Coelho, H. and Cotta, J. C. 1985. *"Temporal Inferencing on Administrative Databases",* Information Systems, Vol. 10, No.2, pp. 197–206.

March, J. G. and Olsen, J.P. 1979. *Ambiguity and Choice in Organizations.* 2nd Edition. Bergen: Universitetsforlaget.

Peters, T. J. 1980. *"Management Systems: The Language of Organizational Character and Competence,"* Organizational Dynamics. Summer.

Putnam, H. 1970. *"Is Semantics Possible?"* in: H. E. Keefer and M. K. Munitz, eds, Language, Belief, and Metaphysics. New York: State University of New York Press. Also reprinted in; Schwartz, S. P., ed., 1977.

Putnam, H. 1978. *Meaning and the Moral Sciences.* Boston: Routledge & Kegan Paul.

Quine, W. V. 1960. *Word and Object,* The MIT Press.

Quine, W. V. 1969. *"Speaking of Objects,"* in Quine, W. V. *Ontological Relativity and Other Essays,* pp. 1–25.

Schwartz, S. P. ed. 1977. *Naming, Necessity, and Natural Kinds.* London: Cornell University Press.

Sibatani, A. 1980. *The Japanese Brain.* Science 80. December: 22–27.

Sterling, L. and Shapiro, E. 1986. *The Art of Prolog.* MIT Press.

Strawson, P. F. 1959. *Individuals.* Garden City, New York: Anchor Books.

Thom, R. 1975. *Structural Stability and Morphogenesis. An Outline of a General Theory of Models.* Reading, Massachussetts: W. A. Benjamin.

van Fraasen, B. C. 1971. *Formal Semantics and Logic.* New York: Macmillan.

Wittgenstein, L. 1953/1958. *Philosophical Investigations.* Translated by G. E. M. Anscombe. Third Edition. New York: Macmillan.

Wittgenstein, L. 1958/1965. *The Blue and Brown Books.* New York: Harper & Row.

Zadeh, L. A., K. S. Fu, K. Tanaka, and M. Shimura (eds.). 1975. *Fuzzy Sets and Their Applications to Cognitive and Decision Processes.* New York: Academic Press.

Margarita Vasquez & Manuel Liz
Univ. La Laguna, Tenerife (Spain)

Designing Systems

Abstract

The point about artificial systems is that men build them. That is what we mean when say that men design. We can consider that designs are conceptual systems with action rules and with models about how these action rules are related with our goals and with representations that we have about the systems that we want to design.

Resumo

A questão acerca de sistemas artificiais é que são construídos por homens. É o que significamos ao dizer que os homens projectam. Podemos considerar que os projectos são sistemas conceptuais com regras de acção e com modelos sobre como estas regras de acção se relacionam com os nossos fins e com as representações que formamos dos sistemas que queremos projectar.

1. Introduction

In other places[1], we have distinguished between natural and artificial systems. The only point in which these two kind of systems are different is in the way they have got the structure that let them to be considered as that. If the man has put this structure, the system will be artificial. In any other case, the system will be a natural one.

The important point about artificial systems is that, because men build, they put in them the right structure to get their goals and purposes. That is what we mean when we say that men design[2]. In this sense, we can talk about technological planning. A technological planning is a design. With this very special design we can change a natural system into an artificial one, an artificial one into another artificial or we can build a new system too.

2. Two ways to talk about technological plannings

First of all, if we want to talk about that kind of plannings that we call technological plannings, we could consider them in two main ways:

1. We could consider a technological planning like the set of rules that we should use in order to get some goals and purposes, that we have previously chosen.
2. We could consider a technological planning like a conceptual system. A system compoused mainly by technological rules.

If we consider it in the first way, we are going to get in trouble if the system that we are trying to modify changes in a different way than the one we had thought. This is because we have already decided what rules we are going to use in order to get the goals

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 345/348
1992, APDC & SPF, Lisboa

and purposes, but we have forgotten everything that in the beginning doesn't look like pretty important. Then, in this first case, we have only two options: either we have to decide apriori the formal structure of our planning, or we have to modify continously this structure to put in or take away new elements and relations that become important or stop being so.

We don't find this kind of problem in the second of our ways. In this second case, we consider a technological planning like a system. If we define what a system is, following Bunge[3], the environment is very important, and the environment is going to let us to distinguish perfectly between what is common in every technological planning, its general structure, and what belongs to each particular one.

3. Systems of tchnological planning

Then, we can define a system, S, of technological planning, S=[T,E,R], composed by technological rules, T, the environment of those technological rules, E, and the relations, R, between the components of T and between T and E.

We could consider another system S', S'=[T',E',R'], where

T'=T

E'=E-E

R'=R-Relations between T and E

If we have these two systems, S and S', by one hand, we can understand S' like the representation of a kind of general structures: we have the technological rules and the relations between those technological rules. That is, the way in which those technological rules are going to relate in order to get our goals and purposes.

By the other hand, we can understand S like the representation of a particular system structure. Each particular system structure will have different important elements in the environment. If the environment elements are important, we are not going to find the problem that we could find in the first way, in which we had to choose between giving a-priori formal structures to every planning and changing this formal structure when we find anything unexpected.

This advantage of systems, that let us to make the difference between system and environment elements, is typical of General Systems Theory, that wants to:

1. give us a general methodology to build up conceptual models, that are able to represent, explain, and sometimes predicte too, real systems. Real systems that can be both natural or artificial.
2. stay in the middle of the formal and the empirical, between mathematical structures, without empirical meaning, and empirical hypothesis, that want to represent and explain.

4. Kinds of relation and behavior

Going on with our main case, by one side, the relations, R, R', between the technological rules could be logical or empirical. By the other side, the relations, R, between the components of the technological planning and its environment would be empirical ones. By this way, the system view let us to distinguish two kinds of relations:

1. between the elements of T: general logic and empirical relations;
2. between T and E: empirical relations;

and, so, to describe two kinds of behavior in a technological planning:

1. its logical behavior with some general empirical conditions;

2. its particular empirical behavior.

Then, we could say (1) that so some things are logical consequences of a technological planning system under some empirical conditions, and (2) that some other things are particular empirical results of the technological planning system.

The logical behavior under some empirical conditions (1) could be studied with the formal tools of the operational research, of the modal, temporal, deontic, and so on, logics, and, perhaps, we should find new logics of the technological speech (technological probabilities).

The particular empirical behavior of technological systems (2) could be studied by the mathematical tools of system dynamics and its simulation with the help of computers.

5. System Dynamics view

System Dynamics have a great power to deal with these empirical behavior. This is easily understood if we think in the newest studies about qualitative change and "chaos"[4].

The problem of the unexpected future changes in the system that we were going to design, that we could find in the first way of understanding designs like a set of action rules, doesn't appear here. A design with adequate models in System Dynamics gets representations of all the different kinds of behavior of the systems that we are going to design.

In the same way, and this is another important point in systems view, it could have dispositives to regulate automatically action -technological- rules by these changes. In other words, it could have feedback mechanisms.

The ways we should change our designs in relation with changes in the systems we want to design, would belong to the own design. If we understand designs only like sets of action rules, all of these could do nearly impossible to ennumerate the set of rules for any action. But, in the systemic view, models could have implicitely all the important knowledge about what we have to do if something changes in the object of our design. It is important to point out here that the most important thing in most of designs is the "time" that we give to make the design. If we put into it a lot of feedback structures, it can be useful in many cases.

6. "Natural" and "artificial". Conclusions

Coming back to the "natural" and the "artificial", Artificial Systems get their artificiality in relation to the way they get their structure. Its beginning must be human. The man makes a system artificial when he designs it. In this just moment, it is very important, we could say necessary, to talk about intentions, patterns, goals, purposes, etc. Because of that, nobody should think that a TV set, justly the same that our TV sets, but casually created, for example, through accidental and strange combinations of vulcanic magma, is something artificial. I think that all of us should say that this very special TV set is a natural thing. Its beginning wasn't human in any way. It was not a human goal. But, by the other side, if I throw a stone to my neigborough, I don't make, by this action, an artificial stone. It doesn't matter if I hate my neigborough more than any other person in the whole world.

This is because in some way the design must belong to the structure of the real system that we want to be artificial. We have to structurally change its natural constitution. This is the technological action that brings out artificiality.

If we understand designs only like sets of action rules, we can't understand these

thoughts. But, we can consider that designs are conceptual systems with action – technological– rules and with models about how these action rules are related with our goals and with representations, it doesn't matter if they are scientific or not, that we have about the systems that we want to design. If we consider designs in this way, some of the relations, that belong to R, should be representations of empirical and real relations.

They should be the relations that give the structure to the real systems that we want to design and through which we can describe them related with our goals and purposes.

Notes

(1) In Vásquez Campos, Margarita (1988a). *"Sistemas naturales, sistemas artificiales"*. Actas del III Congreso de Lenguajes Naturales y Lenguajes Formales, PPV ED, Barcelona. And in Vásquez Campos, Margarita (1988b) "Planificatión y control en los sistemas artificiales". Revista ARBOR, n.507, 1988.

(2) Simon, Herbert (1973). *Las Ciencias de lo artificial.* ATE, Barcelona.

(3) Bunge, Mario Augusto (1979). *Treatise on Basic Philosophy. Vol. 4. A World of Systems.* Reidel Publishing Co.

(4) Aracil, Javier (1986) *"Análisis cualitativo de sistemas dinámicos"* and Toro, Miguel (1986) *"El colapso de la civilización maya desde la perspectiva del análisis cualitativo en dinámica de sistemas"* in *Modelado y simulación por ordenador de sistemas dinámicos complejos.* Servicio editorial Universidad del País Vasco, Bilbao.

Ioanna Tsivacou

Univ. of the Aegean, Athens (Greece)

Generation and Transference of Meaning in Generic Design Activity

Abstract

This paper uses the semiotic square of A.J. Greimas as an abstract structural model to represent the generation and transference of meaning in two different theoretical constructs. The first is the science of Generic Design founded by J.N. Warfield. The second is the critical theory of J. Habermas. Both constructs despite their differences, derive their legitimization from the linguistic and communication action. The Greimas model when applied to the systemic science of Generic Design makes transparent its structural exigencies. In the case of Habermas the structural representation is applicable because of its rational and teleological nature.

Resumo

Este estudo utiliza o quadro semiótico de A.J. Greimas como um modelo estrutural abstracto para representar a geração e transferência de significações em duas construções teoréticas diferentes. A primeira é a ciência do "Generic Design" fundada por J.N. Warfield. A segunda é a teoria crítica de J. Habermas. Ambas as construções, não obstante o que as diferencia, fazem derivar a sua legitimação da acção linguística e comunicativa. O modelo de Greimas aplicado à ciência sistémica do "Generic Design" torna transparentes as suas exigências estruturais. No caso de Habermas, a representação estrutural é aplicável devido à sua natureza racional e teleológica.

Just as we are not in position to define electricity, we are not in position to define meaning. On the contrary, we can talk, as Greimas does, of the phenomenon of meaning, of the sense of its presence, of the way that meaning is manifested, or about "meaning effect"[1], just as we can talk about the electricity phenomenon and its manifestations.

Meaning to be generated needs language, but language to express meaning needs the communicative function. It is in the communicative activity that the collective hypostatises the social semantic experiences. The collective builds them through the system language, with its words and its sentences, with its sounds and its silences, with its ambiguities and its equivocations, all of which aim, not only at representing the world, but at reflecting experience in such a way that the world is signified.

The language as instrument but also as a structural system of communication, contributes to the exchange of experiences and to their "socialization", therefore to an understanding of the reality and consequently to the consciousness of the desire for the change of that reality, that is to say for the design of its change. So, language produces and nourishes the design, which in turn controls and leads the desire.

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 349/357
1992, APDC & SPF, Lisboa

In the science of Generic Design, founded by J. Warfield[2] the language is one of the universal priors together with the human being and reasoning.

The generic design science is a product of 20 years of research on the properties and behaviour of large scale sociotechnical systems. It is operating at the level of participative and consensus–building solutions applicable to complex situations which characterize the actual sociotechnical organizations.

J. Warfield and his colleagues in the development of the science of generic design do not wish to be critics of philosophy neither metamodern thinkers who bring forward the communicative competence and the consensus methodologies as a form of interference in social life. Rather, they face the methodologies as scientists and they use them as productive means capable of being applied in a context where is already assumed: a) the "systemic acceptability" (i.e. the participants in the design procedure feel as parts of a system) and b) the "cultural availability"[3] (i.e. the participants accept to temporarily suspend their apparent differences on cultural values), necessary conditions for the common confrontation of a problems situation. Design, as the natural or the conceptual representation of the desired reality, consists of the object of a science with its own theory and methodology, ruled by concrete postulates, laws and principles.

The methodology steers towards the domain of the applications from which the science draws the semantic experience in order to interact on the theoretic corpus, but its special purpose is to serve the needs of groups of people working collaboratively. According to Warfield the methodologies operate on a dual–basis condition: (a) The behavioral that induces from the universal prior of the human being and (b) the communication & documentation that results from the universal priors of language & reasoning. The dual–basis may be represented by two axes.

The roles of the pilot/facilitator and the participants in the design process, and also the role of the environment in which the design takes shape being on the first axis. The consensus methodologies that enable the discipline of the discourse and a special software which aims to improve the natural language and to augment the human intelligence belong to the second axis.

The Generic Design Methodology attemps to secure the best prerequisites of communication in order that the development and the transference of meaning be more complete. The final intention of this procedure is the construction of conceptual models in accordance with the world–views of the participants with their idiosyncratic ideas since the discourse, under the guidance of the facilitator and of some endogenous self–control rules is driven to logical inferences.

The aim of generic design in to control scientifically the human actions which transform the future through the use of the methodologies that lead to the designing of those actions, is of great theoretical & practical interest. Through generic design the science expresses its ambitions to conceive the "spontaneous" and to rationalize it without reducing it to nothingness.

The term "spontaneous", in this case refers to the free unfolding of the personality of every speaker, or the unhindered pronounciation of his desires, beyond the constraints of a purposeful self–control that the conditions of the organization in particular, or of the society in general, impose on the individual for the maintennance and the evolution of his/her position in them.

In generic design methodologies we invent certain prerequisites that govern the terms for the anticipation of an ideal speech situation as it is conceived by J. Habermas. We make the comparison with J. Habermas because we believe that despite his opposition to Systems

Theory, the modern Sociotechnical Systems approach belongs to the Action theory tradition just as Habermas does. Habermas'thesis, as also that of Generic Design is an educative model, that tries to emancipate the human existence through rationalization i.e. acting on the level of generation of meaning.

According J. Habermas it is during the communicative action that the so-called "practical" interest is manifested and symbolic and intersubjective actions are generated among human beings. These actions are determined by rules or world-views that govern the collective behaviour and that constitute the institutional context inside of which the "technical" interest functions. A language is required for the expression of the "practical" interest which will also reassume the necessary validity claims for an intersubjective communication. The disturbance of validity claims forces the individuals to establish debates, in order to make possible through the dialogue, the creation of the needed consensus context for the restoration of the validity claims. The dialogue presumes the pre-comprehensibility of participants, a pre-comprehensibility supported by commonly accepted behaviours for the pursuing of collective final objectives and which also provides the basis for the increase of comprehension to higher levels of deduction.

In the institutionalized dialogue of J. Habermas there are no accepted truths verified from a primary experience, as every experience is the product of a previous one, but there are commonly accepted norms, that function as truth until the moment of their verification from the praxis result. J.Habermas puts in continuous doubtfulness and critique the proper data of the dialogue, that is the reality or not of the events, the correctness or not of the normative principles.

The pre-understanding and the understanding of J.Habermas are developed not only from the function of "practical" interest, but, also, from the simultaneous activation of a third interest, named "emancipatory" interest. The last one emerges in a process of self-reflection, reveals to the thinking individual his/her false conscience and leads him/her from the ideological thinking to the rational thinking.

As is well known, J.Habermas criticizes the system approach (and surely he refers to the Hard System's)[4] as an outlined structure, as a logical and hierarchical model, that sits on the social body as a narthex and the empirical world must a posteriori confirm its rightness. According to Habermas, the theory is not verified a posteriori on the basis of the empirical control, but it is governed by a pre-transformed experience of the society as a whole. These pre-transformed experiences find their signifying support in the signifying potential of the past (and at this point Habermas agrees with Benjamin) from which the humanity draws and therefore absorbs and lives its experiences. In this ancient potential are found the "semantic universals which precede all socialization and semantic universals which are linked to the conditions of potential socialization (monological/intersubjective)"[5].

Here, i.e. on the question of language, there is a profound difference between Habermas and the scientists of Generic Design and of Soft Systems in general, though all of them try to realize conditions of communicative competence. J. Warfield, f.e., does not regard for the language as source of meaning, rather he sees it as a device. He is concerned with the accuracy of language so that it may provide maximum flexibility in communication. With this as his main objective, he seeks for a type of scientific language, which, "will promote the development of high quality in the area of large-scale system design"[6]. This kind of language cannot be only the prose, but "a graphically integrated language system (GRAILS). Such a mix correlates greatly with the requirements of interpreting claims of formal logic. Correlating the limitations of the human being with requirements for language design sheds further insight upon the design of languages, allowing the development of a

criterion set for language design to be applied to design and assess a language for use in articulating and applying a science of generic design''[7].

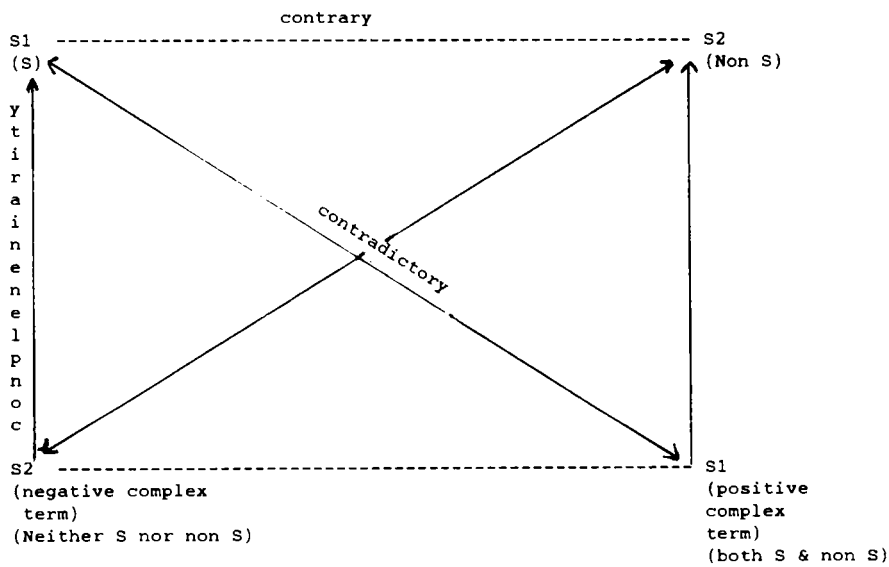
We observe that while in Habermas the meaning pre-exists and emerges in the discourse opposed to the technological-instrumental position, in generic design there is an attempt, through an interactive relation between discourse & information technology, for the development and the transfer of meaning that finally will be stabilized in articulated models.

In order to make more comprehensive the semantic content of generic design its importance during meaning generation and transference and simultaneously to demonstrate its similarities and differences with the communication function of J. Habermas, we are going to use the semiotic square of A.J. Greimas. The semiotic square, an elementary complex structure of signification, can express the duality of language as it has been introduced by F. de Saussure in linguistics between contrast and combination.

In the semiotic square there are three logical relationships. A contrary relationship, a contradictory relationship and a complementary relationship. As R. Schleifer[8] reports: ''The relationships exhaust the logical possibilities of binary oppositions. A contrary (or ''privative'') relationship creates a double relation of conjuction and disjunction in terms of the presence or absence of some shared feature; a contradictory (or ''equipollent'') relationship creates that double relation in terms of a shared function; and a complementary (or ''gradual'') relationship creates that double relation in terms of an (arbitrarly defined) implication''.

Figure 1 is an abstract semiotic square and represents the above mentionned relationships.

FIGURE 1

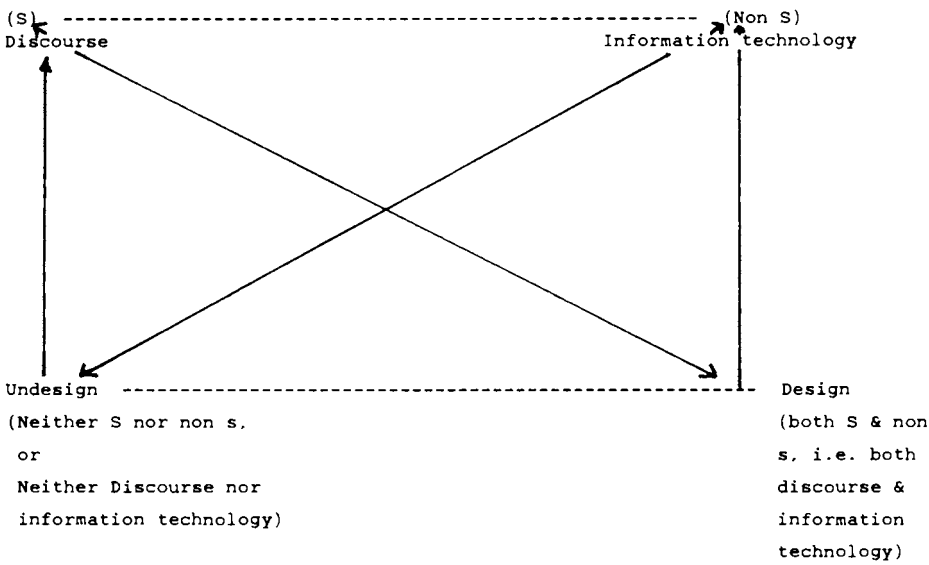


Source: R. Schleifer, "A.J. Greimas and the nature of Meaning", p. 25.

Figure 2 suggests the signification of "generic design", if we consider as S and non S of semiotic square the terms "discourse" and "information technology". Both of them are on the same positive axis but differ in their semantic content and they are connected in a contrary relationship.

The semantic content of "Design" is composed by "Discourse" and "Information Technology", therefore is the third angle of semiotic square and expresses the combination of the two contrary terms: "Discourse" & "Information Technology".

FIGURE 2



The forth term, (neither s nor non s), is the contrary term of Design, that means the undesign, or neither "Discourse", nor "Information Technology".

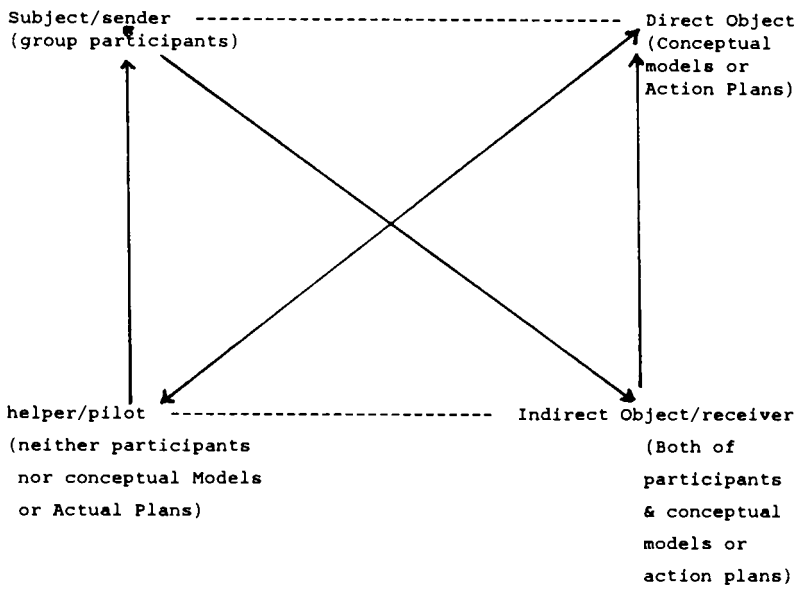
If we accept that the "meaning effect" is manifested only through discourse, then the "Undesign" is equivalent with the non-meaning or non-sense, or with the collective unconsciousness before acquiring signifying expression and feeds with meaning the discourse according Habermas or Benjamin. Thus the "undesign" is a complementary term of "discourse". The world of "Undesign" is a world without signification, it is a non signifying reality, but it is not the absence of the reality. It is the world of desires of Lacan before to be recognized in the language. The object of desires cannot be designed, is undesigned and only when it will be captured by the language and only through the enunciation will acquire meaning. But the discourse, under its double articulation between expression and content, between enunciation and utterance, rationalizes the desire, and objectifies it, by separating it from the subject. The utterance, in order to fully rationalize the desire, to deduct from the sentence the ambiguity and to fully manipulate it, is not satisfied only with the grammaticisation of discourse. It runs to the aid of information technology, so the discourse is articulated in conceptual models and the actants of discourse

become the actors of Design Methodology. The actants of Greimas, deduced from Propp's analysis as Greimas himself explains in structural semantics, are entities which "anthropomorphize somehow the significations" in discourse and represent the double articulation on the duality of language through the pairs:

sender–receiver
subject–object
helper–opponent.

In Generic Design Methodology we have a similar syntactic structure where: sender/subject is the participants in the process of methodology, receiver/object is the construction of Conceptual Models or Action Plans and helper is the pilot/facilitator of Consensus Methodologies. The model of this methodology can also be represented in the form of semiotic square, as the figure 3 suggests:

FIGURE 3



With this syntactic interpretation, the pilot has the key role of generic design methodology. He is not an entity out of the system, that, as Habermas fears, risks to manipulate the participants of the communicative function to his own reasoning, but he is the only part of the system capable of resisting to the logic of this system. He is the negation both of positive and negative that there is in the system, he is the point of neutralisation of meaning. This is exactly the function of pilot that helps the other angles of the square to produce and to exchange messages, he is the pilot, who, as negation of the negation neutralizes the flattening character of the information technology and contributes to the saving of the desire from the affirmative negative statements of the aristotelian logic. In the extent that the pilot, due to his position, obtains the final messages inscribed in the conceptual models to be more informative than the announcements during the process of

the methodology, he has contributed to the development of the communicative function, which in its turn deliberates the hidden meaning. The pilot as a complementary term of the sender, i.e. of the participants, is the helper of the sender for the generation of a new meaning.

The direct object, i.e. the conceptual models or the action plans, is a representation of meaning, which as messages, circulate through the participants and are establishing in the object. From this modelized, articuled and stabilized meaning the sender and the receiver seem to be excluded. If the subject/sender is, as Greimas says, "the incarnation at the level of anthropomorphic grammar of the universe of values",[9] then the object is the objectivization of these values.

In semiotic square of generic design methodology of great interest is the contradictory relationship between subject/sender and indirect object/receiver. As the figure 3 suggests, both terms not on the syntactic level but on the semantic level, contain the group of participants. The procedure of designing is a learning process where the participants are both, senders & receivers of messages. If the direct object of the design is the production of conceptual models, the indirect receivers of these models are the individuals who participate to their construction. The learning process is a closed process where the producers are also the users of the product. Through the contradictory relationship which develops between the same participants as senders vs receivers during the transfer of meaning, the linguistic phenomenon of combination (both S and non S), or the semantic contradiction, which combines the participants with the conceptual models intervenes, and anticipates of the transforming the individuals to the enlightened participants of Habermas. This is possible through the spoken language, the more it is emancipated from the structural forms of the written language, the more permits the presence of the subject in the conceptual models/action plans the more the later, under the pilot's influence, could react as the catalyst of meaning in the system. The contradictory relationship between pilot & conceptual models is a relationship which, in certain circumstances could react as catalyst for the generation of a new meaning. The new meaning will transcend the equilibrium of semiotic square, will explode the equilibrium and will create another plane of content, that means another square, where the point of the negation of the negation will be occupied by another unmarked term.

Conclusion

By employing the semiotic square of Greimas, we tried to demonstrate that the methodology of the Science of Generic Design is capable of meaning representations in a functional manner with the aid of information technology. The same technology on the other hand, incorporates components that are capable of acting not only at functional level but at the evolutionary level. These components are the role of the "Pilotos", the spoken language and the process learning that takes place during the application of the methodology of the Science, since the subject and the indirect object of the process are the participants. The more the participants are enabled to exchange meanings from the undesign world and the a priori experiences of J. Habermas and to embed them in conceptual models and action plans, the more they are capable of acting in a catalyst role in the content of the relevant system and its self-realization.

Notes

(1) A.J. Greimas, *"Semiotics and Language: An Analytical Dictionary"*, Indianna University Press, 1982.

(2) J. Warfield, *"The Science of Generic Design"*, unpublished.

(3) Dr. Ch. Atkinson, poses similar presumptions in his dissertation, *"Metaphor and Systemic Praxis"*, University of Lancaster, 1985, p. 339: "..The teams in other words had to arrive at a methodology that was both systemically desirable and culturally feasible..."

(4) The terms "Hard Systems" – "Soft Systems" have been introduced by P. Checkland [*Systems Thinking, Systems Practice*, J. Wiley, 1981] to define a system according its objectives, decisions taking and measures of performance:
Hard System:– easy to define objectives
– clearly defined decision taking
– quantitative measures of performance.
Soft System:– Hard to define objectives
– Uncertain decisions taking
– qualitative measures of performance.
According P. Checkland, B. Wilson, and others, the hard systems use engineering type methodologies to face well–structured problems, meanwhile the soft systems use soft systems methodologies to approach unstructured problems situations.

(5) J. Habermas, *"Towards a Theory of Communicative Competence"*, Inquiry 13, 1970, p. 363.

(6) J. Warfield, *"On the design of Language for System Design"*, paper presented at the European Meeting for Cybernetics and Systems Research, Vienna, 1988.

(7) J. Warfield, *"A Science of Generic Design"*, op. cit., chapter 2.2.

(8) R. Schleifer, *"A.J. Greimas and the Nature of Meaning: Linguistics, Semiotics and Discourse Theory"*, University of Nebraska Press, 1987, p. 28.

(9) A.J. Greimas, *"Du Sens"*, V. 2, Seuil, Paris, 1983, p.221.

References

–Ch. Atkinson, *"Metaphor and System Praxis"*, Ph'd, thesis, University of Lancaster, 1985.
–P. Checkland, *"Systems Thinking, Systems Practice"*, J. Wiley, Chichester, U.K., 1981.
–B. Fay, *"How People change themselves; the relationship between Critical Theory and its audience"*, in T. Ball (ed.), *"Political Theory and Practice: New Perspectives"*, Minnesota, u.p., 1976.
–A.J. Greimas, *"Structural Semantics: An Attemp at a Method"*, University of Nebraska Press, 1983.
–A.J. Greimas, *"Semiotics and Language: An Analytical Dictionary"*, Indiana University Press, 1982.
–A.J. Greimas, *"Du Sens"*, Seuil, Paris, 1983.
–J. Habermas, *"Consciousness Raising or Redemptive Criticism: J. Habermas on W. Benjamin"*, New German Critique, 17, 1979.
–J. Habermas, *"Towards a Theory of Communicative Competence"*, Inquiry 13, 1970.
–J. Habermas, *"Knowledge and Human Interest"*, Beacon, Boston, 1971.
–M.C. Jackson, *"Systems Methods for Organizational Analysis and Design"*, Systems Research, V. 5, nº 3, 1988.

–J. Lacan, *"The four fundamental concepts of Psychoanalysis"*, Penguin Books, 1978.
–J.C. Oliga, *"Methodological foundations of Systems Methodologies"*, Syst. Practice 1, 1988.
–F. de Saussure, *"Course in General Linguistics"*, Mc Graw Hill, N.Y., 1959.
–R. Schleifer, *"A.J. Greimas and the Nature of Meaning: Linguistics, Theory, Semiotics and Discourse Theory"*, University of Nebraska Press, 1977.
–J. Warfield, *"On the design of Language for System Design"*, paper presented at the European Meeting for Cybernetics and Systems Research, Vienna, 1988.
–J. Warfield, *"The Science of Generic Design"*, unpublished.

David Keever

Center for Interactive Management, George Mason Univ.
Fairfax, Virginia (USA).

Consensual Design of an Organization's Management Information System

Keywords:
Participative design, information systems, paralell learning in organizations.

Abstract

This paper reports on a nine-month case study about the application of a well-tested, participative design system in a small federal government agency. The objective of the design project was to develop a robust, conceptual prototype of the next generation information system for the agency. The design activities were conducted by creating a temporary parallel learning structure involving three stakeholder groups and applying a science-based design system containing special features not found in conventional approaches. Senior managers developed a consensus on goals and issues associated with the new information system. Representatives from the agency Users' Group collectivelly developed 260 functional requirements (organized into 19 categories) and selected short-term and long-term options for the conceptual prototype based on senior managers' goals. A third group, composed primarily of computer consultants, attempted to develop participatively several technical prototypes which were consistent with senior managers' goals and Users' Group requirements. The outcomes of the design activities and the value of the Generic Design Science-based system for the conceptual design are discussed.

Termos-chave:
Projecto participado, sistemas de informação, aprendizagem paralela nas organizações.

Resumo

Esta comunicação refere-se ao estudo de um caso, com a duração de nove meses, de aplicação de um sistema, testado, de projecto participado numa pequena agência do governo federal. O objectivo do projecto consistiu no desenvolvimento de um prototipo conceptual robusto para um sistema de informação de próxima geração destinado à agência. As actividades do projecto foram conduzidas por uma estrutura temporária de aprendizagem paralela compreendendo três grupos de controlo, aplicando-se um sistema de projecto de base científica com aspectos especiais que se não encontram nas abordagens convencionais. Os gestores seniores alcançaram uma situação de consenso acerca de objectivos e virtualidades do novo sistema de informação. Representantes do Grupo de Utentes da agência desenvolveram 260 requisitos funcionais (organizados em 19 categorias) e seleccionaram opções de curto e de longo prazo para o prototipo conceptual baseado nos objectivos dos gestores seniores. Um terceiro grupo, constituído primariamente por consultores informáticos, procurou desenvolver participativamente vários prototipos técnicos que fossem consistentes com os objectivos dos gestores seniores e os requisitos do Grupo de Utentes. Discutem-se os resultados das actividades do projecto e o valimento, para o projecto conceptual, do sistema baseado na ciência do "Generic Design".

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 358/366
1992, APDC & SPF, Lisboa

I) Introduction

The use of information technologies in organizational settings have had mixed reactions. Goldberg, for example, has observed that the "majority of computer installations in the past decade have been deemed 'failures' ranging from disappointment in cases where the system did not yield the benefits expected to disaster where financial control of the company was irretrievably lost" [Goldberg, 1982] A range of complications resulting from the introduction of information technologies in organizations have occurred, including lower wages and higher stress [Olson, 1982], isolation of departments in organizations created by private databases and restricted access [Guimarães, 1981], and damage to interpersonal communication networks [Crozier, 1981], to mention a few.

The objective of this paper is to draw attention to the potential benefits of parallel learning structures for the consensual design of an organization's management information system. Parallel learning structures are supplemental organizations structures for problem-solving and learning created explicity to overcome the well-documented problems of poor adaptability and employee dissatisfaction associated with formal bureaucratic organizations. The use of parallel learning structures is posited as a vastly improved alternative to conventional design practices which ignore critical ingredients and tend only to perpetuate the costly consequences of ill-designed and improperly implemented information technologies in organizational settings. [Perrow, 1983]

The presentation begins with a brief introduction of the concept of parallel learning structures as draw from Bushe and Shani. [Bushe and Shani, 1989] Next, the participative design system used in the case is discussed. A synopsis of the group design activities are presented up to the point of implementing the conceptual design. An informal assessment of the consequences and value of the Generic Design Science-based system for the conceptual design is discussed.

II) Studying Organization Information System Design

The relationship between information technology and organizational change has provided a fertile area for research in the past thirty-five years. Unfortunately much research is the product of disciplinary focus with each researcher employing his or her own preferences for concepts, theories, and methodologies within these two broad topics. Consequently, cross-comparisons and generalizability, among other research attibutes, are hindered. [Huber, 1984; Rousseau, 1985]

Most studies of information system design have been based on a functionalist or mechanistic school of thought to the exclusion of organization theory or culture. For example, the design of information systems is treated as a linear, feedback, or circular system where certain inputs (users' needs, equipment constraints, etc.) are identified, a design process (linear/nonlinear optimization, decomposition methods, etc.) is applied in isolation by experts, and an output is evaluated against pre-established design criteria (expected utility theory, variance analysis, etc.). [Cross, 1984; Vick and Ramamoorthy, 1984]

In studying the interactions of information system design and organizations, the research community is struggling for a research paradigm which will not only allow inquiries to build-upon previous work but also move away from strict disciplinary focuses. Recent work by Bushe and Shani have provided a framework for analyzing these design activities through the use of the concept of "parallel learning structures." [Bushe and

Shani, 1989] This conceptualization focuses attention beyond disciplinary-based activities and can help explain some of the sucesses and failures of information system design and organizational change.

According to their definitions, the key characteristics of parallel learning structures (PLS) are:

1) The purpose of a PLS is to improve existing practice by establishing a time and place where attention is taken away from daily activities and placed on possible futures.
2) The outcomes of PLS are ideas, solutions to problems, innovations, and new understandings of organizational reality.
3) The operating norms and procedures are different from those in the formal organization; an emphasis is on cooperative, non-confrontational, non-hierarchical relations and effective group problem-solving.
4) Members of the PLS are also members of the formal organization. Within the PLS, individuals operate outside the normal chain of command.
5) A senior steering committee is used to control and legitimize the PLS.
6) Mechanisms are used to insure linkages between the PLS and the formal organization. [Bushe and Shani, 1989]

The PLS distinctions move beyond conventional bureaucratic mechanisms which include hiring an external consultant (standard response), establishing task forces and committees (to follow the norms and procedures of the formal organization), or implementing project/matrix management methods (to follow the formal organization norms and to create the dilemma of dual authorities). [Greiner, 1983] Moreover, the PLS allows researchers and practitioners to reach a balanced focus between the type of parallel structure to create (operating rules, sources of authority, etc.) and the design processes to apply (idea generation techniques, dataflow models, etc.)

The design methodologies applied in this case study are embedded in a theory-based science of generic design, a component of which is a participative planning and design system entitled Interactive Management. Briefly, Interactive Management refers to a system of carrying out the important problem-solving functions of Intelligence (Problem Finding), Design (Generation of Alternatives), and Choice (Selection of the Preferred Alternative) participatively, through sessions which involve a studied mix of five synergistic ingredients. These five ingredients are the i) Participants or Group of Stakeholders, ii) the IM Facilitator, iii) computer equipment (software, hardware, and display equipment), iv) a specially-designed "Situation Room", and v) a set of carefully selected and tested Consensus Methodologies. (The Consensus Methodologies used in this case are discussed in more detail in a latter section.) The IM system is based on over 18 years of research in psychology, management, systems science, and computers, resulting in worldwide applications throughout a variety of fields necessitating group interaction and task-oriented results. Underpinning this system is an emerging Science of Generic Design which guides the design activities. [Warfield, 1983; Warfield, 1988]

III) Generic Design Science Applied to the Design of an Organization's Management Information System

The agency in this case[1] is a U.S. government organization entrusted with sponsoring approximately $1 billion worth of industrial and university research projects in its mission area to keep the U.S. alert to new and emerging sciences and technologies. To assist the 100 program managers and 50 support staff members in tracking and managing several

hundred, multi-year, multi-million dollar research projects, a computerized information system has been used since the early 1970s. Program managers' administrative needs include an accurate tracking of their budgets, an ability to "hide" their uncommitted funds from the agency's deputy director (for fear of the director's redirecting the funds for other purposes), the desire by some to be a showcase for new information technologies, and ready access to project data from off-site locations, since on average three of five days are spent on out-of-town business. The agency support staff is composed of trained professionals who support purchasing activities, report preparation, travel arrangements, budget tracking and projections, and routine secretarial and mail support. Some of these activities are performed with the agency's computer-based information system.

The design and redesign of the agency's information system can be accurately characterized as a festering management problem for over 15 years. Issues of standardization, security, employment of the latest technology, networks, and hardware-software capabilities, to mention a few, are in a continuous state of planning and re-evaluation, rendering the current information system unsatisfactory to many of the users. After the installation of the original information system, two separate consulting efforts, one internal and one external to the agency, were attempted to design an organization information system. Both designs were based on classical methods of designing information systems, i.e., mechanistic models of needs assessments through interviews and surveys, a translation of these needs into technical configuration (with cost and schedule ramifications), and presentation to the agency Users' Group[2] for a decision. [Vick and Ramamoorthy, 1984] In neither case were parallel learning structures employed.

The internal consultant, the director of the agency's data-processing unit, had used a trial-and-error approach for assessing users' needs and providing solutions which would address those needs. The outcome of this four-year effort was a patchwork of computer hardware and software resulting in additional requirements for specialized training (for which there was no full-time staff), the need for customized computer programs to permit data exchange and interconnection (a workload which overwhelmed the data-processing unit and had to be purchased from consultants external to the agency), and frustration among the user community to the point that about one-third of the computers were not being used. In their frustration, the Users' Group had recomended to senior management a more comprehensive planning effort.

A long-range planner was hired and assigned the primary project of developing the next generation information system. After a three-month assessment, the project manager hired an external consultant to assist in the planning and conceptual design. The agency's project manager devoted approximately two man-years and $300,000 applying conventional methods (surveys, time-motion studies, etc.) to develop, in conjunction with the external consultant, a conceptual design of the next generation information system. This design was rejected by senior managers and the Users' Group as incomplete, unresponsive to users' needs, and too complex due to the requirements for massive training and changes in current operating procedures. As of May 1987 when the participative design activities started, the agency project manager had yet to establish a consensus on overarching goals and design requirements among senior managers, program managers, and the agency Users' Group.

A participative design approach for defining users' needs and developing conceptual technical prototypes was advocated by the agency's project manager and Users' Group to overcoming the previous planning setbacks. Several factors lead to this position. The complexity of the issues and situation necessitated special planning and design methodolo-

gies and skills not possessed internally by the organization. Also, relationships among program managers, support staff members, and computer experts were highly sensitive and diverse, necessitating an impartial intervention to organize and conduct group activity. Therefore, there was consensus among the project manager and the Users' Group that a parallel learning structure offering process expertise in complex design situations was required.

Through word-of-mouth contact the project manager at the agency was introduced to the staff of the Center for Interactive Management at George Mason University, Fairfax, Virginia, where the Interactive Management (IM) system of problem-solving and design is being researched and applied under sponsored-contract arrangements with selected client organizations. Based on a review of the agency design activities and authorization for the design activities from the agency's director, a contract was signed between the agency and the Center for Interactive Management to provide design process assistance.

The application of the Science of Generic Design for the conceptual design of the information system involved a three-phase, ten-day strategy involving participative design sessions with different stakeholders. Each session lasted no more than three consecutive days although a considerable amount of time was spent in pre-session and post-session planning and reporting. The entire set of planning and design sessions required about four months to complete. All participants (whether members of the agency or external consultants) engaged voluntarily in this design project. They were given a clear role distinction involving five activities: (1) Engage in focused dialogue; (2) Make judgments on relationships among ideas with computer assistance; (3) Represent the views of some constituency; (4) Adopt the posture of individual and collective learning through iteration; and (5) Contribute to the ownership and applications of products. The specifics of each design session were arranged in consultation with the information system project manager and implemented by the IM Facilitator.

In the first phase, senior managers established, during a one day session, a consensus on the critical issues and goals for the next generation information system through the use of the Nominal Group Technique and Interpretive Structural Modeling. Nominal Group Technique (NGT) is an efficient method for generating ideas in a group, for clarifying the generated ideas, for editing the generated ideas, and for developing a preliminary partition of the ideas. Interpretive Structural Modeling (ISM) is a computer-assisted learning process that culminates in the development of a structural model of an issue, problem, plan, project, or design. [Warfield, 1976]

At the beginning of the first phase, the deputy director of the agency provided the context and two primary criteria for the design project: save the time of the program managers on nontechnical, administrative tasks and develop a sufficiently sophisticated and attractive information system which would aid in the recruitment of high-quality program managers and support staff. Financial constraints for the purchase and implementation of the system were minimal; up to $10 million, and possibly more, could be spent if the design was worthwhile and responsive to users' need. However, it was important for the design team to develop, within the next six to nine months, a reasonable set of conceptual prototype designs against which formal contracts for purchase and implementation could be evaluated.

Given this context for the design activity, 11 senior managers and research program managers generated and clarified using the NGT a list of 32 critical issues affecting the development, installation, and operation of the next generation information system. This process allowed the participants to develop a common understanding of critical concerns

such as the need to be easy-to-learn, to be accessible from remote sites, and to protect the program manager from budget and schedule overruns. The NGT technique was applied again with the same group of managers responding to a different triggering question to generate and clarify a total of 44 goal statements. The Interpretive Structural Modeling (ISM) process was used to develop a supportive structure among 25 of the top vote-getting goal statement. An interpretation of the supportive structure indicated the need for an improved, flexible communication network through the use of integrated databases as one of the major goals of the new computer-based information system. Many senior managers were being hampered by an existing computer system which took too long to learn, had awkward, isolated databases and structures, and was more of a hindrance than a support of their program administration activities.

In the second phase, the Users' Group developed, over seven days, all functional requirements of the proposed next generation information system using the Options Field/ Options Profile methodology. The Options Field is a method for developing, organizing, and displaying the options, dimensions, and clusters that are relevant to a design problem. The Options Profile is a method for developing a top-down design, beginning with an Options Field, and ending with a specification of all options selected to form the set of design choices. An Options Profile is the representation of one design alternative. [Warfield and Christakis, 1987]

A total of 260 options were identified and organized into 19 design dimensions. The design dimensions were also sequenced for choice-making among options. The sequenced design dimensions were:

1A. Program Execution	7A. External Database
1B. Image	7B. Administrative Support
2. Program Control	7C. Tools
3. Programmatics	7D. Document Automatiom
4A. User-Friendly	7E. Database Access
4B. System Response	7F. Storage
4C. Data	8. Security
5A. Access	9. Reliability and Maintainability
5B. Communication	10. Information Exchange
6. Policy	

Serially-numbered categories, such as 4A and 4B, indicate strong interdependence during the selection of options. The sequence of the design dimensions demonstrates the primary importance the Users' Group placed on the ability of the next generation information system to be responsive to the research program administrative needs (Program Execution and Programmatics) as well as the Image of quality and concern for employees that the information system will convey, especially to potential employees. Of secondary concern was the need for user-friendliness and system responsiveness. Of tertiary concern was the more technical issues of databases access, reliability and maintainability, etc.

Using the Options Field, members of the the Users' Group developed two preliminary Options Profiles of functional requirements for the next generation information system. One profile was developed under the context of a one-year planning horizon; the second profile was developed under a five-year planning horizon. Approximately two-thirds of the 260 options were selected under each scenario.

In the third and final phase, a group of eight computer consultants and contractors (internal and external to the agency) started to develop participatively several conceptual

technical alternatives which would be responsive to the functional requirements. The Options Field/Options Profile methodology was expanded and used to develop two preliminary technical alternatives. At this point in the project, the technical consultants would not continue to participate collectively in the design process, for a number of reasons, including the alleged need for more information from the user community, the desire to introduce a different design process which was more compatible with their "normal" design practices, an atmosphere of suspicion and confusion about their role as advisors or implementors, and challenges to the conceptual underpinnings of participative design activities. A more complete discussion of the cultural impediments is contained in Keever. [Keever, 1989]

To keep the design project on track, the project manager hired one of the eight computer consultants as the sole technical consultant for the project. This consultant developed, during a five month period, two conceptual alternatives. These were presented to senior managers and the Users' Group for review.

Selection of the final configuration was made easier for senior managers by the development of only two alternatives. Both configurations used a mini-computer as the central depository for a collection of integrated databases for tracking, budgeting, and scheduling agency research programs. Connected to this central computer was a local area network and approximately 100 personal computers. The two design alternatives differed in the type of personal computers to be selected and the subsequent interfacing software. The preferred alternative was chosen by senior managers primarily because of its capability to access the central database with ease (minimal user training) and its capability for producing business graphics for program manager's presentations. This configuration was significantly different from the one proposed 18 months earlier by the external consultant.

IV) Discussion of this Case

The progression of events in this case may be familiar to individuals involved in the management of computer projects. Similar circumstances have become the source for many studies and books on how to overcome barriers or manage change on interdisciplinary projects. [Bennis, 1966; Dyer, 1984; Kilmann, 1984; Livingston, 1987] The participative planning and design activities had progressed smoothly through the development of the short-term and long-term set of functional requirements for the next generation information system. Integrating the eight computer consultants into the parallel learning structure was not as easy, as discussed elsewhere. [Keever, 1989]

The parallel learning structure satisfied a variety of planning and design factors which could not be achieved through the formal organization structure and processes. First, the complexity of the organization information system was beyond the capacity of the internal data processing manager, the long-range planner, and the first external consultant. The science-based theory and methodologies used in this particular parallel learning structure provided a means to overcome the previous inadequacies in knowledge integration, complexity management (in scope and depth), and lack of creativity (groupthink). [Janis and Mann, 1977] Second, by temporarily redefining roles, the design participants were able to overcome previous underconceptualization of the information system, develop a more balanced understanding of the behavioral versus technical factors in the system design, and overcome the influences of egotistical designers and parochialism. Third, the dual role of the participants in the parallel learning structure and the organization contributed to the ownership, application, and dissemination of the products from the participative design

sessions. The rationale for the information system design was more transparent and readily explained by the design participants to more agency members.

V) Summary

A participative design system was used with a federal agency, in a parallel learning structure mode, to develop a conceptual prototype of the next generation, computer-based information system for the agency. The conceptual design provided the basis for the selection of the final computer hardware and software system which is currently being installed. The parallel learning structure provided a structural mechanism to overcome the previously failed design attempts. The value of the design process, based on Generic Design Science, includes a means for handling substantive complexity through behaviorally-balanced processes addressing idea generation, clarification, and organization, the development of structural models which could be easily interpreted and updated by the design participants, the imposition of more productive design role definitions for participants yielding a commitment and enthusiasm to implement the conceptual design, and the use of a tested, supporting environment to manage and implement the design processes.

Notes

(1) The data for this study was collected while under research contract to provide expertise and services on participative design processes. The data is in several forms, including 50 hours of audio and video tape, direct access to the written products from the planning and design sessions, reports from previous design projects, and direct observations as process analyst and organizational liaison.

(2) The agency Users' Group is composed of 12 staff members and two program managers from various departments throughout the agency, some of whom have served on the group since its inception.

Bibliography

Bennis, W., *Beyond Bureaucracy*, McGraw-Hill, New York, 1966.

Bernstein, A., *"Trading Disasters"*, IEEE Spectrum, April 1987: 8.

Bushe, Gervase R. and Shani, Abraham, B., *"Parallel Learning Structures Interventions In Bureaucratic Organizations,"* to be published in R. W. Woodman and W. A. Pasmore (Eds.), *Research in Organizational Change and Development,* 1990.

Christakis, A. N., *"Technology Participative Design: The Space-Based Laser"*, General Systems Yearbook, 1987: 69–75.

Christakis, A. N., David B. Keever, and J. N. Warfield, *"Development of a Generalized Design Theory and Methodology,"* unpublished National Science Foundation Workshop Report, 1987.

Crozier, M., *"Implications for the Organization,"* New Office Technology: Human and Organizational Aspects, Academic Press, New York, 1981.

Cross, N., *Development in Design Methodology*, John Wiley, New York, 1984.

Dyer, W., *Strategies for Managing Change*, Addison-Wesley Publishing, Massachusetts, 1984.

Goldberg, L.A., *"Information Managers: The Key to Computers Success,"* California Magazine, 8(1982): 39–47.

Greiner, L., *Consulting to Management,* Prentice-Hall, New York, 1983.

Guimaraes, T., *"Understanding Implementation Failures,"* Journal of Systems Management, March, (1981): 203–218.

Huber, G., *"The Nature and Design of Post-Industrial Organizations"*, Management Sicence, 30(1984): 928–951.

Janis, I., and Mann, D., *Decision-Making,* Free Press, New York, 1977.

Keever, D., *"Cultural Complexities in The Design of a Management Information System,"* Proceedings of the International Conference on The Mutual Uses of Science and Cybernetics, University of Amsterdam, Amsterdam, The Netherlands, March 1989.

Kilmann, R., *Gain Control of the Corporate Culture*, Jossey–Bass, San Francisco, 1985.

Lindblom, C. and Cohen, D. *Usable Knowledge: Social Science and Social Problem Solving,* Yale University Press, New Haven, CT., 1979.

Livingston, *Organizations,* F.E.S. Publishing, New York, 1987.

Olson, M.H., *"New Information Technology and Organizational Culture, "* MIS Quarterly, Special Issue, 6(1982): 71–92.

Perrow, C., *"The Organizational Context of Human Factors Engineering,"*, Administrative Science Quarterly, 29(1983): 521–541.

Rousseau, D., *"Issues of Level in Organizational Research: Multi-level and Cross-level Perspectives,"* in Research in Organizational Behavior, JAI Press, Greenwich, CT., 7 (1985).

Saaty, T., *The Analytic Hierarchy Method*, John Wiley, New York, 1982.

Thompson, J., *Organizations in Action*, McGraw–Hill, New York, 1967.

Vick, C. R., and Ramamoorthy, C.V., *Handbook of Software Engineering,* Van Nostrand–Reinhold, New York, 1984.

Vosbury, N., *"Process Design"*, in C. Vick and C. Ramamoorthy (Eds.), Handbook of Software Engineering, Van Nostrand–Reinhold, New York, 1984.

Warfield, J. *Societal Systems*, John Wiley, New York, 1976.

Warfield, J. *"Organizations and Systems Learning"*, General Systems Yearbook, 1983: 5–74.

Warfield, J., *"The Domain of Science Model: Evolution and Design,"* in Proceedings of the Annual Meeting of the Society for General Systems Research, Intersystems, Salinas, CA, 1(1986): H46–H59.

Warfield, J. and Christakis, A., *"Dimensionality,"* Systems Research, 4, 2(1987): 127–137.

Warfield, J., *"Implicit Aspects of Much Systems Thinking,"* Systems Research, 5, 4(1988): 333–342.

F. Ross Janes

Interactive Management Unit,
Dept. of Systems Science,
City University, London (UK)

Planning for an industrial training board team: an application of Interactive Management

Abstract

This paper reports on an application of Interactive Management in planning for a development team within a national industrial training board. Commencing with a brief description of the background to the situation and the components needed to put Interactive Management into practice, the paper describes how idea generation and structuring methods were used by the team in setting and structuring objectives over a two-year time horizon. Initially, Nominal Group Technique was used to generate, clarify, edit and rank a set of over 40 objectives for the nine member team concerned with training development for the engineering industry. These objectives were then classified into groups representing different sub-systems of activity. A subset of the objectives, selected on the basis of the ranking scores, was structured during a computer-assisted Interpretive Structural Modelling session to produce an Intent Structure showing how the objectives interrelated. The various sub-systems of activity within the Intent Structure were subsequently identified and their interactions examined. The products of the sessions gave a clear picture of the participant's thinking on the role and functions of the team and provided a clear basis for future activities. It is concluded that Interactive Management provides both a context and process for such group work. This enables a pooling of content knowledge by the participants and results in both planning products and significant learning amongst the participants

Resumo

Esta comunicação refere-se a uma aplicação de Gestão Interactiva no planeamento de uma equipa de desenvolvimento no âmbito do orgão de gestão da formação de uma indústria nacional. Começando com uma descrição breve dos bastidores da situação e dos componentes exigidos para se pôr em prática a Gestão Interactiva, a comunicação expõe como a equipa usou métodos de geração de idéias e de estruturação para estabelecer e estruturar objectivos num horizonte de dois anos. Inicialmente, a chamada «Nominal Group Technique» foi utilizada para gerar, clarificar, fixar e hierarquizar um elenco de cerca de 40 objectivos para a equipa de nove membros que tinham a seu cargo o desenvolvimento da formação da engenharia industrial. Estes objectivos foram classificados por grupos representativos dos diferentes sub-sistemas da actividade. Um sub-elenco de objectivos, seleccionado na base de uma classificação de importâncias, foi estruturado durante uma sessão de Análise e Estruturação de Modelos, <Interpretive Structural Modelling> com assistência de computador, visando a produzir uma Estrutura de Intenções demonstrativa da inter-relação dos objectivos. Os vários sub-sistemas de actividade no âmbito da Estrutura de Intenções foram seguidamente identificados e examinadas as suas inter-acções. Os resultados das sessões forneceram um quadro claro do pensamento dos participantes sobre o

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 367/374
1992, APDC & SPF, Lisboa

papel e as funções da equipa e facultaram uma base clara para as actividades futuras. Concluíu-se que a Gestão Interactiva assegura o contexto e o processo de trabalhos de grupo desta natureza. Consegue-se, deste modo, que os participantes ponham em comum conteúdos de conhecimento e, ao mesmo tempo, obtêm-se quer resultados de planeamento quer níveis significativos de aprendizagem entre quem participa.

1. Introduction

The work described here was carried out for a training development team of a national industrial training board. The Board is responsible for advising on, and monitoring, the training provided by companies in the engineering sector throughout the country concerned. The appointment of a new team manager presented an opportunity to rethink the direction of the work of the team. At his request, it was agreed that an Interactive Management approach [Warfield, 1984; Broome and Keever, 1986] be used to initiate planning for the team through a two-day planning workshop. After exploring various possibilities with the manager, it was decided to develop a set of objectives for the team with regard to what they should be trying to achieve over the next two years to examine how those objectives were interrelated.

The team consisted of nine members, including the manager, and it was decided to involve them all in the planning workshop. The workshop sessions were designed to incorporate five components for effective group work in the context of Interactive Management [Christakis and Keever, 1984]:

Participant Group – the training development team members, being stakeholders with content knowledge relevant to the issue,

Facilitator – to manage the work of the group and guide participants through the steps of the methods used,

Methods – in this case Nominal Group Technique and Interpretive Structural Modelling for generating and structuring objectives,

Computer and Peripheral Equipment – to run software associated with the methods, and to provide display facilities and word processing capability,

Decision Support Room – offering an environment equipped for effective group work which would support the four components above.

The relations between these components are shown as a transitive digraph in Figure 1.

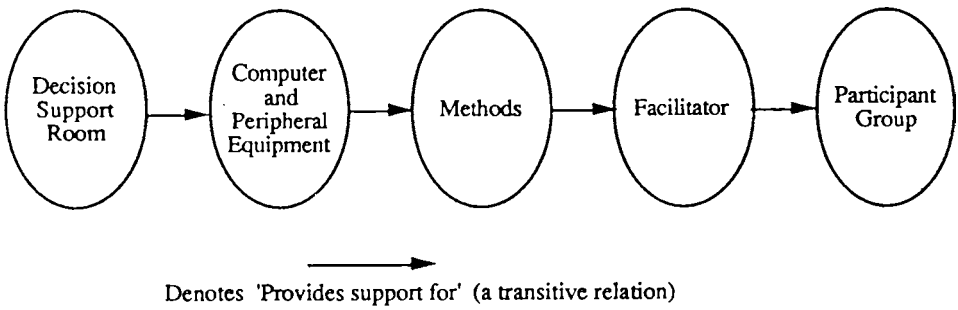


Figure 1. Relationships between the five components for effective group work

2. Generation of Objectives

2.1. Method

On the first day, the team participated in an idea–generation method, Nominal Group Technique (NGT) [Delbecq, 1975; Moore, 1987]. This was used to generate, clarify, edit and obtain a preliminary ranking of a set of objectives for the team. A classification scheme for the objectives was also developed.

Steps of Nominal Group Technique

1. Clarification of a trigger question to focus the idea generation.
2. Silent generation of ideas in writing in response to the trigger question.
3. Round–robin recording of ideas, whereby the ideas of participants are recorded on a flip chart and displayed.
4. Serial discussion of ideas for clarification, whereby each item is discussed in turn with the group in order to clarify and edit it.
5. Voting in order to enable participants to express their value judgements with regard to the importance of the ideas generated.

2.2. Resulting Objectives

The trigger question posed to the participants was:

'What should the training development team be trying to achieve over the next two years?'

A total of 45 objectives were generated and clarified during the NGT. They are not listed in full here, but a subset of them is shown in Figure 2. The numbers appearing before the objectives reflect the sequence in which the ideas were generated and are not of any other significance. The NGT voting scores obtained in step 5 do not give an accurate ranking. They provide a rough and ready assessment of the participant's judgements concerning relative importance of the objectives. Given time constraints this enables the facilitator to select a subset of items to be subsequently structured in the time available.

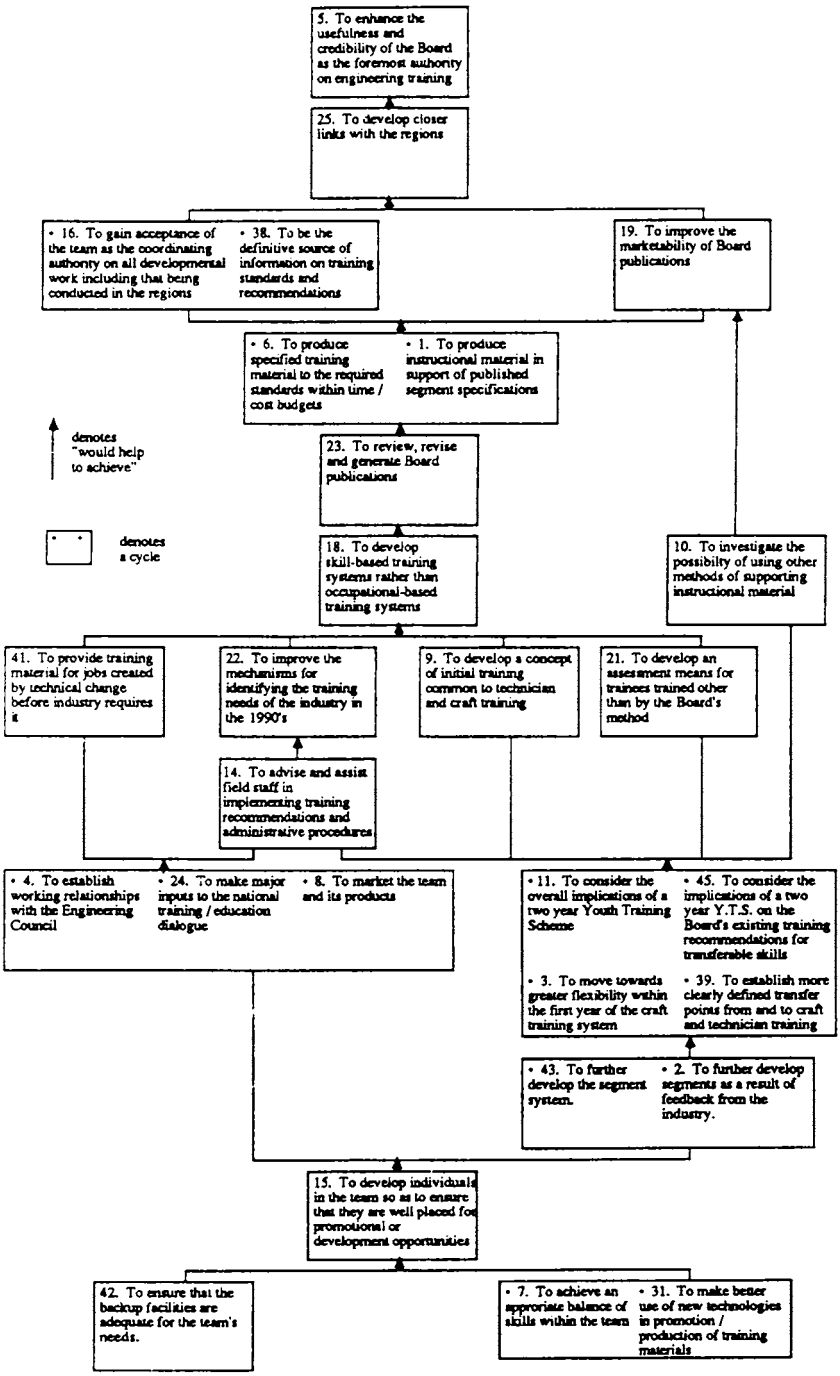


Figure 2: Intent Structure: Training Development Team

Following the NGT, the entire set of objectives was classified into groups as shown below:

. Ideals
. Public Relations
. Marketing
. Thinking Ahead
. Team Building
. Production of Material
. Development of Training System
. Regional Linkages

3. Structuring the Objectives

3.1. Method

In order to examine the relationships amongst the objectives, the team participated in a process called Interpretive Structural Modelling (ISM) [Warfield, 1976; Janes 1988]. This is a computer-aided method which helps participants to examine the interactions between the elements in a complex problem or situation. It enables a group to pool their thinking and reach agreement on the structure of the problem. In this case an ISM session was carried out on the second day, taking the objectives developed during the NGT as the elements to be structured. The resulting Intent Structure is outlined below.

3.2. Developing an Intent Structure

An Intent Structure shows how the accomplishment of one objective might contribute to the achievement of another. In this case, given the limited time available, 28 of the objectives were selected for structuring on a basis of the NGT voting scores. The ISM process required the team to respond to a series of questions put by the computer of the form:

'WOULD ACCOMPLISHING THE OBJECTIVE
23. To review, revise and generate
Board publications
HELP TO ACHIEVE THE OBJECTIVE
6. To produce specified training
material to the standards required
within time/cost budgets?'

As with the NGT, a two-year time horizon was assumed. After discussion, the answer to each question was fed in to the computer, the results of a majority vote being used whenever the participants were not unanimous.

During an ISM session, the responses of a group are fed into the computer which draws logical inferences from the answers as the process proceeds. Gradually, the computer builds up an ISM 'map', portraying the group's perceptions of the interrelations between the objectives. The map can be displayed at any stage of the process, examined by the group and amended by hand as desired. The Intent Structure in Figure 2 shows the completed 28 element map constructed by the training development team.

3.3. Interpretation of the Intent Structure

The boxes in the map contain the objectives with the original NGT numbering scheme. The arrows between the boxes represent the relation 'would help to achieve'. The Intent Structure thus shows what helps to achieve what.

Paths. A sequence of objectives connected by arrows is known as a path on the map. For example, the path 14–>22–>18 is explicitly shown. This may be interpreted as a statement that objective 14 helps to achieve objective 22 and that 22 helps to achieve 18. However, the transitive nature of the map means that 14 may help to achieve directly any elements which it reaches via a path of one or more arrows, e.g. objectives 22, 18, 23, 6, 1, 16, 38, 19, 25, 5.

Cycles. There are a number of cycles on the map indicated by multiple objectives in a box. These denote highly interdependent clusters of objectives which tend to reinforce each other.

Objective groupings. The Intent Structure may be partitioned (Figure 3) into a number of sub–systems as indicated by the shaded areas around groups of the objectives. It can be seen that the sub–systems broadly agree with the groups into which the objectives were originally classified following the NGT.

3.4. Uses of the Intent Structure

Within the context of the training development team, the uses of such an Intent Structure include the following:

Making explicit the multiple objectives of the team and their management implications.

Clarifying thinking, through asking the team systematically to think through the complex interactions between the objectives.

As a way of explaining the purpose of the team those within the training development division, the Board and outside agencies.

As a way of assessing and reporting progress. Many objectives may be followed at the same time and effort may be switched between them.

As a base from which to change objectives as new ideas are developed or as circumstances change.

As a basis of resource allocation. Given limited resources, the structure can assist in deciding on the allocation of resources to those activities which will be most productive.

To assist in organizational design. The Intent Structure can provide a basis for developing an appropriate organizational design.

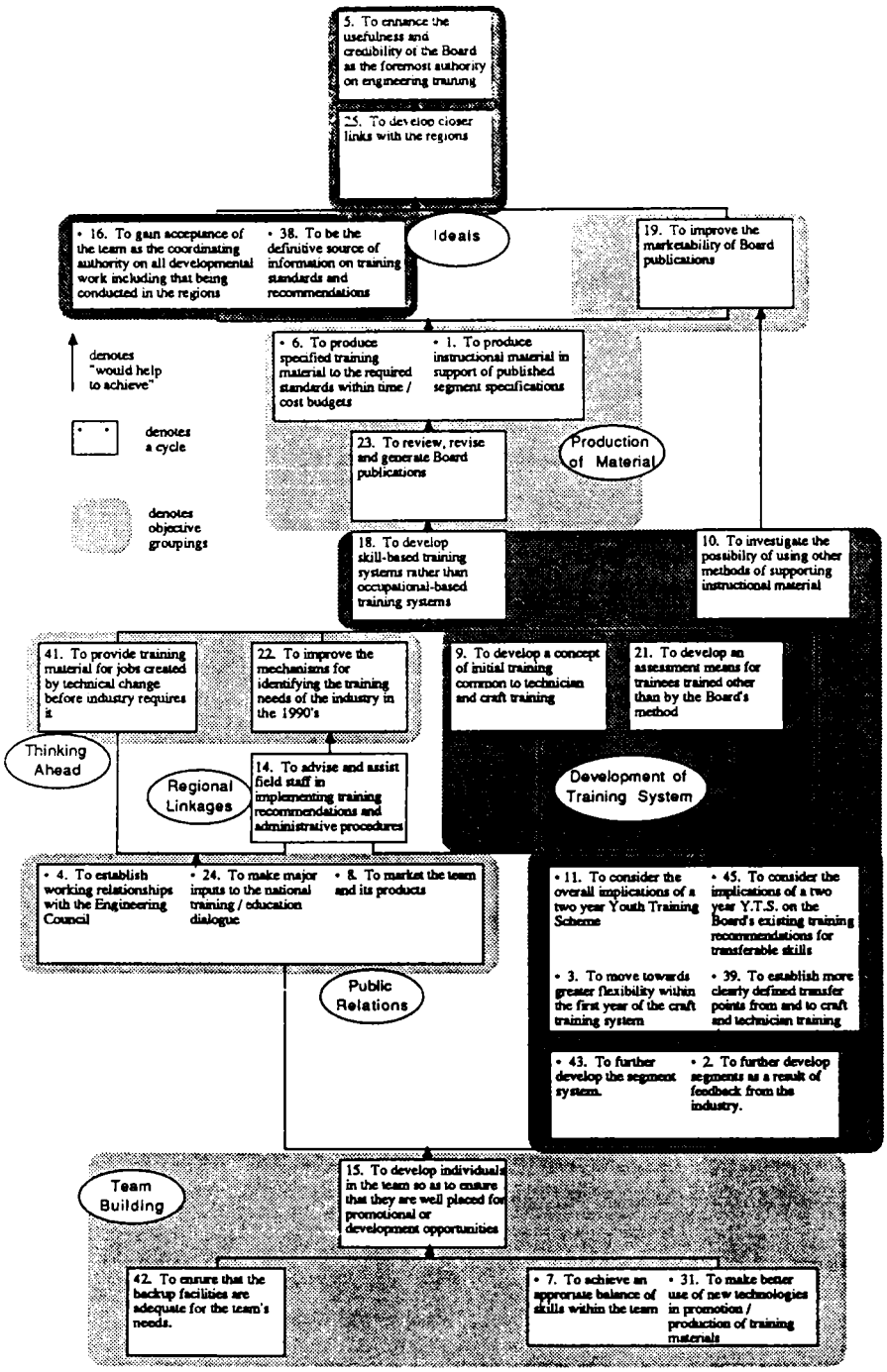


Figure 3: Sub-systems in Intent Structure

4. Conclusions

The five components consisting of: participant group; facilitator; methods; computer and peripheral equipment; and decision support room, provided a *context* for effective group work. The idea generation and idea structuring methods in the form of Nominal Group Technique and Interpretive Structural Modelling enabled participants to pool their *content* knowledge and work efficiently. The methods provided a *process* which focussed debate, encouraged structured argument and clarified thinking. This enabled *products* to be generated in the form of objectives for the team and an Intent Structure showing the interrelations between the objectives. The resulting learning and team building may also be regarded as products of the workshop. The usefulness of these various products in the subsequent work of the team was later confirmed by the team manager.

References

Broome, B.J. and Keever, D.B. (1986) *Facilitating group communication: the interactive management approach.* Paper presented at the annual convention of the Speech Communication Association, Chicago, November, 1986.

Christakis, A.N. and Keever, D.B. (1984) *Management's quest for new solutions.* Unpublished paper, The Centre for Interactive Management, George Mason University, Virginia, USA.

Delbecq, A.L., Van de Ven, A.H. and Gustafson, G.H. (1975) *Group techniques for program planning.* Scott, Foresman and Company, Glenview, Illinois, USA.

Janes, F.R. (1988) *Interpretive Structural Modelling: a methodology for structuring complex issues.* Trans. Inst. Meas. and Cont. Vol. 10, Nº. 3, 145–154.

Moore. C.M. (1987) *Group techniques for idea building* Sage Publications, London, England.

Warfield, J.N. (1984) *Progress in interactive management.* Proc. 6th Int. Cong. World Organisation of General Systems and Cybernetics, AFCET: Paris, XXIX–XXXV.

Warfield, J.N. (1976) *Societal systems: planning, policy and complexity,* John Wiley & Sons Inc, New York, USA.

James Terence Coulter Wright

Fac. de Economia e Administração, Univ. de São Paulo
São Paulo (Brasil)

Análise e Estruturação de Modelos (ISM) aplicado à seleção de projectos de pesquisa

Resumo

Este trabalho apresenta um caso de aplicação da técnica de Análise e Estruturação de Modelos (ISM) no estabelecimento de um sistema de avaliação e seleção de projectos na Divisão de Pesquisa e Desenvolvimento de uma empresa do setor energético. A técnica ISM é um método de trabalho em grupo auxiliado pelo computador para a estruturação de modelos de problemas complexos. A partir da estruturação de uma árvore de objetivos pelos gerentes da Divisão de P & D, foram identificados fatores e seus respectivos pesos para um sistema de avaliação de projetos, com grande grau de participação, compreensão do problema e consenso sobre os objetivos da Pesquisa entre os gerentes envolvidos.

Abstract

This paper reports on an application of <ISM> Interpretative Structural Modelling tecnique in the implementation of a projects evaluation/selection system in the R & D Division of a company operating in the energetic sector. The ISM technique is a work group method aided by computer for structuring complex problems models. The factors for a projects-evaluation system and their respective relevance are identified on the basis of a goals tree structured by the R & D Division's managers, with high levels of participation, of problem understanding and of consensus amongst all the managers concerning the research goals.

I. Apresentação e objectivos

Este artigo apresenta uma aplicação da técnica de Análise e Estruturação de Modelos ao desenvolvimento de uma sistemática para avaliação e seleção de projetos de pesquisa e desenvolvimento. A sistemática foi desenvolvida para aplicação na Divisão de Pesquisa e Desenvolvimento de uma empresa do setor energético, tendo possibilitado a realização da seleção de projetos a partir de uma visão consensual dos gerentes de projetos da empresa acerca dos critérios de avaliação e seleção dos mesmos. Pretendemos aqui dar uma visão geral do sistema de avaliação desenvolvido e, em especial, da contribuição a este resultado dada pela aplicação da técnica de Análise e Estruturação de Modelos.

II. A técnica de Análise e Estruturação de Modelos

A Análise e Estruturação de Modelos (ISM) – Interpretative Structural Modelling – é uma técnica para o trabalho em grupo em situações complexas. Um computador é utilizado para auxiliar um grupo de trabalho a comunicar, compreender e modelar uma questão

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 375/383
1992, APDC & SPF, Lisboa

complexa, sendo extremamente útil em situações como a estruturação de objetivos, definição de prioridades e a hierarquização de políticas e diretrizes (Warfield, 1976). Esta técnica permite estruturar um grande número de informações não quantitativas, contribuindo para a análise e tomada de decisão ao nível de políticas e diretrizes, onde os modelos quantitativos em geral não captam muitos dos condicionantes críticos dos problemas que representam.

Uma das maneiras tradicionais de lidar com estes problemas é de formar grupos de trabalho incluindo indivíduos capacitados nas diversas disciplinas envolvidas. Uma dificuldade que surge nesta situação é a de comunicação entre especialistas que, em função de sua formação académica e experiência profissional, aplicam aos trabalhos conceitos e linguagem que não são comuns a todos os membros do grupo. Por outro lado, especialistas de alto nível em geral têm pouco tempo disponível para uma efetiva integração em um grupo de trabalho. Ademais, a estrutura de comitê apresenta uma série de deficiências para o trabalho de investigação coletiva. A capacidade de persuasão aliada ao prestígio acadêmico ou profissional de um indivíduo tende a dar peso desproporcional a seus argumentos, independentemente de sua validade real; as pressões de grupo também tendem a abafar as posições minoritárias e o próprio processo de comunicação verbal entre os elementos do grupo torna-se bastante moroso quando é grande o número de participantes (Wright, 1984).

Surgem também dificuldades inerentes à própria complexidade, pois o ser humano tem a capacidade de lidar simultaneamente com um número limitado de variáveis, estimando-se que a nossa memória de curto prazo usualmente consegue "manipular" cerca de 5 a 7 "itens" de informação onde cada "item" compõe-se de uma variável e suas relações com as demais variáveis.

Com vistas a superar estes problemas, a técnica ISM busca, inicialmente, extrair do contexto em estudo um conjunto de componentes ou elementos do problema. Usando a seguir um computador para ordenar, armazenar e fazer inferências lógicas sobre o relacionamento entre os elementos identificados constrói-se um modelo estrutural do problema considerado. O modelo resultante é representado por um gráfico direcionado, obtido através da aplicação de noções básicas de topologia, somadas ao julgamento humano do grupo e a capacidade dos computadores de processarem grande número de dados e fazerem inferências lógicas. Os gráficos direcionados são na verdade estruturas de uso bastante comum; apenas, em geral, não as consideramos desta forma, como mostra o quadro a seguir:

QUADRO I
ESTRUTURAS COMUNS, USUALMENTE REPRESENTADAS
POR GRÁFICOS DIRECIONADOS

ELEMENTOS E RELAÇÕES COMPÕEM: ESTRUTURAS		
Pessoas	subordinado a	organograma
Variáveis	em função de	modelo matemático
Atividades	precede	rede PERT
Objetivos	apoia	árvore de objetivos

O conjunto das informações necessárias à definição de uma estrutura deste tipo pode ser matematicamente representado por uma matriz binária, e uma operação matemática simples (potenciações booleanas sucessivas) permite transformá-la de forma a evidenciar de modo claro todas as inferências e níveis hierárquicos contidos nas informações originais (Warfield, 1976). Numa nova abordagem ao problema de estruturação, a técnica computacional utilizada realiza inferências lógicas diretamente sobre a árvore de inferências existente, e utiliza um procedimento de busca heurística para definir a sequência de perguntas para coleta de informações.

A atuação do grupo de trabalho utilizando ISM consiste em, definido o contexto do problema, extrair deste um conjunto bem definido de elementos pertinentes.

A seguir, o grupo analisa a existência de uma relação transitiva entre pares de elementos, respondendo a perguntas colocadas pelo computador, que faz as inferências lógicas possíveis, produzindo assim o modelo estrutural. A definição clara dos componentes do problema e de suas relações, a interpretação e o aperfeiçoamento do modelo estrutural são de extrema valia para melhorar o entendimento do grupo sobre a questão, sistematizar o raciocínio dos integrantes e facilitar a comunicação sobre o problema. É interessante observar que o grupo não precisa ter nenhum conhecimento sobre os procedimentos matemáticos realizados, analisando apenas o relacionamento entre pares de elementos do problema. Por sua vez, o computador faz inferências lógicas com base apenas nas informações fornecidas pelo grupo, minimizando o número de perguntas necessárias à determinação do modelo estrutural.

III. Condicionantes do caso

A avaliação e seleção de projetos de pesquisa e desenvolvimento apresenta uma série de condições específicas que a diferenciam do processo utilizado para seleção de outros tipos de projetos. Uma característica fundamental desse processo é que ele necessariamente deverá apresentar a sensibilidade necessária aos aspectos estratégicos e à multiplicidade de objetivos inerentes à seleção de um projeto em um ambiente de pesquisa. Neste caso, em que desenvolveu-se um sistema específico para a Divisão de Pesquisa e Desenvolvimento de uma grande empresa do setor energético, os múltiplos objetivos que devem ser contemplados quando da formação de uma carteira de projetos na área de energia alternativa implicam na necessidade de um sistema de avaliação e seleção capaz de lidar com uma grande variedade de projetos de natureza e escalas distintas e que visam atender em diferentes graus os múltiplos objetivos da organização (Twiss, 1972).

Em adição a estas condicionantes há de se considerar que a grande parte das ações de desenvolvimento dos projetos selecionados decorrerão por prazos dilatados no futuro, trazendo ao processo de análise todas as dificuldades e incertezas, associados à previsão de resultados futuros, particularmente difícil em uma área de pesquisa, em que o conhecimento de resultados futuros é por definição incompleto. Outro aspecto a considerar é que além das diferenças de escala, há diferenças significativas entre os projetos em consideração, com relação a seu estágio no ciclo de inovação, o que traz necessidades distintas para análise dos projetos que se encontram em diferentes estágios. Finalmente há a condicionante de que o processo desenvolvido seja praticável em prazos relativamente curtos, de forma que este tipo de avaliação e seleção possa ser repetido periodicamente, em geral em base anual, servindo assim de apoio direto ao próprio processo de planejamento, orçamentação e controle das atividades da organização.

IV. Resultados alcançados

Foi desenvolvido neste trabalho uma metodologia, que é especificada a seguir, para o processo de avaliação e seleção de projetos para a divisão de pesquisas de uma empresa do setor energético. O método proposto baseia-se no sistema de "Escores", onde cada projeto é avaliado em relação a um conjunto de atributos para os quais são definidos pesos relativos de forma a permitir uma avaliação geral, ordenação e comparação entre os projetos propostos. O processo desenvolvido procura superar limitações geralmente observadas nos métodos de "Escore" que são a dificuldade e subjetividade associado à definição dos pesos dos diferentes atributos e a subjetividade na avaliação destes atributos em cada projeto. Essas limitações são superadas pela amarração da sistemática de definição de pesos e de avaliação de atributos à prévia definição e discussão detalhada de uma estrutura de objetivos. Essa estrutura de objetivos permite a ordenação e análise da importância relativa de cada um dos objetivos da organização, permitindo um estabelecimento mais direto dos pesos dos atributos. Estes por sua vez foram selecionados de listas de atributos mais comumente utilizados neste tipo de processo, agrupados e relacionados aos diferentes conjuntos de objetivos da divisão de forma que cada atributo foi diretamente relacionado ao objetivo específico para o qual deverá contribuir.

Este procedimento, que envolve também a discussão detalhada dos objetivos e clara definição dos atributos, permite redução substancial do grau de subjetividade em que normalmente é feito este tipo de avaliação.

Eliminar totalmente a subjetividade significaria em princípio, trabalhar com critérios estritamente objetivos, quantificáveis e conhecidos, o que numa área de pesquisa implica em abandonar qualquer conceito de adequação estratégica, atendimento a objetivos não quantitativos e não financeiros, passando a trabalhar com níveis de certeza que na realidade inexistem.

O processo de avaliação proposto permite pois uma análise individual por projeto em relação a cada um de uma série de atributos uniformemente definidos e relacionados com os objetivos próprios da organização. Por outro lado, o processo viabiliza o estabelecimento de um sistema de ponderação dos atributos, essencial para a utilização do modelo, e que é feito de maneira bastante objetiva, relacionando-se diretamente os atributos à importância relativa dos diferentes objetivos.

Finalmente o processo apresenta como resultado adicional, por si só de grande utilidade, o estabelecimento da estrutura de objetivos através da metodologia de Análise e Estruturação de Modelos (ISM), resultado da participação e da visão conjunta de todos gerentes da empresa em um processo de discussão e troca de informações que esclarece e encaminha as etapas subsequentes de definição de prioridade e seleção dos projetos. A estrutura completa de objetivos da Divisão é apresentada na Figura 1 (infra) a partir da qual foram definidos os 8 Macro-objetivos utilizados no processo de seleção e avaliação de projetos.

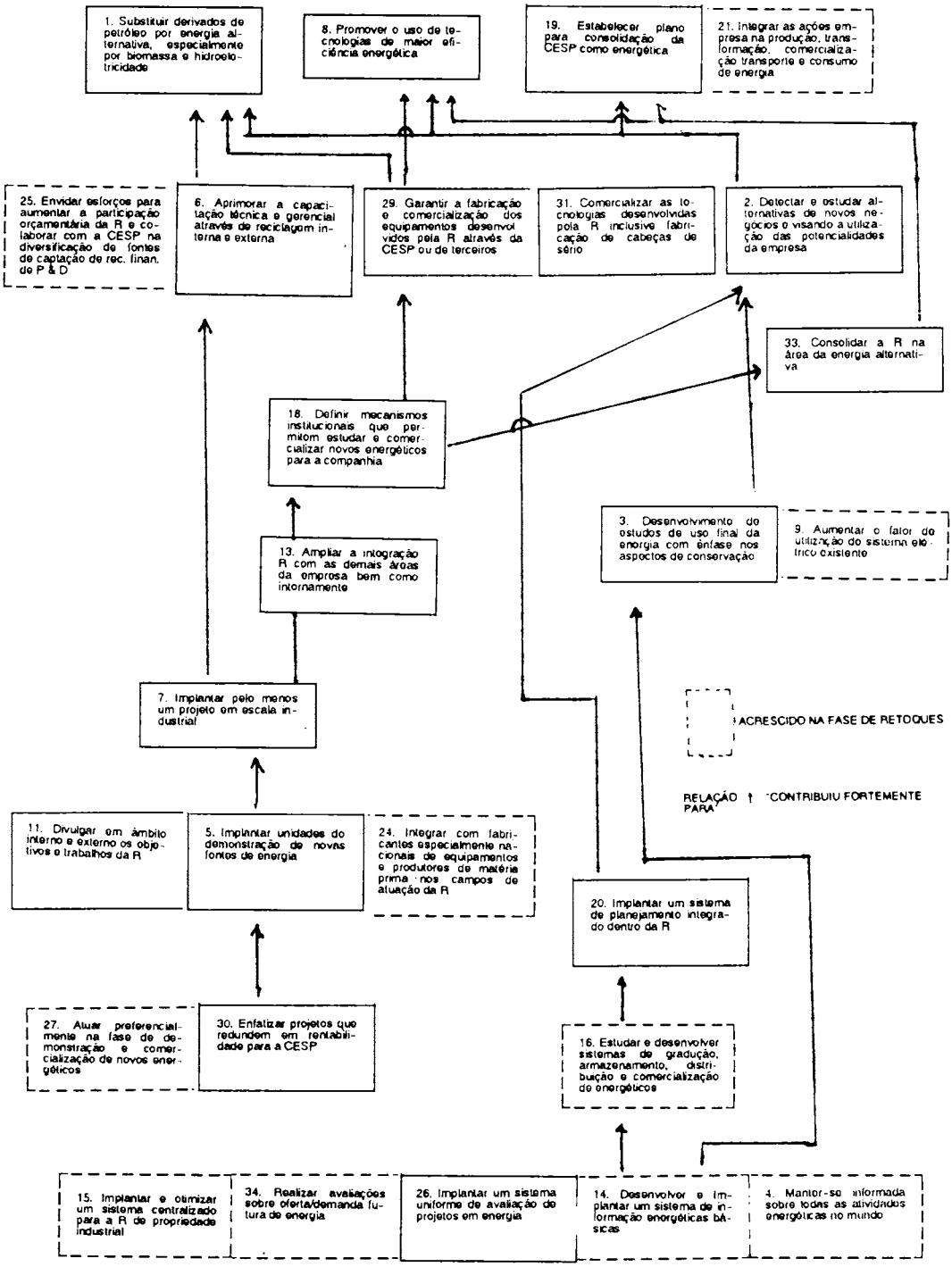
V. Sistema de avaliação resultante

A figura 2 ilustra o fluxo de atividades necessário para a aplicação da metodologia desenvolvida. Apresentemos a seguir as principais atividades desse fluxo.

1. Revisão de Diretrizes/Objetivos

O ponto de partida para a aplicação da sistemática de avaliação e seleção de projetos é a revisão de diretrizes e objetivos da organização, realizada mediante o processo ISM.

FIGURA 1

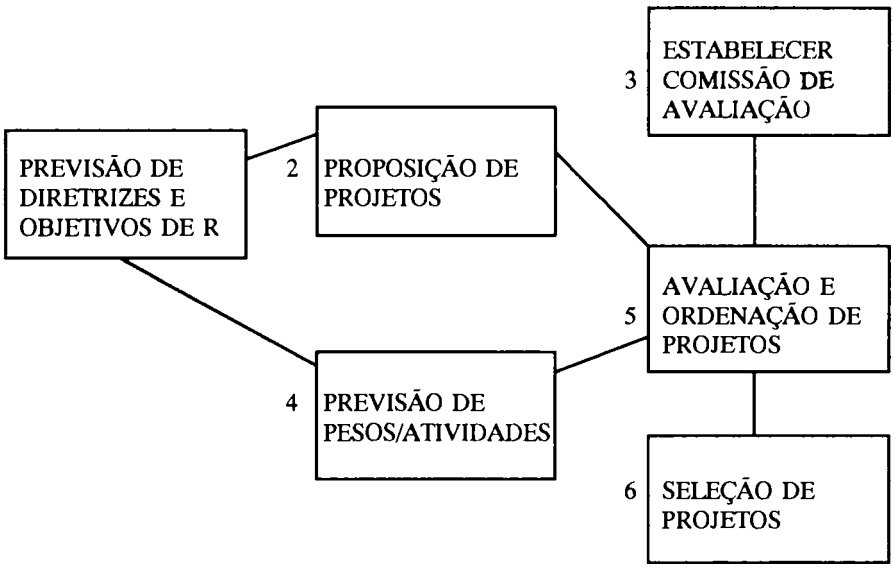

1. Substituir derivados de petróleo por energia alternativa, especialmente por biomassa e hidroeletricidade
8. Promover o uso de tecnologias de maior eficiência energética
19. Estabelecer plano para consolidação da CESP como energética
21. Integrar as ações empresa na produção, transformação, comercialização transporte e consumo de energia
25. Envidar esforços para aumentar a participação orçamentária da R e colaborar com a CESP na diversificação de fontes de captação de rec. finan. de P & D
6. Aprimorar a capacitação técnica e gerencial através de reciclagem interna e externa
29. Garantir a fabricação e comercialização dos equipamentos desenvolvidos pela R através da CESP ou de terceiros
31. Comercializar as tecnologias desenvolvidas pela R inclusive fabricação de cabeças de série
2. Detectar e estudar alternativas de novos negócios e visando a utilização das potencialidades da empresa
33. Consolidar a R na área da energia alternativa
18. Definir mecanismos institucionais que permitem estudar e comercializar novos energéticos para a companhia
3. Desenvolvimento de estudos de uso final da energia com ênfase nos aspectos de conservação
9. Aumentar o fator de utilização do sistema elétrico existente
13. Ampliar a integração R com as demais áreas da empresa bem como internamente
7. Implantar pelo menos um projeto em escala industrial
ACRESCIDO NA FASE DE RETOQUES
RELAÇÃO ↑ "CONTRIBUIU FORTEMENTE PARA"
11. Divulgar em âmbito interno e externo os objetivos e trabalhos da R
5. Implantar unidades de demonstração de novas fontes de energia
24. Integrar com fabricantes especialmente nacionais de equipamentos e produtores de matéria prima nos campos de atuação da R
20. Implantar um sistema de planejamento integrado dentro da R
27. Atuar preferencialmente na fase de demonstração e comercialização de novos energéticos
30. Enfatizar projetos que redundem em rentabilidade para a CESP
16. Estudar e desenvolver sistemas de gradução, armazenamento, distribuição e comercialização de energéticos
15. Implantar e otimizar um sistema centralizado para a R de propriedade industrial
34. Realizar avaliações sobre oferta/demanda futura de energia
26. Implantar um sistema uniforme de avaliação de projetos em energia
14. Desenvolver e implantar um sistema de informação energéticas básicas
4. Manter-se informada sobre todas as atividades energéticas no mundo

Deve-se verificar possíveis mudanças no macro-ambiente da empresa, e alterações de objetivos e estratégias que possam afetar as diretrizes de pesquisa e desenvolvimento, introduzindo novos elementos e eliminando ou modificando elementos existentes. Não se espera que as diretrizes e objetivos sofram grandes modicações de um ano para outro, embora a importância relativa dos objetivos deva ser revisada anualmente.

A revisão de diretrizes e objetivos, feita pela gerência com ou sem a participação dos técnicos da divisão, deve ser amplamente divulgada e discutida com estes, a fim de orientar a atividade 2, de proposição de projetos, para que estes se tornem cada vez mais consistentes com as diretrizes da divisão.

FIGURA 2

FLUXO DE ATIVIDADES DO PROCESSO DE AVALIAÇÃO E SELECÃO DE PROJETOS



2. Proposição de Projetos

O segundo passo na análise e seleção de projetos requer a elaboração e consolidação de uma lista de projetos de pesquisa aplicada e desenvolvimento para o período em consideração. A lista deve conter proposições para projetos novos, sugeridos de forma descentralizada pelos técnicos, pelos gerentes da organização ou por outras áreas da empresa. Também deve-se listar projetos existentes, onde se propõe continuidade para o próximo exercício. A geração desta lista deve refletir os objetivos e prioridades estabelecidas no primeiro passo.

3. Estabelecimento de uma Comissão de Avaliação dos Projetos

Para realizar o processo de avaliação propriamente dito dos projetos, isto é o processo de atribuição de valores ou notas relativas a cada atributo dos projetos, deverá ser estabelecida uma comissão encarregada desta atividade. Recomenda-se que esta comissão seja formada por gerentes dos diversos departamentos ou pelos próprios chefes desses departamentos.

4. Revisão dos Pesos dos Atributos

Uma vez analisada a estrutura de objetivos da divisão, os pesos alocados a cada conjunto de atributos deverão ser revistos, particularmente se houver mudanças de conteúdo ou de prioridade entre os objetivos da Divisional.

A revisão deverá ser feita mediante processo de análise colectiva dos pesos dos atributos realizados pela gerência do Departamento. A Figura 3 apresenta o resultado obtido em uma aplicação indicado o peso relativo dos objetivos e dos atributos a serem avaliados que são associados a estes objetivos.

FIGURA 3
PESOS DOS ATRIBUTOS PARA AVALIAÇÃO DE PROJETOS

OBJETIVOS	PESO	ATRIBUTOS ASSOCIADOS AOS OBJETIVOS DA R	AVALIAÇÃO 0 - 5
Consolidar a empresa como companhia energética	32	ESTRATÉGICOS O projeto é compatível com o objetivo estratégico de consolidação da empresa?	
Substituir derivados de petróleo por outras fontes de energia, inclusive hidroelétrica	26	MERCADO Qual a penetração no mercado possível em termos do volume equivalente de petróleo?	
Aumentar a eficiência do uso final dos energéticos			
Aumentar o fator de utilização do potencial elétrico existente			
Contribuir para a rentabilidade da Empresa	17	FINANÇAS Qual o nível de retomo sobre o investimento? (TIR, Benefício anual líquido, valor atual, pay-back, etc)	
Estudar e desenvolver sistemas de produção, armazenamento, transporte, distribuição e comercialização de energéticos	12	PESQUISA E DESENVOLVIMENTO Qual a probabilidade de sucesso técnico? Qual o custo de Pesquisa e Desenvolvimento?	
Identificar e estabelecer programas de uso múltiplo das águas	9	USO MÚLTIPLO DAS ÁGUAS O projeto pode contribuir para o uso múltiplo das águas em sua área de aplicação?	
Não degradar o meio ambiente	5	ECOLOGIA O projeto contribue para minorar os impactos ecológicos do uso de energia?	

5. Avaliação dos Projetos

Cada um dos projetos em consideração deverá ser apresentado à comissão de avaliação pelo gerente proponente ou responsável pela proposição do referido projeto. Esta apresentação deverá ser feita através de uma descrição por escrito, cobrindo os aspectos considerados pelos atributos de julgamento, juntamente com uma apresentação verbal sucinta pelo proponente para a comissão de modo a facilitar a troca de informações que dará o embasamento à avaliação do projeto.

A comissão de avaliação fornecerá notas referentes a cada um dos atributos do projeto de maneira que seja possível o estabelecimento de uma nota total ponderada pelo peso de cada atributo. A nota dada a cada atributo será representada pela média aritmética das notas dos avaliadores.

Com base na somatória ponderada das notas dadas aos diversos atributos dos projetos será possível fazer uma ordenação simples dos projetos. Há de se considerar nesse processo que há um certo grau de imprecisão de modo que as diferenças muito pequenas em notas de projetos não devem ser consideradas como indicações definitivas de prioridade relativas.

6. Seleção de Projetos

A tarefa de seleção de projetos será feita com base na lista ordenada de projetos propostos para início ou manutenção, com justificativas bastante objetivas para a sua ordenação referentes à estrutura de objetivos da organização. Por outro lado há de se considerar que em um processo seletivo será necessário identificar um ponto de ruptura abaixo do qual projetos não serão aceites. Esta situação ocorre fundamentalmente em função da limitação de recursos, admitindo-se que a maioria dos projetos propostos originalmente tenham qualidades positivas. O procedimento adotado deverá certamente estabelecer avaliações mais elevadas para aqueles projetos que melhor se coadunem com os objetivos da organização.

VI. Conclusões

Uma característica fundamental que distingue o processo usado é o estabelecimento prévio de uma árvore de diretrizes e objetivos. Esse trabalho, realizado em conjunto pelos gerentes da Divisão de P & D, permitiu que o resultado fosse diretamente relacionado com as peculiaridades e necessidades específicas da Divisional. O uso da metodologia de ISM para o trabalho de grupo que produziu essa estrutura viabilizou um processo de discussão, análise e formação de consenso e respeito da topologia da estrutura que contribui fortemente para a aceitação futura do processo de avaliação e seleção de projetos.

A estrutura completa de objetivos e diretrizes é apresentada na Figura 1. Desta estrutura com 27 objetivos, os 8 objetivos da Figura 3 foram extraídos para representar a síntese dos objetivos e diretrizes gerais da empresa. Vários objetivos foram eliminados por se tratar de objetivos específicos de um só projeto, ou de estratégias específicas de ação. A relação de objetivos e diretrizes foi apresentada e discutida na última reunião de metas do ciclo de planejamento, quando os pesos também foram atribuídos a estes.

Há ainda duas considerações importantes a fazer sobre aplicações desta metodologia ao processo de seleção de projetos. Primeiro, a aplicação sistemática deste processo poderá orientar o próprio processo de identificação e formulação de projetos. Os projetos desenvolvidos com vistas à sua adequação aos critérios de seleção serão necessariamente projetos que orientam-se para o atendimento dos objetivos da organização e consideram explicitamente os importantes atributos de potencial do mercado, retorno financeiro,

adequação estratégica, etc. Consideramos que a proposição contínua de projetos com esta orientação durante um período de tempo significativo deverá trazer contribuições importantes no aprimoramento da carteira de projetos da organização.

A segunda consideração é de que este aprimoramento da carteira de projetos resultará também na adoção de um procedimento em que a ordenação dos projetos e a definição de prioridades entre os projetos implicará necessariamente em que alguns projetos que vêm sendo realizados e continuamente executados no decorrer dos anos venham a ter sua realização questionada. Consideramos que este questionamento será bastante saudável a medida em que o estabelecimento de uma linha de corte abaixo da qual os projetos não serão aprovados, mesmo que esta linha não possa ser inicialmente imposta de maneira rigorosa, significa que o processo de alocação de recursos será claramente orientado no sentido de reservar os recursos disponíveis para aqueles projetos que são mais adequados aos objetivos da organização, obrigando-a a deixar de executar aqueles projetos que não mais mostram contribuição aos objetivos explícitos da organização, que tornaram-se bastante claros para todos os envolvidos, através da aplicação da técnica de Análise e Estruturação de Modelos.

Bibliografia

Twiss, Brian. *Managing Technological Innovation*, Longman Group Limited. London, 1972.

Warfield, John N. *Societal Systems, Planning, Policy and Complexity.* John Wyley e Sons Ltd., New York, 1976.

Wright, James T.C. *A Técnica de Análise e Estruturação de Modelos Aplicada ao PROALCOOL.* Revista de Administração Vol. 19 (3) Julho/Setembro 1984 pp. 58 a 62.

Ohuchi, Azuma et alii. *An efficient Procedure for Transitive Coupling in ISM, IEEE Transactions on Systems*, Man and Cybernetics, vol. SMC-15, nº 3, May/June 1985.

Ilona Kovács

Núcleo de Ciências Sociais, Inst. Sup. de Economia-UTL, Linha de Acção "Implicações Sociais" do Grupo de Robótica Inteligente - UNL, Consultora na CESO I&D.

António Brandão Moniz

Dep. de Ciências do Desenvolvimento, Fac. de Ciências e Tecnologia - UNL,
Linha de Acção "Implicações Sociais" do Grupo de Robótica Inteligente-UNL, Director da CESO I&D.

As tendências tecnocêntricas e antropocêntricas na concepção de sistemas integrados de produção

Resumo

Nesta comunicação são apresentadas as características da automação flexível, nomeadamente as do sistema de Produção Integrada por Computador (CIM). Confrontam-se as perspectivas tecnocêntrica e antropocêntrica no desenvolvimento de sistemas integrados de produção. Procura-se evidenciar as razões da persistência da perspectiva tecnocêntrica.

Analisa-se ainda a relação entre tecnologias flexíveis e desenho organizacional. Procura-se demonstrar as vantagens do modelo organizacional flexível face ao modelo burocrático-rígido no que se refere à realização das potencialidades de automatização flexível e à capacidade de enfrentar novos problemas.

Por fim apresentam-se algumas considerações sobre as potencialidades da intervenção teórica e prática do sociólogo no desenvolvimento de sistemas integrados de produção.

Abstract

This paper reports on the features of flexible automation, concerning namely the CIM - Computer Integrated Manufacturing system. We compare the tecnocentric and antropocentric views within CIM development. We try to show why the technocentric view persists.

We analise the relationship between flexible technology and organizational design. We try to show the advantages of the flexible organizational model compared with the rigid-bureaucratic one for the accomplishment of flexible automation potential and the capability to face new problems.

Finally we consider the sociologist's potentialities to intervene theoretically and pratically in CIM systems development.

> "Não deveremos repousar-nos, esperando ingenuamente que talvez no futuro a Tecnologia Avançada de Produção seja tão automatizada que tornará irrelevantes os aspectos humanos. Trata-se precisamente do contrário; os problemas humanos e organizacionais serão cada vez mais importantes"[1].

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 384/393
1992, APDC & SPF, Lisboa

I - Características da Produção Integrada por Computador (CIM)

A automatização flexível dos sistemas de produção significa uma modificação profunda na maneira de produzir no que diz respeito quer ao aspecto tecnológico quer ao organizacional. O Sistema de Produção Integrada por Computador, mais conhecido por CIM-*Computer Integrated Manufacturing* constitui o estádio mais avançado de duas das tendências fundamentais da modernização industrial:

a) a integração e controlo global e sistemático por computador das diversas operações da concepção, do processo produtivo e da gestão;

b) a extensão da automatização para além das operações de fabricação.

Apesar de este tipo de sistema produtivo ainda se encontrar muito pouco divulgado no mundo inteiro[2], como é sublinhado por diversos autores, existem diversos estudos retrospectivos - e alguns prospectivos - referentes aos aspectos humanos e organizacionais das tecnologias avançadas de produção, entre elas do sistema **CIM**.

Quais são os objectivos estratégicos a atingir quando se pretende desenvolver um projecto CIM? Por um lado, pretende-se identificar padrões de integração produtiva baseada em conceitos de sistemas abertos e, por outro, desenvolver métodos e instrumentos associados a estes conceitos. Pretende-se desenvolver ainda subsistemas **CIM** que sejam capazes de funcionar neste quadro. Por fim, outro dos objectivos estragégicos deste projecto é demonstrar o sucesso desta abordagem, demonstrando os benefícios deste sistema **CIM** mesmo que seja implementado desde já num tipo já largamente variado de ambientes de produção. É possível implementar um sistema CIM numa empresa já existente, mas um tal sistema apenas poderá funcionar numa unidade mais autonomizada dessa mesma empresa[3].

Este tipo de estragégia é o que aliás tem vindo a ser seguido sobretudo pelo Programa **ESPRIT**, no Sub-programa **5 CIM**[4], até porque os parceiros mais importantes de consórcios **ESPRIT** são em geral empresas com significativo poder económico e que poderão assim fazer desenvolver experiências tecnológicas e organizacionais nessas unidades automatizadas. Apesar disso, é possível implementar de raíz experiências de automação integrada a partir do amadurecimento de experiências assim feitas. Essa implementação só poderá vir a ser feita no final deste século.

Um dos aspectos fundamentais do sistema **CIM** está relacionado com o seu *carácter integrado*. Assim, este sistema possui uma lógica própria (contrária à do taylorismo) que caracteriza todo o fluxo informativo e produtivo: a integração de todas as actividades da empresa (comercial, administrativa, produtiva, de aprovisionamento, de marketing, etc.). Como é referido por J. Child, a integração abrange duas dimensões. Uma será "a *dimensão física* em que a transformação e transferência de material e componentes pode aproximar ao conceito de fluxo contínuo. A segunda, a *dimensão informacional* ou de gestão, em que as actividades de marketing, concepção, engenharia, produção e contabilidade atingem um elevado nível de coordenação no interesse de se atingir uma resposta flexível e económica às oportunidades do mercado"[5]. Por outras palavras: as diversas áreas de actividade assistidas por computador encontram-se interligadas em rede. Deste modo o **CIM** representa o nível de integração mais elevado das diversas actividades da empresa.

A interligação das diversas áreas de actividade é feita com base em tecnologias de informação (TI), desde os primeiros níveis de integração (concepção assistida por computador, máquinas-ferramentas, **AGV's** e robots, armazéns flexíveis, etc.)[6] até, obviamente, ao nível da integração final **CIM**. Com efeito, outro autor afirma que "a tecnologia aplicada em **CIM** faz uso intensivo da distribuição de redes de computadores e

técnicas de processamento de dados, Inteligência Artificial e Sistemas de Gestão de Base de Dados''[7].

O outro aspecto básico do CIM é a *extensão da automatização*. Esta refere não apenas à automatização dos próprios processos de fabricação, mas também aos sistemas de informação e controlo através do qual o trabalho é planeado, organizado e levado a cabo[8].

O sistema CIM significa ainda o aperfeiçoamento de uma nova tendência que surgiu com a utilização de máquinas programáveis: *o aumento conjunto do grau de automatização e da flexibilidade*. Anteriormente, o aumento do grau de automatização engendrava maior rigidez, limitando-se à produção de um número reduzido de produtos estandardizados segundo o imperativo da quantidade e produtividade (economia de escala) - foi a era da chamada automatização convencional ou rígida. Pelo contrário, o CIM é a forma mais avançada da automatização flexível com a vantagem de permitir uma maior «performance» económica produzindo produtos diversificados, individualizados, frequentemente renovados com a melhoria simultânea da produtividade e da qualidade, cumprindo prazos de entrega. Torna-se assim possível uma espécie de *neo-artesanato* baseado numa tecnologia sofisticada que permite responder a uma procura caracterizada pela grande diferenciação e individualização. Uma das vantagens capitais da automatização flexível é precisamente essa flexibilidade técnica, uma vez que no actual contexto económico é uma questão vital a capacidade de adaptação rápida às mudanças do mercado numa concorrência acrescida.

II - O factor humano nos sistemas integrados de produção

O desenho do sistema produtivo implica opções quanto aos factores humanos e aos aspectos organizacionais. Concretamente implica opções quanto à repartição de tarefas, informações e decisões entre gestores, engenheiros, técnicos (programadores) e operadores. Coloca-se, por conseguinte, o dilema da especialização/polivalência, e da centralização/descentralização, quer de bases de dados, quer das linhas de comunicação. Assim, a implementação de sistemas de controlo descentralizados implica nomeadamente necessidades para o estabelecimento de uma arquitectura específica de sistema aberto para CIM.

Se consideramos o conceito de ''factores humanos'' em toda a sua amplitude (ergonomia e sociologia geral), facilmente compreendemos que há uma variedade infinita de questões e de problemas que podem ser colocados, pois se entende normalmente por *factor humano* a interface homem/máquina. De igual modo, também poderemos compreender que os mais diversos tipos de disciplinas científicas tentarão responder a essas questões (tal como a psicologia social, a sociologia industrial, a economia do trabalho, a engenharia de sistemas, a sociologia das organizações, ou a antropologia industrial, etc.).

No estado actual do debate podemos dizer que um dos problemas que preocupam neste momento os engenheiros e outros cientistas em projectos CIM tem a ver, não apenas com as responsabilidades sociais dos out-puts deste projecto, mas com o saber até que ponto estarão os engenheiros alertados para os factores humanos (nas suas diversas dimensões) quando pretendem estabelecer eficientes sistemas integrados de produção assistida por computador.

E, de facto, ao nível dos projectos CIM podemos encontrar estes dois tipos de abordagem: um, que compreende a interface indivíduo-máquina apenas do ponto de vista instrumental, ou ergonomista, e outro, que procura entender esta interface do ponto de vista da participação nas organizações.

O primeiro tipo de abordagem tende a minimizar o papel do factor humano e das suas condições sociais. Esta abordagem tende para o controlo centralizado do sistema. O

segundo tipo de abordagem tende, pelo contrário, a valorizar o factor humano, entendendo-o como essencial para o funcionamento eficaz do sistema.

Como exemplo desta última abordagem, podemos referir uma área do projecto **ESPRIT** sobre "Factores Humanos e Interfaces Homem-máquina"[9], tendo como objectivo o melhoramento destas interfaces de modo a facilitar o controlo local de sistemas de fabricação (FMS-Flexible Manufacturing Systems, FAS-Flexible Automated Systems, sistemas de controlo de processo, etc.). Subentende-se, por conseguinte, uma opção pela polivalência e por uma maior descentralização das decisões partindo duma perspectiva segundo a qual "a integração de um sistema TI[10] é um problema de ordem social e técnica que afecta profundamente o elemento humano. São necessários modelos e métodos para analisar e conceber sistemas sócio-técnicos integrados e para avaliar a relação organização-sistema"[11].

Ainda de acordo com o Programa de Trabalho mais recente do projecto **ESPRIT** sublinham-se estes aspectos quando se refere que "a área **Gestão e Controlo de Processos de Produção Industrial** diz respeito a Tecnologias de Informação que apoiam uma maior mudança nos métodos de gestão da produção, dos sistemas de controlo centralizado da produção para estruturas de controlo flexíveis e distribuídas. Isto também envolve importantes mudanças organizacionais que deveriam ser conduzidas com cuidadosas considerações do seu impacte sobre ambientes de operação, estruturas organizacionais e relações de trabalho entre organizações"[12].

O desenho organizacional dos sistemas **CIM** constitui, por conseguinte, um desafio aos cientistas sociais, nomeadamente aos sociólogos. É, por isso, fundamental a consciencialização para a importância dos factores humanos e sociais e para a existência e as consequências das opções organizacionais possíveis. Principalmente, é importante que os sociólogos se dêem conta das possibilidades e potencialidades da sua participação em projectos de tecnologias avançadas de produção (AMT's).

O desenho dos sistemas produtivos depende da maneira como os indivíduos e a empresa são encarados, concretamente do quadro sócio-organizacional em que os equipamentos são utilizados. Os resultados obtidos, quer em termos psico-sociais quer económicos, serão bem diferentes de acordo com a perspectiva adoptada. Quando se concebem e/ou adoptam sistemas técnicos, há opções basicamente em torno de duas perspectivas diferentes. Uma é a *perspectiva tecnocêntrica* que privilegia os aspectos técnicos e marginaliza os aspectos humanos, encara a empresa como um mero local de produção de bens e de serviços. O próprio trabalhador não passa de um instrumento de produção programável por especialistas e controlável por chefes. Assim a tecnologia e a organização tendem a ser desenhados de acordo com princípios como: especialização e simplificação do trabalho, controlo hierárquico, centralização de informações e decisões, formalização de conhecimentos e procedimentos. Por conseguinte, as capacidades especificamente humanas para reagir ao imprevisto e inesperado, a capacidade de inovação não são valorizadas ou aceites[13].

A outra é a *perspectiva antropocêntrica* orientada para a valorização do factor humano que considera o factor técnico numa perspectiva instrumental e encara a empresa como não apenas um local de produção de bens e de serviços mas igualmente como espaço de produção social e cultural, de expressão de subjectividade, de identidade. O trabalhador é encarado como um ser activo e autónomo que em vez de ser submetido a um controlo autoritário, é capaz de criatividade e auto-controlo. Assim a tecnologia e a organização procuram-se desenhar com base em princípios como: trabalho interessante e criativo, autonomia, desenvolvimento pessoal e profissional, descentralização, participação.

O quadro seguinte indica as principais diferenças entre as duas perspectivas:

Perspectiva tecnocêntrica	Perspectiva antropocêntrica
Meios técnicos adquirem-se para concentrar o potencial de controlo sobre a produção (sistema de auto-controlo)	Meios técnicos adquirem-se como ferramentas para realizar funções rotineiras deixando a regulação superior à intervenção humana
Marginalização do factor humano por via da automatização máxima	Valorização do factor humano por via da combinação das capacidades humanas com as potencialidades da automatização
Polarização das qualificações, redução do espaço de qualificação	Aumento do espaço de qualificação
Redução da autonomia dos operadores	Aumento de autonomia e responsabilidade no nível operacional
Simplificação do trabalho	Diversificação do trabalho
Crescente formalização do saber e sua incorporação no software	Espaço de intervenção e de utilização do saber para prevenir incidentes
Não utilização dos conhecimentos e capacidades dos trabalhadores	Conjunção do saber empírico e teórico
Centralização das informações, do controlo	Descentralização das informações, do controlo

A nosso ver, apenas com base numa perspectiva antropocêntrica se torna possível desenvolver *sistemas produtivos caracterizados* por uma *nova lógica organizacional*: fluidez; mobilidade; adaptabilidade; fluxo de produção e de informações–comunicações; interligação e integração de operações, de postos de trabalho, de funções, de unidades funcionais, do saber–fazer empírico com o saber teórico; cooperação, participação, formação contínua. Porém existe uma forte tendência para a persistência da perspectiva tecnocêntrica e para a utilização de tecnologias flexíveis em quadros organizacionais rígidos, isto é, caracterizados pela forte especialização do trabalho, hierarquização, centralização e formalização.

Quais as razões dessa persistência?

Por um lado, o desenho do sistema produtivo tende a ser feito numa perspectiva que privilegia os aspectos técnicos (estes planeiam–se primeiro e os aspectos humanos posteriormente têm de ser ajustados) devido principalmente ao tipo de *formação* daqueles *especialistas* (informáticos, analistas de sistemas) que concebem os sistemas técnicos avançados e/ou adoptam elementos técnicos a formas organizacionais existentes. A formação destes especialistas é deficiente, uma vez que "o lado social ou sociorganizador da sua tarefa não foi considerado", e, por conseguinte, não os capacita para a aplicação de conhecimentos relativos ao sistema social quando concebem ou adoptam um sistema técnico. Essa deficiência deve–se à "consciência predominantemente tecnocrática dos estabelecimentos tradicionais que são as universidades, nas disciplinas técnicas e na

informática''[14]. Além disso, os especialistas em ciências sociais não são chamados para participarem no desenho do sistema produtivo por serem considerados pelos empresários, gestores e engenheiros como ignorantes dos imperativos técnicos e económicos.

Essa tendências para o desconhecimento do factor humano parece ser tanto mais forte quanto maior fôr o grau de automatização das actividades produtivas. No caso do sistema **CIM** parte-se geralmente da visão de um sistema produtivo ''sem pessoas''[15] que justifica a concentração nos aspectos técnicos. É frequente encontrar-se esse tipo de pressuposto nos modelos de integração automatizada da produção. Trata-se de tentar criar *um sistema produtivo informatizado, integrado* controlado *centralmente* por meios informatizados, em que a auto-regulação tem por base a formalização e incorporação máxima de todo o saber-fazer empírico no *software*. Procura-se substituir o trabalho humano pela tecnologia que incorpora saberes regularizados e formalizados.

A perspectiva tecnocêntrica e a organização rígida tendem a perpetuar-se também devido às *relações de poder*; particularmente nas grandes empresas, os detentores do poder de controlo têm interesse em manter ou até em aumentar esse poder com os novos sistemas produtivos. O modelo organizacional tradicional justifica-se frequentemente como uma necessidade inevitável, uma vez que os grandes investimentos em novos sistemas produtivos devem ser controlados de modo rigoroso por especialistas e de modo algum podem ser confiados aos operadores. Geralmente os próprios vendedores destes sistemas sublinham esse facto.

São, portanto, esses os mecanismos essenciais que agem no sentido da opção tecnicista e da persistência da lógica da rigidez apesar das exigências de flexibilidade organizacional. Não é, no entanto, a única estratégia possível e realizável.

Com efeito, como diversos estudos demonstram, a importância dos aspectos humanos e organizacionais aumenta na medida em que o sistema se torna mais complexo e integrado e, por conseguinte, mais vulnerável, mais frágil e mais sensível quanto aos factores de incerteza (não previstos, não rotineiros) incontroláveis por automatismos. O reconhecimento deste facto indica uma abordagem não-tecnicista que parte de outro género de pressupostos.

Neste caso, trata-se de criar um *sistema produtivo altamente flexível e descentralizado* em que se complementam as potencialidades da tecnologia informatizada com as capacidades específicamente humanas, tais como a capacidade de gestão do imprevisto realizada pela intervenção humana baseada em informações não formalizáveis e/ou a capacidade de compreensão de situações complexas incertas e não estruturadas. O equipamento informático é entendido como uma ferramenta de informação e de organização para tarefas rotineiras que deixa espaço para o saber e iniciativa dos indivíduos e grupos. A flexibilidade aumenta na medida em que se utilizam o saber-fazer individual e colectivo e as capacidades existentes dos trabalhadores com alto nível de qualificação (conjunção de uma formação teórica e de conhecimentos tácitos mais amplos).

Um exemplo disso pode ser o projecto Esprit 1217 (1199) ''Human Centred **CIM** Systems'', quando se refere explicitamente nos seus objectivos que ''um sistema **CIM** será mais eficiente, mais económico, mais robusto e mais flexível se uma pessoa é directamente responsável, do que num sistema comparável sem pessoas''[16].

III - Desenho organizacional dos sistemas CIM

A relação entre as *novas tecnologias* e o tipo de *organização* constitui uma forte controvérsia no domínio da Sociologia Industrial[17]. Podemos, por conseguinte, distinguir três posições que têm prevalecido nesse debate:

1. – A introdução de sistemas automatizados leva a uma organização neo-taylorista, neo-fordista e, por conseguinte, a uma maior centralização das decisões.

Os ritmos de trabalho, as prescrições dos modos operatórios, as regras de funcionamento integram-se no logicial (ou "software"). A produção do logicial feita por engenheiros de sistemas e serviços informáticos aparece como a objectivação superior da concepção[18].

2. – A introdução de sistemas integrados é incompatível com as formas tradicionais de organização. Assim passa-se de uma organização rígida para uma organização flexível.

A introdução de sistemas integrados é incompatível com as formas tradicionais de organização uma vez que implicam a dissolução de fronteiras entre trabalho manual e intelectual, entre fabrico e manutenção, a imprecisão e falta de rigidez na definição das tarefas, bem como o fluxo de informação. Assim passa-se de uma organização hierarquizada, formalmente integrada e rígida para uma organização menos hierarquizada, orgânica e flexível (equipas móveis e polivalentes com mobilidade espacial e intercâmbio de funções)[19].

3. – É possível escolher entre diferentes modelos de organização com base na mesma tecnologia

Existirão sempre alguns constrangimentos nas opções organizacionais do sistema **CIM**, no entanto, como sempre, é também possível escolher entre diferentes modelos de organização com base na mesma tecnologia. As diversas opções implicam não apenas diferentes consequências sociais, mas também resultados económicos diferentes. A organização mecânica-burocrática tem uma lógica que, como vimos, caracteriza-se pela rigidez, forte formalização, centralização, polarização de qualificações (forte divisão entre especialistas altamente qualificados que organizam e exploram o sistema e operadores desqualificados cujo trabalho é pobre, reduz-se à vigilância). A grande fraqueza dessa opção reside nos seguintes aspectos:

a) fraca motivação e empenhamento dos operadores,

b) inflexibilidade, isto é, na sua incapacidade de fazer face a situações complexas e em permanente mutação, uma vez que não faz uso do saber produtivo não formalizado, não regularizado dos indivíduos, não deixa espaço para uma intervenção autónoma,

c) fraca capacidade ou até incapacidade de evitar falhas, avarias e/ou acidentes. As deficiências latentes que o desenho do *software* implica não podem ser corrigidas por operadores limitados por conhecimentos especializados e fragmentados ligados a postos de trabalho e tarefas isoladas; não são capazes de enfrentar situações incertas e novas num contexto complexo.

Diversos estudos – hoje já considerados como clássicos – têm demonstrado que quanto maior fôr a instabilidade tecnológica e do mercado (quanto mais diversificados os produtos e quanto menor for a previsibilidade), melhores "performances" podem ser obtidas com base numa escolha por um tipo mais flexível de organização. Este tipo de organização é caracterizável por uma fraca hierarquização, por um baixo nível de formalização de regras, de procedimentos e do sistema de controlo, pelo trabalho em grupo, pela polivalência, descentralização e participação nas decisões[20]. As suas grandes vantagens residem nos seguintes aspectos:

– elevada motivação e empenhamento dos operadores;
– rápida solução dos problemas;
– capacidade de evitar avarias e/ou acidentes por parte de operadores qualificados e polivalentes com aptidão para um diagnóstico adequado com base numa percepção global, um controlo amplo e uma compreensão teórica profunda.

Assim não poderemos dizer que o sistema **CIM** comporta um determinado tipo de organização social do trabalho. Evidentemente, existirão alguns constrangimentos nas opções organizacionais deste sistema, mas, como sempre, é possível existirem diversos modelos de organização de trabalho sendo dada a mesma tecnologia. No entanto, o aproveitamento da eficácia económica potencial deste tipo de sistema exige também a implementação de inovação social[21] orientada para o desenvolvimento de organizações mais flexiveis e participativas com base numa perspectiva antropocêntrica. Esta solução responde melhor as exigências das novas tecnologias nomeadamente à necessidade de "adquirimos uma cultura de aprendizagem, uma compreensão dos conhecimentos tácitos, uma receptividade face aos processos interpessoais e uma capacidade de valorizar as nossas próprias opções em matéria de organização. Não por paradoxal deixa de ser certo que, na medida em que aperfeiçoamos as tecnologias mais avançadas, matemáticas e abstractas, dependemos mais de tipos não estruturados de formação, desenho e comunicação"[22].

Porém, a escolha entre essas duas orientações não é uma questão da tecnologia, mas depende dos valores que orientam o desenho tecnológico e organizacional, dos interesses e forças políticas e sociais em presença, nomeadamente das estratégias seguidas pelos actores sociais (poder político, sindicatos, empresários) e do conteúdo das negociações. Por outras palavras, a divisão do trabalho, o conteúdo das tarefas, a organização do trabalho, as qualificações requeridas, as relações de trabalho não são determinadas pela tecnologia mas constituem *várias estratégias*.

IV – O campo de intervenção possível dos sociólogos

O campo de intervenção dos sociólogos no contexto do desenvolvimento de novos sistemas produtivos pode incidir sobre:

. a participação na concepção dos novos sistemas produtivos e em experiências piloto;
. a avaliação segundo a perspectiva sociológica de sistemas avançados já em funcionamento **(CAD/CAM, FMS)**;
. a formulação de conclusões generalizáveis quanto aos critérios humanos/sociais para sistemas integrados.

O objectivo dessa intervenção é influenciar a direcção do desenvolvimento de novos sistemas produtivos no sentido de promover a qualidade de vida no trabalho e contribuir para a melhoria da eficácia global (técnica, social e económica) destes sistemas com base numa perspectiva antropocentrica. Esta acção exige a colaboração dos sociólogos com empresários, gestores, engenheiros, operadores e sindicalistas. Exige-se para isso uma capacidade de comunicação, de aprendizagem e de cooperação em abordagens multidisciplinares.

O facto dos aspectos humanos e sociais tenderem a ser subvalorizados tem a ver não apenas com a subsistência de uma mentalidade tradicional profundamente marcada pelo taylorismo, mas também com as atitudes dos próprios sociólogos. De acordo com Chris Clegg e Martin Corbett, "tem de ser dito que os próprios cientistas sociais contribuiram substancialmente para a sua escassa influência. Os cientistas sociais falharam na persuasão dos outros quanto à legitimidade e significância dos aspectos humanos da Tecnologia

Avançada de Produção. (...) Estas falhas por parte dos cientistas sociais têm a ver com os objectivos da investigação, com as ênfases e com o seu estilo''[23].

Os projectos de investigação sobre **CIM** - que na Europa se desenvolvem apenas ao nível internacional no âmbito do Programa **ESPRIT**, como já foi focado - constituem uma oportunidade para a contribuição das Ciências Sociais. Vários grupos de investigação têm demonstrado a possibilidade dessa participação[24].

Notas

(1) Clegg, Chris e Corbett, Martin: ''Research and Development into 'Humanizing' Advanced Manufacturing Technology'' in Wall-Clegg-Kemp (eds): *The Human Side of Advanced Manufacturing Technology*, Chichester, John Wiley & Sons Ltd, 1987, p. 192.

(2) Entre os diversos autores que têm estudado esta questão referiremos, por exemplo, os que desenvolveram mais este aspecto: Blumberg, Melvin e Gerwin, Donald: *Coping With Advanced Manufacturing Technology*, (Discussion papers IIM/LMP 81-12), Berlim, IIMV-WZB, 1981. Brödner, Peter: ''Skill Based Manufacturing vs. 'Unmanned Factory'- Which is Superior?'', *International Journal of Industrial Ergonomics*, 1 (1986), Amesterdão, Elsevier Science Publ., pp. 145 - 153. Jones, Bryn: ''Segmented Isles in a **CIM**-less Sea - Social Aspects of Flexible Automation: The British Case'', comunicação apresentada à *International Conference on ''Social Problems of the Introduction of Flexible Automation''*, Turim, Vienna Centre/ILO Turin Centre, 3-5 Sep 1987, 36 pp. Martin, T. (ed.): *Design of Work in Automated Manufacturing Systems*, Oxford, IFAC-Pergamon Press, 1984. Meredith, Jack R.: ''The Strategic Advantages of the Factory of the Future'', *California Management Review*, 29 (3), 1987, pp. 27 - 41. Ranky, Paul G.: ''FMS in **CIM** (Flexible Manufacturing Systems in Computer Integrated Manufacturing)'', *Robotica*, Vol. 3, 1985, pp. 205-214.

(3) A maior parte das experiências **CIM** descritas têm sido implementadas deste modo, e não criadas de raíz como normalmente deveria acontecer.

(4) Neste Sub-programa 5 estão já envolvidas algumas equipas portuguesas, em particular o Grupo de Robótica Inteligente da Universidade Nova de Lisboa, nos projectos 278 ''Integrated Sensor-Based Robot System'', e 623 ''Operational Control for Robot System Integration Into **CIM**''.

(5) Child, John: ''Organizational Design for Advanced Manufacturing Technology'' in Wall, T.B.; Clegg, C.W.; Kemp, J.J. (eds): *The Human Side of Advanced Manufacturing Technology*, Chichester, John Wiley & Sons Ltd, 1987, p. 117 [itálico nosso].

(6) Sobre alguns conceitos referentes a tecnologias de sistemas automatizados: cf. Moniz, António B. e Teixeira, J. Pamies: ''Pequeno glossário de robótica'' in Moniz, A.B. et al.: *Efeitos sociais....*, op. cit., 1986, pp. 21 - 26; cf. Scott, Peter B.: *The Robotics Revolution*, Oxford, Basil Blackwell, 1984, pp. 312-331; cf. Hunt, V. Daniel: *Industrial Robotics Handbook*, Nova Iorque, Industrial Press, 1983, pp. 391-44; cf. Bonetto, Roger: *Les ateliers flexibles de production*, Paris, Hermes Pub., 1985, pp. 263-268.

(7) Ranky, Paul G.: ''FMS in **CIM**...'', op. cit., p. 205.

(8) Jones, Althea e Webb, Terry: ''Introducing Computer Integrated Manufacturing'', *Journal of General Mangement*, Vol. 12, Nº 4, 1987, p. 61.

(9) cf. *Plano de trabalho ESPRIT 1987*, Bruxelas, ponto III.3.1.

(10) TI-Tecnologias de Informação - [nt].

(11) cf. *Plano de trabalho ESPRIT 1987,* Bruxelas, ponto III.3.1.1., "Objectivos".

(12) cf. *1989 ESPRIT Workprogramme (Draft),* Bruxelas, p. III.2.

(13) CF.: Kidd, P.T. e Corbett, J.M.: Towards the joint social and technical design of advanced manufacturing systems. *International Journal of Industrial Ergonomics*, Amsterdam, 2/1988, pp. 305–313.

(14) Wobbe, Werner: "Tecnologia, trabalho e emprego – Novas evoluções da reestruturação social". *Formação Profissional*, Nº 1/1987., Berlim, CEDEFOP, pp. 4–5.

(15) A "unmanned factory".

(16) Documento de apresentação do referido projecto Esprit 1217 (1199). Esta questão já foi mais desenvolvida no texto de Moniz, A.B.: "O Projecto **CIM–EUROPE** (Esprit II) e as possibilidades de intervenção sociológica na concepção e simulação de processos automatizadaos", *CTS–Ciência, Tecnologia e Sociedade*, Lisboa, ACTD, no prelo.

(17) cf. Kovács, Ilona e Moniz, António B.: "Organização do trabalho" in Kovács, Ilona (resp.): *Tendências actuais da mudança tecnológica e organizacional na indústria – Um Estado da Arte*, Lisboa, JNICT/FCT–UNL, 1987, pp. 101–110.

(18) cf. Braverman, Harry: *Trabalho e capital monopolista,* Rio de Janeiro, Zahar, 1977; cf. Chanaron, Jean–Jacques e Perrin, Jacques: "Science, technologie et modes d'organisation du travail" in *Sociologie du Travail,* n. 1/1986, Paris, Gauthier–Villars, pp. 23–40.

(19) cf. Veltz, Pierre: op. cit., 1986; cf. Coriat, Benjamin: *La Robótica,* Madrid, Revolución, 1985.

(20) cf. Burns, Tom e Stalker, G.M.: *The Management of Innovation*, London, Tavistock Publ., 1961; cf. Perrow, Charles B.: *Análise Organizacional – Um enfoque sociológico*, São Paulo, Atlas, 1981.

(21) cf. Kovács, Ilona: "Mudança tecnológica versus mudança organizacional ou processo de inovação integrado na indústria" in CISEP/ISE–UTL: *O comportamento dos agentes económicos e a reorientação da política económica*, Vol. I, Lisboa, CISEP/ISE–UTL, 1986, pp. 271–288.

(22) Ibid. p. 256.

(23) Clegg, Chris e Corbett, Martin: op. cit., p. 181.

(24) Entre outros, o grupo liderado por H. Rosenbrock na Universidade de Manchester–Institute of Science and Technology (UMIST), ou o grupo de Psicologia Social e Aplicada da Universidade de Sheffield (SAPU).

José N. Lacerda Fonseca
Univ. de Évora, Évora (Portugal)

Algumas Relações entre Tipos de Distribuição dos Conhecimentos e as Estruturas da Cooperação e Realização Material

Resumo

Discute-se a possibilidade de uma estrutura do conhecimento que possibilite uma distribuição social dos conhecimentos conducente à minimização dos processos de conflito socio-económico, relacionando essa estrutura com a possibilidade de realização material do indivíduo e os mecanismos de motivação para as actividades económicas. As idéias apresentadas são inseridas numa filosofia geral das acções, brevemente exposta.

Abstract

This paper reports on the possibility of a knowledge structure which could provide a social knowledge distribution preventing social and economic conflictive processes. The relations with individual material accomplishment and economic motivation are discussed. These ideas are considered within a general philosophy of action, presented in brief.

1. Introdução

Pretende-se aqui discutir a viabilidade de uma estrutura do conhecimento que possibilite uma distribuição social dos conhecimentos conducente à minimização dos processos de conflito socio-económico e à possibilidade de cabal realização material do indivíduo. Discute-se os seus efeitos nos mecanismos de motivação para as actividades económicas.

Tenta-se tornar patente a relação entre o referido tipo de distribuição do conhecimento e a telemática (fluidez informativa), pedagogias baseadas nas novas ciências cognitivas e a cibernitização do trabalho (restrição do trabalho humano a trabalhos criativos).

Com o fim de facilitar uma mais cabal explicação dos limites e dimensões das ideias a expôr elas são inseridas numa "filosofia global das acções", que aqui brevemente se apresenta.

2. A Filosofia das Acções

2.1. Grandes formas da filosofia das acções: Uma política de desejo e uma estrutura de função

No estado de sua máxima dimensão ideal uma filosofia global das acções pretenderia englobar todas as acções concretas possíveis enquanto input.

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 394/407
1992, APDC & SPF, Lisboa

Enquanto output terminal pretenderia englobar os estados materiais que directamente se possam relacionar com designações como "bem-estar" ou "objectivo global do indivíduo".[1]

2.2. A abordagem do ponto de vista abstracto das condições gerais de "construção"

Neste texto abordar-se-á a filosofia das acções na perspectiva das condições abstractas e gerais para que se tornem interessantes trajectos de "construção" por oposição a trajectos de "redisposição" (estagnação, mera redistribuição, ou destruição).[2]

2.3. A dominância do problema da cooperação

É hoje em dia patente um crescimento exponencial da capacidade de influenciarmos a natureza no sentido da nossa comodidade (tecnologia). É contudo duvidoso o desenvolvimento da nossa capacidade de organizar a sociedade da forma cooperativa passível de usar convenientemente essa tecnologia.

Tenta-se aqui estudar o desenvolvimento da cooperação através do estudo de processos práticos que o possam motivar e tornar patente.

2.4. Pequenas formas da filosofia das acções. Um discurso pragmático. Um discurso de proposta.

Apesar da filosofia das acções não se apresentar lógico-matematicamente, nem ser passível de comprovação experimental, tenta ser um discurso veiculador de propostas concretas, expressas com conceitos e sintaxes comprovadamente objectivos (ou melhor, largamente inter-subjectivos).

Tenta ser um discurso de propostas que possam ter os seus benefícios avaliados na prática, de forma a orientar a insistência nelas ou o seu abandono mais ou menos definitivo.

Tenta veicular propostas controláveis pelo desenrolar dos seus efeitos, e cujos eventuais efeitos negativos se encontrem dentro de limites comportáveis.

Ao inscrever-se em áreas de grandes benefícios eventuais torna-se um discurso cuja eficácia prática não terá de ser necessariamente inferior à do discurso científico.

Em suma é um discurso que para provar a sua utilidade não sente a necessidade de demonstrar que é um discurso comprovado experimentalmente (ou comprovável), nem que é estruturado de forma lógico-matemática[3].

3. Inserção dos temas a apresentar numa Filosofia Global das Acções

A inserção dos temas a apresentar numa filosofia das acções faz-se, como já foi referido, através da orientação da filosofia das acções para a questão das opções entre construção e redisposição. Esta orientação desenvolve-se na consideração de quais as condições materiais que optimizam o rendimento pessoal dos esforços feitos em trajectos construtivos e quais as que, pelo contrário, optimizam o rendimento de trajectos redispositivos.[4]

Como condições optimizadoras dos trajectos construtivos[5] podem-se à partida considerar as seguintes condições:

a) Existência de níveis suficientemente altos de capital (bens de investimento e matérias primas).
b) Informação e formação referentes a técnicas "construtivas" (técnicas produtivas, conhecimentos comerciais, técnicas relativas a serviços, etc.)

c) Convirá distinguir à parte a informação conducente à motivação psicológica do indivíduo para o trajecto construtivo. Por exemplo uma estrutura cultural conducente a uma pressão social que, através de imposição de sanções para os trajectos não construtivos, aumente relativamente o ratio benefício/sacrifício envolvido nos trabalhos construtivos.
d) A existência de organismos e processos ideológicos capazes de arbitrar a repartição pelos indivíduos dos bens que produziram em conjunto ou dos bens que trocam.
e) A inexistência de condições optimizadoras de redisposições é obviamente um factor optimizador dos trajectos construtivos. A consideração destes factores é pois sempre vista num quadro em que os níveis de optimização, de uma ou de outra tendência, são sempre relativos uns aos outros e portanto, em rigor, têm de ser simultaneamente pensados.
f) O valor pessoal e social dos bens que é possível produzir é obviamente de considerar no ratio de rentabilidade dos trajectos construtivos.

Um conjunto de bens vale frequentemente enquanto expressão de sucesso na luta de competição hierárquica[6], isto é, vale como símbolo de status ou de poder. O valor de um conjunto de bens é pois relativo à distribuição social dos bens. Obviamente não se pode omitir esta perspectiva "relativista" quando se reflecte sobre o valor dos benefícios proporcionados por um bem.

O cerne desta comunicação consiste em pensar os efeitos que a evolução de certas estruturas do conhecimento tem nos fenómenos expressos nos itens "c" (motivação para a construção), "d" (estruturas de acordo sobre a repartição social dos bens) e "f" (valor dos bens de consumo).[7]

4. Estruturas de Acordo e a Fluidez de Mercado

Interessa aqui mostrar quais os factores de que estará dependente um sistema de produzir acordos, sobre a repartição social dos bens, que não envolva os custos sociais provocados por "suspensões conflituais" (greves, lock-out, retracção do investimento).

4.1.Processos de conflitualidade

São os seguintes os factores que actualmente podem tornar possível a uma classe, ou sub-classe, profissional impôr conflitualmente um certo tipo de contrato social (contrato de repartição dos bens entre os seus produtores):

a) Capacidade de conluio entre si, que dependerá da facilidade que têm em comunicar uns com os outros.
b) Capacidade de aguentarem períodos de "suspensão conflituosa" da sua tarefa (greve, lock-out, retracção do investimento).[8]
c) Capacidade dessa classe impedir que indivíduos de outras classes passem a efectuar as tarefas que lhes cabem (capacidade de "impedimento). Capacidade esta que é obviamente fundamental durante períodos de suspensão conflituosa.
d) Possibilidade de alguma forma de ser favorecida pela lentidão do sistema de informação-formação (que prepara essa classe profissional) em responder ao mercado. Esta lentidão provoca subida das remunerações dessa classe devido a eventuais pressões da procura dos seus serviços sobre a oferta fornecida pelo sistema de informação-formação.
e) Capacidade de controlar directamente os preços dos produtos finais sem que os compradores estejam suficientemente organizados (associações de consumidores) para

responderem (com suspensão temporária de compra) ao preço caso o considerem excessivo. Embora este processo não seja em si um processo de suspensão conflitual vai, como se sabe, provocar como resposta suspensões conflituais (greves para reposição do poder de compra).

4.2. O conceito de fluidez de mercado e a geração de acordo

Chamemos fluidez de mercado ao cenário onde se verifica o seguinte:

– Quando uma classe profissional está a exigir uma remuneração "excessivamente" alta então indivíduos de outras classes passarão a efectuar a tarefa dessa classe (com remunerações mais baixas), evitando assim que essas outras classes se sintam prejudicadas pela "excessiva" parte do rendimento social que a primeira pretendia.(9)

Trata-se pois de um cenário em que a oferta para desempenhar as diversas profissões será sempre tão grande que, por sua vez, as suas remunerações serão sempre as mínimas necessárias para que alguma pessoa aceite efectuá-las.

O nível das remunerações será assim automaticamente regulado sem os prejuízos relativos às suspensões conflituais tornadas impossíveis pela fluidez.

Esta impossibilidade resulta do facto de qualquer greve poder ser automaticamente "furada" por indivíduos de outras classes e de qualquer retracção do investimento suscitar o aparecimento de neogestores oriundos de outras classes.(10)

4.3. Fluidez de mercados e grupos equilibrantes

Entenda-se por conjunto equilibrante, de uma dada tarefa, o conjunto de indivíduos que, ao poder passar rapidamente a desempenhar essa tarefa, tornam impossível que a classe profissional que a estava a desempenhar possa desenvolver "impedimentos" (vidé 4.1.c). A totalidade dos conjuntos equilibrantes constitui a condição para que o mercado seja fluido.

Note-se que se o conjunto equilibrante de uma dada profissão estiver todo concentrado numa só outra profissão e vice-versa então será difícil obstar a conluios entre essas duas classes profissionais. Ou seja esse "conjunto equilibrante" não seria realmente um conjunto equilibrante. Assim para evitar esses conluios o conjunto equilibrante terá de estar distribuído por diversas classes profissionais.(11)

4.4. Obstáculos à fluidez

4.4.1. Obstáculos à fluidez de mercado: capital

Além do óbvio obstáculo à fluidez do mercado que constitui a dificuldade em indivíduos passarem imediatamente para outras profissões temos de considerar os conhecidos obstáculos de capital. Isto é, mesmo que o indivíduo tenha acesso a toda a informação e eficácia de um capitalista isso não lhe dá acesso ao seu capital.

Contudo num sistema financeiro moderno não parece impossível remover esses obstáculos sem uma alteração radical do conceito de propriedade.

De facto se o capitalista considerar que a sua tarefa (gerir e arriscar o seu capital) se encontra a ser mal remunerada poderá conseguir-se que fique aberto aos projectos de outros neo-gestores que se proponham usar esse capital. Os neo-gestores pagariam aos possuidores do capital uma taxa que seria a taxa mínima que incentivasse as poupanças até aos níveis socialmente necessários.

4.4.2. Obstáculos à fluidez do mercado: estrutura assimilativa da informação activa de mercado

Agora a questão é saber como será possível conseguir que cada indivíduo possua outras formações profissionais além da sua.

Ou então conseguir que os desempregados constituam os grupos equilibrantes, sem que essa situação de desemprego constitua um problema individual e social.

4.4.2.1. Disponibilização da informação aos grupos equilibrantes

Para conseguir que cada indivíduo possua outras formações além da sua terá obviamente de se resolver a questão dos custos sociais de formação profissional[12].

Ora é notório que não se aplicam profissionalmente muitos dos conhecimentos aprendidos na escola. Contudo eles (conhecimentos de flexibilização) não deixam de ser importantes no sentido em que aumentam o leque de tarefas que o discente pode vir a desempenhar, isto é aumentam a sua capacidade de arranjar um emprego, aumentam o seu poder no mercado. Por outro lado parte do conhecimento fornecido na escola constitui base que permite a mais fácil reciclagem profissional. A questão é pois a de saber até que ponto uma escola mais próxima das reais necessidades de mercado poderia deixar de fornecer e medir os conhecimentos de flexibilização. Até que ponto não será rentável fornecer e medir os conhecimentos pró-reciclagem só quando eles forem realmente precisos para a reciclagem?

Questão próxima destas será a de saber (questão orientada a partir de uma mudança das estruturas das tarefas profissionais e não a partir da escola) até que ponto será possível estruturar as tarefas de forma a minimizar o conhecimento realmente utilizado por cada indivíduo no exercício da sua profissão. Isto é, até que ponto é sempre mais rentável que seja efectuado por um grupo de especialistas, a tempo parcial, aquilo que pode ser desempenhado por um indivíduo menos especialista, a tempo inteiro.

Finalmente note-se que para uma real fluidez não basta que a escola licencie indivíduos numa certa profissão. É necessário que ela prepare os indivíduos para as tarefas que em cada instante estão a ser socialmente exigidas, de forma a que, ao sair da escola o indivíduo esteja, imediatamente, apto a substituir outro indivíduo que executa um tipo específico de tarefas. É necessário que nesta escola (nomeadamente numa escola de empresários, que será vital) se efectuem análises aos processos que levam ao sucesso dentro de uma dada tarefa e que se tentem implementar estes num número suficiente de indivíduos (para que alguns comprovadamente os absorvam). Trata-se pois de uma pedagogia centrada no mimetismo do sucesso concreto. Isto pressupõe que a escola pode acompanhar os indivíduos que estão a exercer as profissões tão de perto que pode criar seus "alter-ego". Tal pressupõe obviamente mecanismos telemáticos sofisticados e custos consideráveis que permitam esta interface de transparência entre escola e mercado.

A questão da fluidificação informativa do mercado não se esgota contudo nestas questões da escola e da estrutura profissional das tarefas, abrange também a questão da possibilidade física de disponibilizar informação muito local ou temporal (efémera). Abrange ainda a questão do custo da informação com a consequente necessidade de implementar linhas de crédito facilmente acessíveis para a compra de informação.

Quer estes últimos problemas quer o resultante da existência de secretismos parecem problemas para os quais se encontram já implantadas dinâmicas sociais que tenderão progressivamente a resolvê-los.

4.4.2.2. Disponibilização de capacidades intelectuais nos grupos equilibrantes

Obviamente ligada à questão da formação profissional está a problemática levantada pelas diferenças intelectuais entre os indivíduos.

Para apurar da viabilidade de um mercado fluido ter-se-á de saber até que ponto, nas tarefas mais exigentes intelectualmente, será possível dispôr de um número de indivíduos (possuindo essas capacidades) consideravelmente maior do que o socialmente necessário para o desempenho efectivo dessa tarefa.

Note-se que para viabilizar uma fluidez de mercado também será necessário que os sobredotados, para uma dada profissão, estejam distribuídos por várias outras profissões (para que os grupos equilibrantes sejam efectivos).

Assim temos pois também de saber até que ponto será possível que sobredotados para alguma tarefa executem profissões onde não exerçam as suas capacidades especiais. Esta questão poderá aliás ser paralela à de saber até que ponto será aceitável para um sobredotado uma situação de desemprego (exercer "profissão 0") ou sub-emprego (nomeadamente através de um sistema de rotatividade desempregados-empregados).

Embora possa parecer difícil tornar muito menos raros os "sobredotados para alguma tarefa" devemos considerar que só agora estamos a entrar numa fase de grande expansão de modelos pedagógicos e cognitivos com óbvias repercussões na realização dos potenciais intelectuais humanos. Nomeadamente as ciências cognitivas podem não estar longe de cumprir os seus programas de compreensão dos mecanismos humanos de aprendizagem. O quadro da evolução dos conhecimentos médicos e genéticos também é obviamente de considerar.

Assim como será de considerar uma nova pedagogia, aplicada desde o momento da concepção do indivíduo (ou mesmo antes), e centrada no levantamento das características de casos concretos de sucesso individual.

Obviamente que, à partida, não se podem demonstrar teses sobre a futura "nivelação por cima" do nível intelectual dos indivíduos. Contudo o facto do sucesso de uma fluidificação do mercado depender deste fenómeno de nivelação talvez seja factor importante para que se venha a implementar uma nivelação intelectual que, embora longe de absoluta, possa ser suficiente para uma fluidificação do mercado.

Não se trata de transformar todos os indivíduos em génios, trata-se de conseguir que, em relação a cada tarefa intelectualmente exigente, existam mais indivíduos capazes de a exercer do que os realmente necessários para a exercer. Aliás estudos sobre os processos mentais envolvidos nas tarefas intelectualmente mais exigentes, com vista a torná-las mais simples, parecem ser viáveis no quadro do desenvolvimento das ciências cognitivas.

Retomemos agora a questão da distribuição de um grupo equilibrante pelas diversas profissões.

Como já vimos o grupo equilibrante terá de estar distribuído pelas diversas profissões sem o que serão difíceis de evitar conluios entre algumas profissões.

É pois necessário conseguir que indivíduos dotados exerçam profissões onde esses dotes não são relevantes. Ou então é necessário conseguir que esses "sobredotados" estejam desempregados sem que isso constitua um problema. Para discutir a viabilidade desta situação consideremos o seguinte:

a) Num cenário de fluidez de mercado as tarefas são remuneradas principalmente em função do esforço que envolvem e só secundariamente pela raridade dos dotes intelectuais que implicam (como se tenta mostrar em 5).
b) Num cenário de informatização e cibernitização das tarefas é admissível pensar que as

tarefas intelectualmente menos exigentes passem a ser automatizadas. Isto é, o sobredotado terá um leque suficiente de tarefas interessantes para escolher.

c) Num cenário de desenvolvimento tecnológico que possibilite part-time (ou rotatividade desemprego-emprego) suficientemente agradável o sobredotado poderá exercer, extra profissionalmente, criação artística e científica através da qual melhore a sua situação devido à melhoria social geral que poderá provocar pela sua criatividade.[13]

5. Fluidez, Critério de Esforço, Sistema de Oportunidades para Realização Material de cada Indivíduo, Incentivos Materiais

5.1. Critério de esforço

Sendo impossível o processo conflitual de negociação (e considerando também uma resposta muito rápida na formação de profissionais face às exigências do mercado) então o nível básico de remuneração de cada profissão está em proporção directa com os esforços inerentes ao seu exercício e os benefícios inerentes aos bens que cria.

Não fica pois a remuneração dependente das capacidades de negociação conflitual das classes envolvidas, nem das pressões relativamente conjunturais na sua procura.

O pagamento de um bem é em princípio apenas o suficiente para que o balanço entre esforço do seu criador e benefício desse pagamento seja suficiente para o levar a criar o bem. Aparentemente o benefício do comprador não entra em consideração a não ser no caso em que não cobre o mínimo que ele tem de pagar para o criador ser levado a criar o bem. Aliás mesmo neste caso aparentemente o benefício do comprador não influi no preço do bem, apenas determina se ele é comprado ou não.

A remuneração de cada profissão parece pois apenas em proporção directa com os esforços inerentes ao seu exercício. Contudo a rentabilidade relativa entre produtores é também determinante. De facto os compradores (ou empregadores) vão preferir os produtores mais eficazes e assim pelas leis da oferta e procura, vão pagar em função das rentabilidades relativas dos produtores.

Assim este sistema de distribuição de conhecimentos regula automaticamente os rendimentos das diversas classes, e determina o seu rendimento, primariamente em função do esforço desenvolvido em cada profissão, secundariamente em função das rentabilidades relativas dentro dessas profissões.[14]

5.2. Sistema de oportunidades para a realização das ambições materiais

Consideremos as seguintes asserções:

a) Se fôr suficientemente importante para ele então um indivíduo poderá desempenhar o esforço e o sacrifício exigido pela mais penosa das profissões existentes.

b) Se for suficientemente importante para ele então um indivíduo poderá desempenhar a intelectualmente mais difícil das profissões, ou aquela que exige maior capital, ou aquela que menos indivíduos podem, por outras razões, neste momento efectuar.

O mercado fluido possibilitará tanto mais a realização material pessoal de cada indivíduo quanto maior o número de casos em que se verifique a asserção a) e não se verifique a asserção b).

Note-se contudo que num mercado fluido o que determina "a penosidade" de cada profissão é o nível de remuneração que basta para aliciar para ela o número de indivíduos produtivamente necessários para a sua efectuação. A "penosidade" para um dado indivíduo poderá não coincidir com a referida determinação social da "penosidade". A

eficácia do mercado fluido na questão das possibilidades de realização material dos indivíduos dependerá pois da existência de suficientes semelhanças psicológicas entre os indivíduos. Mais exactamente bastará que pelo menos numa das profissões mais penosas (em termos da determinação social do mercado fluido) o indivíduo possa ser dos mais eficazes.[15]

5.3. Recompensas pecuniárias e não pecuniárias. Sociedade socialista

Recompensas não monetárias devem ser também consideradas como recompensas tão efectivas e eficazes como as pecuniárias. De facto qualquer crédito social, como a reputação ou posição de "chefia", é tão desejado e eficaz como créditos sociais fornecidos através de numerário.

Embora sem base num conhecimento concreto das sociedades socialistas parece razoável colocar as seguintes hipóteses.

Nas sociedades socialistas as necessidades de repressão sobre as correntes capitalistas levou à instauração de um sistema repressivo relativamente tão importante que os seus constituintes (militares, polícias, ideólogos com uma função de crítica às tendências capitalistas) usufruiram dos máximos créditos sociais. Como o valor de um crédito se determina pelas relações de status ou poder social então difícil se torna dar créditos sociais, aos agentes económicos, que expressem um status aliciadoramente alto sem que por isso o sistema repressivo perca poder e eficácia. É aliás também por isto que na liberalização económica se sente a necessidade de ser feita conjuntamente com uma liberalização política e com uma orientação pacifista que minimize a necessidade social do sistema repressivo.

6. Modo Impositivo de conseguir um Sistema sem Custos Conflituais

Em princípio parece possível que um governo implemente uma política proibindo suspensões conflituais, desbloqueando o capital, retirando ou limitando o controle dos preços aos empresários. Difícil será que assim, prescindindo das indicações fornecidas pelo sistema de mercado, este governo possa saber a que nível deve fixar as remunerações de cada classe a fim de aliciar para o seu desempenho um número suficiente de indivíduos. E será difícil pois difícil será evitar conluios entre os que se estão a formar como profissionais e os profissionais já existentes na mesma profissão, com o fim de dar a ideia de que os níveis de remuneração fixados para ela são insuficientes.[16]

Por outro lado um sistema repressivo poderá ocasionar os problemas de incentivo encontrados nas sociedades socialistas. Aqui não se discute mais pormenorizadamente as possibilidades de atingir os objectivos de uma política de fluidez de mercados por meios mais repressivos que os necessários à implementação da fluidez que tem vindo a ser referida. Contudo, em princípio, a via mais repressiva será uma via possível, que conviria estudar em pormenor, esperemos que apenas para melhor argumentar contra ela.

7. Nota Final

Em conclusão, a nível programático e de forma simplista, poder-se-á dizer que o tempo que a escola demora a fornecer a preparação técnica é factor decisivo no conflito social, podendo ser factor muito mais importante do que a preparação técnica, social e humana que possa dar. Mudanças decisivas dependem de mudanças não só na pedagogia mas também na divisão técnica do trabalho, o que depende sobremaneira das ciências

cognitivas e das possibilidades telemáticas.

O que se disse para o tempo da formação poder-se-á dizer em relação à tendência do sistema educativo em aceitar como um facto inultrapassável a raridade dos sobredotados.

Uma segunda ordem de conclusões diz respeito à dificuldade de superar o conflito social sem alguma alteração no conceito de propriedade, embora não pareçam necessárias as soluções socialistas extremas de apropriação social do capital. Essa restrita alteração no conceito de propriedade parece aliás consentânea com a teoria clássica do mercado livre.

A forma repressiva de superar os custos do conflito social parece ser tecnicamente difícil e prejudicar excessivamente o sistema de incentivos materiais, cuja eficácia deve ser preservada por quaisquer medidas de superação do conflito e transparência informativa.

Por último a situação de desemprego maciço além de não ter de constituir um problema psicológico e material para os desempregados (não ter de trabalhar pode ser agradável) não tem de ser fonte de conflitos sociais. Pelo contrário pode ser condição essencial para a superação dos conflitos sociais.

Tratou-se aqui de fazer o levantamento de algumas situações (pedagógicas, informáticas, financeiras) de que poderá depender a resolução de alguns dos principais problemas da relação social. Sobre como se pode atingir essas situações pouco se disse, estando aliás algumas delas algo distantes. Contudo o facto de se mostrar a importância social do desenvolvimento destas situações pode até certo ponto contribuir para uma mobilização e cooperação, social e científica, que além do mais seja desde já suficientemente entusiasmante para modificar a qualidade da relação social.

Contribuir para essa cooperação propondo ideais sociais que de alguma forma estabeleçam pontes entre os sistemas socialistas e capitalistas, com a mínima agressividade e limitado choque social, foi a intenção deste trabalho.

Notas

(1) De facto no quadro actual do conhecimento desenvolveram-se meios intelectuais passíveis de racionalizar e dialogar sobre o desejo humano. Encontramo-nos presentemente num quadro social em que a mutabilidade de certos níveis de desejo (vidé publicidade, vidé psicanálise) é patente. A produção de desejos orientados para uma adequação aos meios sociais disponíveis parece ser tarefa fundamental para a nova teoria social.

Assim a reflexão social, ou individual, sobre as acções humanas tem vantagem em assumir a forma explícita de uma reflexão meios-fins (isto é, de uma função relacionando input-output) em que a problemática dos fins é a problemática do desejo.

Entre os referidos input e output da filosofia das acções se inscreveriam multiplas funções locais, em que os outputs de umas funcionariam como inputs de outras e assim sucessivamente, até se percorrer o caminho desde o input básico até ao output terminal.

À partida a forma da filosofia global não consiste pois num conjunto de axiomas ou leis a partir dos quais o conhecimento se constituiria mediante a combinação, ou particularização, desses axiomas ou leis. Em aparência formalmente mais próximo estará de classificações englobando toda a realidade nas suas classes mais vastas, que se vão dividindo em subclasses cada vez mais finas. Contudo numa filosofia das acções a extensão de uma classe não é necessariamente constituída pela soma das extensões das suas subclasses. De facto numa filosofia das acções uma "classe" pode ser o efeito para o qual concorrem as suas "subclasses". Trata-se pois frequentemente de uma relação de causalidade e não de inclusão, assumindo uma filosofia das acções o aspecto de uma função constituída por vários ramos, eventualmente em parte sobrepostos, nos quais entre o output terminal e o input básico se podem diferenciar vários níveis de input e outputs intermédios.

(2) Abordar-se-á pois a questão das opções entre trajectos de "construção" social e trajectos de "redisposição" social, entendendo-se como trajecto construtivo aquele que é interpretado como melhoria pessoal por alguns indivíduos, sem que outros indivíduos o interpretem como tendo sido um trajecto que para eles próprios piorou algo. Caso estes últimos indivíduos interpretassem esse trajecto como sendo algo que para eles piorou alguma coisa então estariamos perante um trajecto de "redisposição".

(3) Note-se de passagem que é de colocar a hipótese (desenvolvida pelo autor noutros textos) de que o sucesso da ciência não se deve a especificidades do seu método investigacional, mas sim à especificidade comunicacional que possibilitaria uma comunicação eficaz entre os cientistas de todos os tempos, acarretando assim as vantagens da divisão intelectual do trabalho. Vantagens estas que podem bastar para explicar os sucessos da ciência. Nesta perspectiva a preocupação com a inter-subjectividade da comunicação feita no âmbito de uma filosofia das acções constituiria um meio de aproximação à ciência, e à sua eficácia, muito mais vasto do que poderia parecer à primeira vista.

(4) Isto é, pressupõe-se que a opção individual entre construção e redisposição depende de condições materiais exteriores ao indivíduo, histórica e socialmente mutáveis. Pressupõe-se pois que o cerne dos problemas da teoria social consiste na detecção de quais as condições externas que tornam a "construção" um investimento pessoalmente mais rentável do que a "redisposição".

(5) Como optimizadores dos trajectos redispositivos podem-se à partida considerar as seguintes condições:
a) Capacidade física de submeter outros indivíduos, ou de manipular armas que o façam.
b) Capacidade demagógica. Capacidade de iludir ou enganar outros indivíduos. Trata-se pois da existência de diferenças de informação e de formação, entre os indivíduos, que possibilitam trajectos demagógicos.
c) Capacidade apelativa. Capacidade de mobilizar indivíduos para que obriguem outros a fazer algo que não pretendem. Trata-se pois aqui de uma semelhança de informação e formação entre um número de indivíduos suficientemente maior, ou com melhores condições redispositivas e apelativas que os outros grupos. Trata-se em suma de uma situação em que existe uma grande diferença entre os grupos de semelhanças informativas (culturais).
d) Capacidade de mercenarização. Capacidade de um indivíduo levar ao conhecimento de outros indivíduos as suas capacidades comprovadas de "construção", ou de redisposição, de forma a que os outros obedeçam a ordens redispositivas a efectuar sobre terceiros. Subentende assimetrias sociais na capacidade de projectar capacidades pessoais (reputação, dinheiro) e um grande número de indivíduos potencialmente mercenários. Note-se que estes últimos factores são afinal argumentos para as conveniências da igualdade social, assim como são uma caracterização do que terá de ser essa igualdade.
e) Alguns factores citados para a construção (factores a,b,c,d da construção referidos no corpo do presente trabalho) são factores de redisposição quando são aplicáveis à produção de armas, ou de bens, que facilitem os factores redispositivos considerados.

(6) Por luta de competição hierárquica entenda-se o processo em que cada indivíduo tenta obter o máximo crédito social (dinheiro e reputação) de forma a superiorizar-se aos outros com vista a poder ter primazia no acesso a bens e serviços escassos.

(7) Dentro de um quadro de inputs que têm como output a opção por trajectos constitutivos ou redispositivos é possível pensar os efeitos sociais de diversas estruturas materiais pensando-as enquanto factores que se repercutem nestes inputs.
Por exemplo a tradicional importância social da religião pode ser compreendida nomeadamente através dos seus efeitos pró-construtivos a nível dos factores c,d,f de "construção" (referidos

no corpo deste texto):
c) pressão social e psicológica para trajectos construtivos, d) igreja como mecanismo social de arbítrio da repartição social dos bens, f) valorização do valor pessoal de bens facilmente produzidos através da valorização da pobreza ou da sobriedade, ou através da valorização de "bens" espirituais facilmente produzíveis, como os relativos ao seguimento dos ritos, ou prossecução das virtudes comportamentais.
Dentro do quadro da filosofia das acções parece assim ser possível pensar os efeitos das diversas estruturas e instituições sociais.

(8) Esta suspensão conflitual é nomeadamente um meio de demonstrarem a real importância que para eles têm as diferenças que são cerne do desacordo.
Assumindo actualmente o aspecto tradicional de "luta de classes" as suspensões conflituais poderão vir a assumir o aspecto de conflitos entre associações de consumidores e produtores.

(9) É também condição de fluidez que indivíduos dessas outras classes possam reverter para as suas anteriores actividades, assim minimizando os custos sociais de um excesso de oferta. A possibilidade de os profissionais, que por exigirem remunerações altas se viram assim desalojados, poderem reverter para outras profissões é também elemento de fluidez minimizador dos custos sociais destes movimentos.

(10) "Como no mercado fluido se pode responder à determinação do nível de remuneração de uma classe quando ela controla os preços dos produtos".
No vertente caso as respostas habituais são as greves ou o aumento dos preços por outros produtores. Contudo em princípio a resposta poderia surgir logo no acto da compra dos produtos, nomeadamente através das associações de consumidores. Estas não precisariam de efectuar boicote aos produtos que considerem de preço excessivo, bastaria que dessem o seu parecer sobre os preços. Caso o seu parecer não fosse passado à prática então as associações teriam de motivar o grupo equilibrante para assumir esse projecto de produção.

(11) Além disso poderá ter de ser numericamente maior do que a classe profissional que pretende substituir. Isto evitará que alguns conluios entre classes, ou entre classe e indivíduos pertencentes ao seu grupo equilibrante, restrinjam a um número excessivamente pequeno o conjunto equilibrante disposto a efectuar a tarefa de que são grupo equilibrante.
De qualquer forma não é condição para a fluidez que "todos saibam tudo e sejam igualmente inteligentes". O que se discute aqui são as condições e os níveis de aproximação a esta situação que poderão ser suficientes para minimizar o conflito social, mantendo simultaneamente as condições gerais de "construção".

(12) Não se trata só do custo directo da educação mas também dos custos que advêm do facto dos indivíduos não estarem a produzir enquanto estão "a estudar". No caso de uma situação de desemprego ou subemprego generalizado estes segundos custos serão obviamente muito menores.

(13) Para se compreender melhor até que ponto poderá bastar ao indivíduo (sobredotado ou não) que a recompensa de um seu esforço criativo seja apenas aquela que resulta automaticamente do avanço social geral que essa criatividade provocou, atente-se nas seguintes perspectivas: Será possível considerar que os dois tipos principais de problemas da humanidade são a sobrevivência (em situação ecológica normal) e as anormalidades eventuais (catástrofes naturais, doença, escassez devido a explosão demográfica excessiva, eventualidade de dor na morte, etc.).
Outras questões, como as de carácter sexual e afectivo, podem ser vistas como consequências destes problemas. Por exemplo a problemática afectiva, ligada à existência de hierarquias selectivas de afeições preferenciais (padrões de beleza, etc.), está intimamente relacionada com processos sociais não conscientes de eugenia. Processos de eugenia que, em última

análise, são modos de luta sobrevivencial e anti–a.e. (anti–anormalidades eventuais).
Com o avanço tecnológico a problemática relativa às a.e. passa claramente a ser a problemática dominante.
Os processos gerais de luta anti a.e. são, em geral, ou tentativas de evitar as a.e., ou tentativas de adaptação às a.e. A adaptação pode ser psicológica (nomeadamente através da religião) ou material (conquista de uma posição na hierarquia social que possibilite sofrer menos os efeitos de uma a.e.).
Apesar das suas limitações a análise marxista evidenciou as consequências da competição hierárquica (vide nota 6) a nível da insatisfação social e do imobilismo face ao progresso social (o velho poder reagindo ao novo poder).
O desenvolvimento das sociedades está criando um quadro em que pelo menos parcialmente, a competição hierárquica poderá ser gradualmente substituída por outras vias anti a.e. com menos inconvenientes que a via hierárquica tradicional.
O desenvolvimento tecnológico fornece novos meios de controle das a.e. Controle à explosão demográfica, medicina preventiva, previsão meteorológica, protecção civil, previsão sísmica, etc. Origina também tempos livres e meios de informação–comunicação.
Assim possibilita novos processos, não "hierarquistas", anti a.e. De facto a partir do momento em que os indivíduos compreendem o sentido da investigação tecnológica e científica (e têm algum controle sobre o seu rumo) o investimento pessoal nesta via passa a constituir uma sua forma possível de luta contra as a.e. As perspectivas de investimento nas vias preventivas podem assim crescer, em detrimento do investimento em vias como as hierárquicas.
Por outro lado, a potencialidade que a arte, articulada com a análise psicológica, tem em alterar a forma individual de reacção emocional às situações difíceis (vide processo de compactação de escala desenvolvido no trabalho do autor "Potencialidades heurísticas e fundamentadoras de modelos cognitivos neurais") pode levar a que se torne interessante o investimento no desenvolvimento da capacidade artística de cada um, enquanto processo de luta adaptativa anti–a.e.
Assim, o investimento nas vias adaptativas psicológicas (arte, análise psicológica) e nas vias preventivas (controle sobre o sentido da investigação científica) poderá ir substituindo com vantagem a competição hierárquica. Uma sociedade em que todos sejam "investigadores" e "artistas" parece assim constituir uma utopia que embora longínqua se afigura coerente e plausível.
Note–se contudo que esta perspectiva dual (sobrevivência/anormalidades eventuais) pressupõe que não existem, nem surgirão, objectos de consumo de difícil acesso económico e que preencham necessidades humanas que não estejam directamente relacionadas com a sobrevivência nem, directa ou indirectamente, com a necessidade humana de segurança face às anormalidades eventuais. Poder–se–á chamar a este tipo de objectos "objectos de prazer em si". Uma cara "máquina de prazer" será um exemplo de um destes objectos.

(14) Mais exactamente as tarefas serão pagas tendo como limiar inferior o necessário para compensar o esforço nelas desenvolvido pelo subconjunto dos menos eficazes do conjunto de indivíduos que desempenha efectivamente essa tarefa. O que será o mesmo que dizer que têm como limiar inferior o quantitativo de recompensa tão baixo quanto o suficiente para aliciar para essa tarefa o número de indivíduos socialmente necessários para o seu desempenho (não estamos pois longe da realização integral do programa da economia clássica). O limiar superior será o resultado da multiplicação do limiar inferior vezes a razão entre rentabilidade máxima nessa profissão/ren. mínima dos indivíduos que realmente a desempenham.

(15) Apesar da grande importância psicológica da posição hierárquica, chegar ao topo da posição social poderá não bastar se esse topo fôr mais baixo do que em outros sistemas sociais possíveis. Mais baixo nomeadamente devido à eventual existência de menores diferenças

socio-económicas.
Se considerarmos as perspectivas enunciadas na precedente nota 13, comportando nomeadamente a dominância do problema das anormalidades eventuais, e um quadro de desenvolvimento vasto das necessidades básicas de consumo e de comodidade então as diferenças no nível do topo social só serão importantes na medida que forem diferenças em relação às tecnologias preventivas. Tecnologias que um sistema de mercado fluido parece tão capaz de produzir como outros.

(16) Para impedir esse conluio poder-se-ia obrigar a que um estudante não pudesse mudar a sua escolha profissional nem suspender os seus estudos. Mas isto equivale a estar a obrigar os estudantes a optar definitivamente por uma profissão sem ter o mínimo de certeza de quais os seus níveis de remuneração futuros (pois estes estarão a ser testados pelo governo).

Bibliografia

ADAMS, J. (editor) - *Institutional Economics: Contributions to the development of holistic economics*, Adams J., Boston, 1980

ALTHUSSER, L. - *Ideologia e aparelhos ideológicos do Estado*, Presença, Lisboa, 1974.

AXELROD, R. - *La evolución de la cooperación*, Alianza Editorial, Madrid, 1986

BATLLE, A. - *Les travailleurs du futur*, Seghers, Paris,, 1986.

BELL, D. - *The coming post-industrial society*, Peregrin Books, Londres, 1974

BETELHEIM, C. - *Planificação socialista da economia*, Edições 70, Lisboa,
Cálculo económico e formas de propriedade, D. Quixote, Lisboa, 1975

BOWLES, S./GINTIS, H. - *The problem with human capital theory - a marxian critique*, The American Economic Review, p. pers and proceedings may 1975, vol. 65, nº 2, pp. 74-82.

BRESSAND, ALBERT - *O mundo de amanhã*, D. Quixote, Lisboa, 1988

CANTO, D. / JOINES, D. / LAFFER, A. - *Foundations of Supply-Side Economics*, New York, 1983

CAZES, B. - *Histoire des futurs*, Seghers, Paris, 1986

CAZES, B. (e outros) - *Vers l'an 2000... et aprés?* La documentation française, Paris, 1987

CETRON, M. - *Jobs of the future*, Mac-graw Hill, London, 1984

CHAMBERLAIN, E.H. - *The theory of monopolistic competition*, Harvard University Press, Cambridge, 1933

DUCROQ, A. - *1985-2000: O futuro é já hoje*, D. Quixote, Lisboa, 1986

DWORKIN, R. - *Taking rights seriously*, Harvard University Press, 1977

FRIEDMAN, M. - *Liberdade para escolher*, Europa-América, Lisboa, 1982

GALBRAITH, J. K. - *A era da incerteza*, Moraes, Lisboa, 1988
- *A sociedade da abundância*, Europa-América, Lisboa, 1984
- *Anatomia do poder*, Europa-América, Lisboa, 1987
- *O novo Estado industrial*, D. Quixote, Lisboa, 1973
- *Economia e bem público*, Europa-América, Lisboa, 1978

GILDER, G. - *Wealth and poverty*, New York, 1981

GINTIS, H. - *The reemergence of marxian economics in America*. - Left Academy. Marxist scholarship on America campuses, Ollman, B., e Vernoff, E., New York, 1982

GOBET, M. - *Demain les crises: de la résignation à l'antifatalité*, Hachette, Paris, 1988

GREENWALD, B. - *Adverse selection in the labour market*, Review of economic studies vol LIII 1986, pp. 325-347

GRUCHY, A., J. - *The current state of institutional economics*, The American Journal of Economics and Sociology, vol. 41 nº 3 1982

HAEFELE, EDWIN T. – *A utility theory of representative government*, The American Economic Review, vol. 61, nº 3, Pt. 1, p. 350, Junho 1971

HEILBRONER, R. – *Between capitalism and socialism*, New York, 1970

ILICH, IVAN – *A convivencialidade*, Europa–América, Lisboa, 1987

JACQUEMIN, A. – *Selection et pouvoir dans la nouvelle économie industrielle*, Economica, Paris, 1985

JENSEN, P. / WESTERGAARD–NIELSEN – *A search model applied to the transition from education to work*, Review of economics studies LIV 1987, pp. 461–472

KUMAR, K. – *The rise of the modern society*, Basil Blackwell, Oxford, 1988

LESOURNE, J. (e outros) – *L'école de l'an 2000*, Le monde de l'éducation, Paris, 1988

LINDER, STAFFEN B. – *The Harried leisure class*, Columbia University Press, New York, 1970

MACHLUP, F. – *The economics of information and human capital*, Princeton, 1984

MACLUHAN, MARSHAL – *Os meios de comunicação como extensão do homem*, Cultrix, São Paulo

MARCUSE, H. – *Liberdade e ordem social*, Lisboa, 1975

NOZICK – *Anarchy State and Utopy*, New York, 1974

POQUET, G. – *Perspectives 2005*, Economica, Paris, 1987

RAWLS, J. – *Theorie de la justice*, Seuil, Paris, 1982

REICH, CHARLES – *The greening of America*, Random House, New York, 1970

REICH, R. – *Human capital and economy policy*, Dissent nº 3 1983

SAMUELSON, P.A. – *Economia*, Gulbenkian, Lisboa

SCHERER, F. – *Industriel market structure and economic performance*, Houghton Mifflin, London, 1987

SCHUMACHER, E.F. – *Small is beautifull*, D. Quixote, Lisboa, 1980

– *Um guia para os perplexos*, D. Quixote, Lisboa, 1985

SMITH, D.R. – *Mobility in professional occupational–internal labor markets: stratification, segmentation and vacancy chains*, American Sociological Review vol. 48 nº 3, Junho 1983, pp. 289–305

VEBLEN, THORSTEIN – *The theory of the leisure class*, Houghton Mifflin, Boston, 1973

WELCH, F. – *Human capital theory: education, discrimination and life–cycles*, The American Economic Review, papers and proceedings may 1975, vol. 65, nº 2, pp. 63–73

2

EDUCAÇÃO

EDUCATION

Duarte Costa Pereira

Fac. de Ciências, Univ. do Porto,
Porto (Portugal)

Modelos como representações privilegiadas nas estratégias de confrontação para aprendizagem da ciência

Resumo

As posições construtivistas estão a colher forte aceitação na Psicologia Cognitiva e na Ciência da Educação, enquanto na Filosofia das Ciências o primado da questão epistemológica, introduzido por Descartes, foi afastado e substituído por uma idéia de verdade científica convencionada. Também a mudança de paradigma na ciência tornou o modelo perceptual inadequado como suporte de modelos científicos conceptuais, verificando-se que o princípio de modelação de anschaulichkeit *foi substituído pelo princípio da incerteza tal como o determinismo e o realismo foram substituídos pelo estocastismo e pelo instrumentalismo.*

Sustenta-se neste estudo que a importância histórica dos modelos no desenvolvimento da ciência não se desvaneceu, principalmente agora que a intuição deixou de ser a fonte única da explanação e a computação está a tomar o lugar daquela. De facto, entre todas as representações possíveis do conhecimento, as analógicas, baseadas na idéia de Modelos Mentais (Johnson Laird, 1983), parecem ser os melhores suportes para as teorias da ciência que são proposicionais e mudam abruptamente em contraste com os respectivos modelos de sustentação. As pesquisas empíricas confirmam que a alimentação de modelos científicos é muito complicada e distante da estrutura e mecanismos perceptuais. Analisam-se também argumentos importantes no sentido de que a melhor maneira de auxiliar efectivamente a genese do pensamento teorético consiste no recurso a uma estratégia de confrontação, usando-se modelos de configuração intermédia entre a complexidade perceptual e a abstracção conceptual que possam ser actuantes. Apresentam-se argumentos acerca da excelência destes modelos pedagógicos e da eficácia do computador para a sua implantação. Por último, apresenta-se uma classificação de modelos pedagógicos computacionais em ciência, com exemplificação de alguns.

Abstract

In both Cognitive Psychology and Science Education constructive views are being strongly held and in Philosophy of Sciences the primacy of the epistemologic question, introduced by Descartes, has been discarded and substituted by a conventional view of scientific truth. Also the paradigmatic shift in science made the perceptual modelling inadequate to support Science conceptual models, the anschaulichkeit *principle of modelling being substituted by the uncertainty principle and determinism and realism being replaced respectively by stochastism and instrumentalism.*

This paper will argue that the historical importance models have had in the development of Science hasn't vanished, particularly when intuition is abandoned as the only source of explanation

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 411/430
1992, APDC & SPF, Lisboa

and computability is taken instead. In fact, from all possible representations of knowledge the analogic ones, based on the idea of Mental Models (Johnson Laird, 1983) seem to be the best supports for the Theories of Science which are propositional and change abruptly in contrast with their supporting models. Empirical research supports that the nurture of Scientific models is very complicated and removed from perceptual structure and mechanisms. Also important arguments are analysed suggesting that the best way to effectively help the genesis of theoretical thought is by means of a strategy of confrontation using models with an intermediate configuration between the perceptual complexity and the conceptual abstraction, that may be acted upon. Arguments are presented about the excellency of these pedagogical models and the suitability of computers for their implementation.

Finally a classification is presented for the pedagogical computer models in Science and a few are exemplified.

1. Introdução

Neste artigo, pretende fazer-se sobretudo uma apreciação crítica do papel dos modelos na Educação Científica actual.

Antes de mais, já que o vocábulo modelo (Rosado, 1986) é muitíssimo vago, convém precisar o seu significado, no âmbito deste artigo:

representação operacional (sobre a qual se pode actuar) e analógica (preservando a estrutura) de um fenómeno e que não precisa de ser exactamente uma réplica dele.

Assim, por exemplo, um relógio é um modelo do movimento de rotação da terra *(modelo externo)* e também se pode dizer que todo o professor possui um modelo *(modelo mental)* de uma sala de aula.

A simples imitação sem preservação das características estruturais nem manutenção da possibilidade de transformação, designa-se por *simulação* e distingue-se nitidamente da *modelagem.*

A tarefa que nos propomos é muito difícil pois começamos por enfrentar um aparente paradoxo:

os modelos tais como eram entendidos pelo paradigma científico (Kuhn, 1970) anterior às revoluções relativista e quântica não são mais admitidos como válidos; no entanto, nunca se falou tanto como nos tempos que correm em modelos, tanto como instrumentos de progresso científico como no seu uso como auxiliares pedagógicos.

Tratar-se-á de outro tipo de modelos?

Sem dúvida. De facto, o princípio de *anschaulichkeit* (Calado, 1985), que preconizava a possibilidade da representação da realidade ser sensível aos mais ínfimos pormenores, foi completamente abandonado e substituído pelo princípio de *incerteza*, como abandonado foi também o princípio da *causalidade*, base do *determinismo* científico, que foi substituído por um *estocastismo*. Também além disso, como veremos, questões filosóficas levaram ao abandono do *realismo* e sua substituição por um *instrumentalismo*. Isto, no entanto, não fez abandonar os modelos, só fez com que o objectivo principal da *modelagem científica* deixasse de ser o *ajustamento ao senso comum* que continua a ser o objectivo da *modelagem perceptual* e passasse a ser a *coerência com as teorias* muitas vezes contra-intuitivas. No caso dos *modelos educativos*, como defenderei adiante, verificar-se-á um compromisso entre as duas tendências, incorporando-se alguns aspectos *intuitivos*, para faciliar a compreensão, entre os quais a própria causalidade.

Como se não bastassem as consequências da mudança do paradigma científico, o estatuto epistemológico dos modelos tem vindo também a ser afectado pelas recentes correntes de pensamento em *Filosofia, Psicologia e Educação.*

Com efeito, motivado em grande parte pela evolução científica, e após o falhanço rotundo do positivismo lógico (Wittgenstein, 1969), assiste-se em *Filosofia* a uma desvalorização da questão epistemológica, que tinha sido introduzida por Descartes como primordial no domínio da Filosofia das Ciências, evoluindo-se não só como foi dito para o abandono do *realismo* (Gilbert & Watts, 1983) em favor de um *instrumentalismo*, mas, muito principalmente, para uma ideia de *verdade científica construída e convencionada* (Habermas, 1987). Isto como que, ao negar a atingibilidade da verdade pela Ciência, implica que os modelos científicos sejam objecto de um contrato social: *veridicção*.

Por outro lado a *Psicologia Cognitiva* ao permitir a abertura da caixa preta do Behaviorismo e admitir a complexidade perceptual e conceptual, refuta o empirismo radical Baconiano e, sem caír num racionalismo estéril e solipsista, advoga para a génese do conhecimento uma posição construtivista (Carey, 1986). Esta posição coincide também com o que por vezes se chama paradigma das concepções alternativas da *Educação* (Driver, 1983). Com efeito fala-se muito hoje em dia, na importância das "concepções alternativas" no processo ensino/aprendizagem das ciências. Segundo este modelo de aprendizagem os alunos constroem versões privadas dos conceitos sobre que posteriormente elaboram às suas interpretações. Nesta perspectiva construtivista do conhecimento é de extrema importância que o professor conheça o ponto de onde parte o aluno, para que possa escolher estratégias adequadas facilitando-lhe a construção de novos modelos conceptuais mais próximos dos modelos científicos. Digamos que o conhecimento das "concepções alternativas" relacionadas com os vários domínios da ciência será condição necessária, ainda que não suficiente, para o sucesso no processo ensino/aprendizagem, e que a Educação Científica será essencialmente um processo de convergência dos modelos privados dos aprendizes para os modelos públicos verediccionados.

Antes de tratarmos mais detalhamente a caracterização dos modelos adaptados às estratégias de confrontação (5) vamos ver brevissimamente alguns dos aspectos mais importantes da questão da representação do conhecimento (2), e algumas questões relacionadas com o uso de modelos em Psicologia (3) e em Ciência (4).

2. A questão da representação do conhecimento

O problema da representação do conhecimento é um problema central, quer em Inteligência Artificial, quer em Psicologia Cognitiva e embora se possam encontrar muitas distinções na sua abordagem há, de facto, muitos aspectos comuns que culminam na refutação do dualismo Cartesiano através da admissão do conceito de computação (Johnson Laird, 1988) que tem por base a teoria matemática da computabilidade, anterior ao aparecimento dos computadores digitais e que é sustentada, como posição chave, pela moderna Ciência Cognitiva. A principal distinção que existe entre os tratamentos nas duas disciplinas, é que a Psicologia se baseia numa teoria de correspondência da verdade, enquanto que a Inteligência Artificial opta por uma teoria de coerência. As principais controvérsias podem resumir-se (Anderson, 1983) num campo e noutro a:

O que é uma representação?

Deve ser analógica ou proposicional?

Deve ser procedimental (de processo) ou declarativa (factual)?

Será que coexistem vários tipos de representações na mente humana?

Representação é uma entidade que está em vez de outra (Brachman & Levesque, 1985). A *Representação do Conhecimento* ocupa-se dos constructos mentais que estão em vez do mundo exterior ou que o *representam* e das regras da sua manipulação.

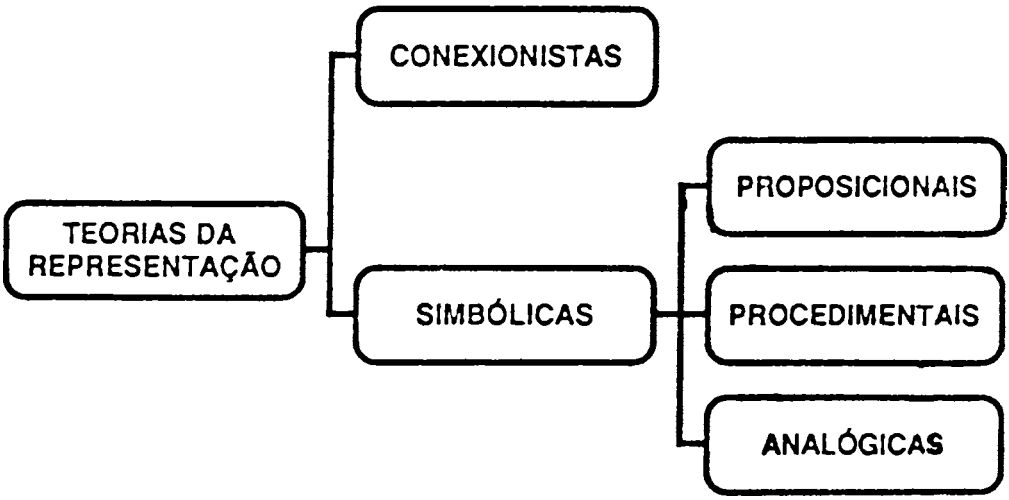


Fig. 1 - Classificação das Teorias da Representação

Estes constructos podem ser símbolos ou ligações conforme as Teorias de Representação sejam *Simbólicas* ou *Conexionistas*, devendo quer num caso quer noutro ser computáveis.

As teorias *Conexionistas* (Hinton & Anderson, 1981) têm vindo a aparecer recentemente, particularmente no campo da Psicologia, procurando emular as verdadeiras ligações neurológicas, responsáveis pela aprendizagem, para tal baseando-se na ideia de processamento paralelo distribuído. Enquanto não se generalizar o uso de computadores com processamento paralelo e o custo das memórias não descer ainda algo mais, tais teorias estão limitadas, por causa da explosão combinatória a que dão origem, a fenómenos bastante restritos. Mesmo assim há a assinalar progressos notáveis como a modelagem eficiente das formas irregulares do passado dos verbos em inglês (Mc Clelland & al, 1986).

Para já a maior parte das teorias de representação são *Simbólicas* i.e. o conhecimento é representado por símbolos que podem posteriormente ser manipulados.

A natureza de tais símbolos é que varia de autor para autor, já que não existe nenhum paradigma unificado nem em Psicologia Cognitiva nem em Inteligência Artificial.

Ao procurar sistematizar a diversidade de teorias Simbólicas existentes, uma primeira possibilidade consiste em classificá-las em *holísticas ou globais e elementaristas ou atomísticas*. Outra também interessante é a da consideração de teorias centradas no *sujeito da representação* ou no seu *objecto*. Na literatura (Johnson Laird, 1983), a sistematização mais encontrada e que seguiremos, é, no entanto, a que considera a trilogia de representações: *proposicionais, procedimentais e analógicas*.

As *representações proposicionais* assumem que o conhecimento é representado por um conjunto de símbolos discreto ou proposições, correspondendo os conceitos a enunciados formais, envolvendo esses símbolos (conhecimento declarativo).

Nas *representações procedimentais* o conhecimento é representado em termos de processo activo ou procedimento.

Nas *representações analógicas* a correspondência entre os mundos (representado e representador) é tão directa quanto possível, usando-se tradicionalmente variáveis contínuas para representar conceitos contínuos no mundo real.

Note-se que muitos dos sistemas de representação são híbridos. Assim é que, por

exemplo, em rigor, os sistemas de produção (Newell, 1973), tão importantes na construção de módulos de Inteligência Artificial, são, de facto, mistos entre proposições e procedimentos (as produções por que são constituídos não passam de pares ordenados de: condições, (que são proposições) e acções (que são procedimentos)); também os enquadramentos a que adiante nos referiremos como proposicionais são sistemas híbridos entre o proposicional e o analógico.

Por outro lado certos autores admitem constructos, como as séries temporais (Aderson, 1983) que não se enquadram em nenhum dos tipos precedentes.

Além disso existem teorias que envolvem todos os tipos de sistemas. Por exemplo, a ACT* (Aderson, 1983), para além das tais séries temporais, envolve proposições, imagens e regras de produção. Estas teorias podem designar-se por híbridas sendo aqui as diferentes formas de representação *cooperantes* (desempenhando um papel bem definido na manipulação umas das outras). Há também teorias (Johnson Laird, 1983) em que se admite um hibridismo *competitivo* (concorrência entre representações analógicas e proposicionais).

Apresentam-se no Quadro seguinte algumas das principais teorias de representação com relevo para as que privilegiam a correspondência (Psicologia) que são mais relevantes nas aplicações educativas.

PROPOSICIONAIS:	
	Lógica
	Atributos Semânticos (Smith, Soben e Rips, 1974)
	Redes Semânticas (Quillian, 1967)
	Dependência Conceptual (Schank, 1974)
	Esquemas e Enquadramentos (Minsky, 1975)
PROCEDIMENTAIS:	
	Sistemas de Produção
	Sistemas Periciais
	Sistemas Tutores Inteligentes
ANALÓGICOS:	
	Sistemas baseados em imagens
	(Shepard, 1982)
	(Kosslyn, 1980)
	Sistemas baseados em modelos
	Modelos de situação (Kintsch, 1977)
	Modelos internos (Collins & Brown, 1980)
	Cenários (Sanford & Garrod, 1982)
	Modelos mentais (Johnson Laird, 1983)

Fig. 2 - Algumas Teorias da Representação

É o último autor referido aquele cuja conceptualização seguiremos mais nos próximos capítulos por se revelar a mais adequada e fértil tanto na Ciência como na Educação Científica.

3. Modelos mentais

Embora, como se viu, não haja unanimidade sobre a questão da representação, em que a Educação se possa inspirar para modelar a aprendizagem, a posição mais fértil e exequível, deste ponto de vista é aquela que (refutando as representações conexionistas [Hinton & Anderson, 1981] por inexequíveis e as procedimentais pela evidência que existe [Johnson Laird, 1983], no campo da Psicologia, sobre o facto de os seres humanos não utilizarem regras como base das suas inferências) admite uma aprendizagem simultânea duma forma proposicional e analógica. Os sujeitos (Johnson Laird, 1983) evocariam modelos mentais pré-existentes que iriam modificando em conformidade com as proposições processadas (nesta perspectiva, processar uma proposição é conhecer como seria o mundo se ela fosse verdadeira) e com as imagens também processadas (que são interpretadas como vistas, a partir de um ponto do modelo). Atingindo-se uma certa discrepância crítica entre a base proposicional e o modelo (que é aceite sempre provisoriamente como hipótese) este é abandonado e ou se recupera um modelo alternativo, ou se reverte para uma aprendizagem exclusivamente proposicional. Como as características mnésicas dos dois formatos de representação, são substancialmente diferentes (as representações proposicionais são rapidamente processadas e lentamente recuperadas exactamente ao contrário dos modelos difíceis de processar mas rapidamente recuperados) esta teoria é cientificamente boa, por ser refutável no sentido Popperiano (Popper, 1972).

Saliente-se o carácter de hipótese que os modelos mentais assumem a ser ou não confirmada por uma estratégia que depende do *estilo cognitivo* (Pask, 1972) dos sujeitos e que pode envolver a activação em paralelo de vários modelos, sucessivamente desconfirmados ou de um único modelo contrastante com a base proposicional. Isto está nitidamente de acordo com as posições construtivistas já referidas da Psicologia e da Educação.

De acordo com a teoria dos Modelos Mentais a representação do meio ambiente faz-se por modelos mentais do dia a dia que organizam os dados sensoriais e que estão em permanente actualização: são os chamados *modelos perceptuais*.

Por outro lado estes modelos servem de base para, através do processamento de proposições, de imagens ou por mecanismos de inferência, serem modificados em modelos, já muitas vezes longe da realidade perceptual, que se designam por *modelos conceptuais*.

Há semelhanças (Gregory, 1981) entre as percepções e as concepções, nomeadamente o recurso à interpolação para suprir lacunas, à extrapolação para representar estados não sentidos e à criação de objectos nos espaços perceptual e conceptual, respectivamente. As principais diferenças (Gregory, 1981) entre as concepções e as percepções são:

1) As percepções são sempre feitas de um *ponto de vista*, enquanto que os conceitos têm um carácter *global*.
2) Os objectos da percepção são *instâncias* enquanto que os conceitos são *generalizações*.
3) As percepções limitam-se a objectos *concretos* enquanto que os conceitos são *abstractos*.
4) As percepções não são *explanativas* enquanto que a conceptualização o deve ser.

4. Modelos em ciência

Em Ciência costuma distinguir-se entre as *teorias* que são proposicionais e os *modelos* que são analógicos e que procuram tornar operacionais as teorias, quer operando a partir

dos primeiros princípios (modelos ab initio) quer ainda a partir de simplificações justificadas sempre só pela complexidade dos problemas (modelos simplificados).

Esta distinção (Johnson Laird, 1983) entre *teoria* (PROPOSICIONAL E "IMUTÁVEL") e *modelo* (ANALÓGICO E AJUSTÁVEL) é muito importante já que a evolução de modelos caracterizaria os períodos de Ciência normal (Kuhn, 1970), enquanto que a mudança de teorias caracterizaria os períodos de Ciência revolucionária.

O carácter hipotético e a possibilidade de serem manipulados dos modelos mentais a que nos referimos, e o carácter global, generalizante, abstracto e explanatório (particularmente este último) dos modelos conceptuais, torna-os nas representações privilegiadas pela Ciência.

O poder explanatório de tais modelos não deve ser confundido com uma procura de intuição (a ideia de Ciência como senso comum organizado está ultrapassada), para o que geralmente a modelagem mental espontânea não é adequada e há que recorrer a *regras de modelagem* (Richmond, Peterson & Vescuso, 1987) que se desviam da modelagem espontânea (particularmente quando há a intervenção da variável tempo) e que têm que ser objecto de aprendizagem (veremos algo sobre isto mais adiante).

Tais modelos podem ser testados pelos resultados experimentais obtidos sendo as teorias em que se suportam refutadas ou aceites conforme estes testes. O grau de sofisticação destes modelos varia substancialmente, podendo quando estão implementados em computador a ponto de gerarem resultados fiáveis (não necessariamente válidos) dar origem ao que se designa por *experimentação numérica* e que é talvez a maior novidade metodológica na Ciência deste século (Madureira, 1984). Várias são as técnicas usadas para a implementação de modelos, entre elas:

- ajustamento a curvas teóricas de dados experimentais
- regressão linear múltipla
- utilização de métodos iterativos de resolução de equações diferenciais
- utilização de métodos estocásticos.

Note-se que estes modelos teóricos verediccionados pela sociedade, por estarem de acordo com o paradigma vigente (Kuhn, 1970) (geralmente estabelecido duma forma proposicional), só têm valor se forem passíveis de refutação (Popper, 1972) e estão geralmente muito distanciados dos modelos perceptuais, sendo para a sua psicogénese preciso admitir uma actividade notável (objecto da educação científica). É esta última questão que procuraremos caracterizar agora.

5. Os modelos na educação científica. Confrontação como compromisso entre os modelos conceptuais e perceptuais

Todo o professor de Ciências sentiu certamente a importância das concepções alternativas e a resistência que elas oferecem à adopção do modelos formais, aceites cientificamente. O grande desafio que se põe é: como provocar a mudança?

Falou-se no afastamento dos modelos conceptuais dos perceptuais e da grande tarefa que representa a psicogénese dos modelos conceptuais para a Educação Científica. Por outro lado muitos autores e em especial R. Driver (1983) falam do papel que o confronto entre os modelos dos alunos e a "realidade" pode desempenhar na construção do conhecimento, levando-os a fazerem previsões, com base nos seus modelos conceptuais, e a confrontá-las com situações (bem escolhidas) que provoquem o conflito. A literatura tem dado muita ênfase à definição de estratégias que ajudem a conhecer os "modelos alternativos". Propostas de confronto não são, todavia, muito frequentes, nomeadamente

no que diz respeito ao ensino das Ciências. Este tem vários outros objectivos (Ogborn, 1988) de que se desempenha razoavelmente bem, como sejam a condução da actividade experimental e a educação de atitudes. É estéril o debate sobre a prioridade dos objectivos em Educação Científica, mas o que é certo é que o mais difícil, geralmente reservado para o professor (quando este não tem também os seus deslizes (Costa Pereira & Ribeiro, 1989), é o conseguir-se atingir o pensamento teórico a que de qualquer forma está (Einstein, 1920) subordinada a observação (nem racionalismo nem empirismo exagerado).

Se há áreas das Ciências em que o confronto é clarificador e é suficiente para provocar a mudança (é o caso de algumas situações no estudo da corrente eléctrica), outras há em que o confronto é problemático, é este o caso da mecânica por exemplo. Esta situação advém do facto de o "mundo" ideal, sem atrito, que queremos conceptualizar ser um "mundo" que nada tem a ver com a experiência do aluno. Nós propomos uma fase de confronto/configuração[1] e consideramos que há áreas da Ciência em que a configuração terá necessariamente um papel muito mais importante que o confronto, por exemplo em áreas em que os conceitos sejam extremamente abstractos como é o caso da energia.

Configurar é dar forma, é ajudar o aluno na construção de um "todo" coerente e em interacção. Não há vários conhecimentos, isto é, não há um conhecimento válido para a sala de aula, na resolução de problemas abstractos e outro válido para as situações quotidianas. O professor na sala de aula não deve separar mas sim ligar. Uma prática pedagógica muito arreigada que se pode considerar contraproducente por anticonfiguracionista é a de chamar a atenção do aluno para as diferenças entre a linguagem corrente e a linguagem científica sem ter a preocupação de estabelecer ligações. Com efeito a linguagem científica interactua necessariamente com a linguagem corrente através da vulgarização científica, e é frequente, hoje em dia, encontrarmos na literatura conceitos científicos trabalhados pela arte da escrita. A tendência será no sentido do conhecimento científico agir sobre o senso comum alterando-o. Não são mundos completamente separados. Como diz B. Sousa Santos (1989): "Na Ciência Moderna, a ruptura epistemológica simboliza o salto qualitativo do conhecimento de senso comum para o conhecimento científico; na Ciência Pós-moderna o salto mais importante é o que é dado do conhecimento científico para o senso comum. O conhecimento científico só se realiza enquanto tal, na medida em que se converte em senso comum."

Embora esta posição não seja universalmente aceite pois há quem se resigne sobre o facto da educação científica e da própria ciência correrem contrariamente ao senso comum, a chave da configuração poderá estar numa verdadeira "acção comunicacional", como diria Habermas (1987).

Como refere Ogborn (1988) "a verdade científica já não se encontra na cabeça (posição racionalista), nem lá fora (posição empirista): a verdade tem de ser construída entre nós!" (veridicção).

A solução para tal problema passa pela adopção de modelos pedagógicos que reflictam um compromisso entre os modelos perceptivos e os conceptuais, tendo tanto quanto possível a intuitividade dos primeiros e o poder explanatório dos segundos. (Note-se que em termos do paradigma computacional (Johnson Laird, 1988), adoptado pela Ciência Cognitiva, os dois conceitos não se confundem, já que a explanação significa apenas computabilidade, i.e. possibilidade de descrição através de um programa).

Para se conseguir a convergência da intuição com a explanação, poder-se-á recorrer a super - simplificações ou mesmo a ideias intuitivas rejeitadas pela Ciência. Note-se que a transigência pode ir desde a adopção de caracterizações pictóricas fantasistas (caso dos modelos atómicos) até à admissão do deposto princípio de causalidade e consequentemente

do determinismo, passando por quantizações ou tratamentos estatísticos simplificados, ou ainda pela simulação da dificilmente modelada dependência temporal.

Quanto ao papel do intuitivismo nos modelos pedagógicos basta, para uma cabal justificação, ir buscar o construtivismo Piagetiano (Piaget, 1950) consubstanciado nas ideias de assimilação acomodação e muito principalmente as ideias de Bruner (Bruner, 1968) quanto à necessidade do conhecimento passar pelas fases: activa (modelo perceptual), icónica (modelo pedagógico) e simbólica (modelo científico).

Relativamente à questão da explanação ou se fica pelo nível intuitivo, o que é geralmente suficiente para os modelos estruturais, ou se recorre a simulações do comportamento aleatório do sistema (modelos estocásticos) ou de relações causais entre as variáveis do sistema (modelos determinísticos). O caso dos modelos dinâmicos (Richmond, Peterson & Vescuso, 1987) é muito mais delicado, em virtude de haver uma tendência que tem que ser corrigida por técnicas adequadas de modelagem (ver 4) para se ser assertivo em vez de explanativo. Estas correcções tenderiam, para além disso, a modelar o pensamento em termos de relações circulares de causalidade (em vez das mais usadas relações causais lineares) e ainda para em vez de destacar factores causais independentes os integrar numa teia de interdependência, devendo também, tanto quanto possível, explicar o comportamento do sistema em função da sua estrutura interna e não de "forças" que lhe são exteriores. Todo este processo levaria a encarar o comportamento dinâmico dos sistemas numa perspectiva de *prossecução de objectivos*, perspectiva cujo valor universal como base do pensamento teórico ultrapassa, em muito, os limites da Educação em Ciência.

Em todos os casos, excepto talvez na modelagem estrutural, o computador como armazenador e manipulador de símbolos por excelência desempenha e desempenhará cada vez mais um papel fundamental muito particularmente pela facilidade da implementação da iteração que é essencial para a prossecução de objectivos.

6. Taxonomia para a utilização de modelos de computador como estratégias de confrontação

Para atingir o objectivo de suporte de modelos que possam servir de base a uma eficaz confrontação (a modelagem em si é uma actividade mental) os computadores podem ser usados de múltiplas maneiras.

Uma primeira classificação pode assentar no facto de os modelos serem de carácter estocástico ou determinístico. Uma outra (Ogborn, 1988) mais prática distinguirá:

A – Uso de um programa escrito em linguagem normal (Basic, Pascal, ...) de computador, com mais ou menos gráficos e/ou animação, com um grau limitado de interactividade.

B – Idem mas recorrendo a linguagens especiais de modelagem, como o MODL.

C – Utilização de uma folha de cálculo comercial (Fig. 3) em que se aproveitam os factos de as células serem calculadas sequencialmente e de haver possibilidade de iteração. Tal folha de cálculo poderá estar associada a um programa de gráficos como acontece em muitos integrados.

	A	B	C	D	E	F
1	Entalpy	92000				
2	Change	298		L	6.02E23	
3	Temp			k	1.38E23	
4				Lk	8.3076	
5			Standard	Standard	Partial	Further
6			entropy	entropy	pressure	entropy
7	Event	moles	per mole	change	atm	change
8	Lose N2	-1	191.4	-191.4	0.0153	-34.74
9	Lose H2	-3	130.6	-391.8	0.0458	-76.83
10	Gain NH3	2	192.5	385	0.9389	1.05
11	Standard state			———	Alter by:	———
12	entropy change of chemicals			-198.2		-110.52
13	entropy change of surronds			308.72		
14	total change of entropy			———	total	———
15	(standard state)			110.52	entropy	0.00
16				———	change	———
17				equil. constant K		599560
18	Gibb's free		ln K			13.30
19	energy change		LK ln K			110.52
20	(standard state)		-32936	LTK lnK		32936

Fig. 3 - Equilíbrio Processo de Haber

D - Utilização de uma aplicação de modelagem que não é mais do que um interpretador em Basic das equações incrementais do modelo ou então uma folha de cálculo com janelas denomináveis e adaptáveis à função que as variáveis que representam desempenham no sistema. Diferentes graus de sofisticação são conseguidos com um DMS (Fig. 4) com as suas três janelas para condições iniciais, equações do modelo e gráficos (do primeiro tipo) ou com o Cellular Modelling System (Fig. 5) ou o Stella (Fig. 10) (do segundo tipo).

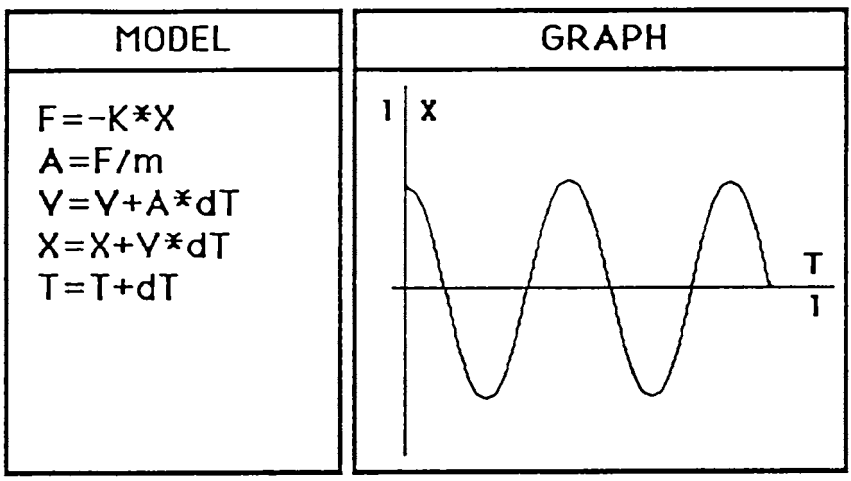


Fig. 4 - Dynamic Modelling System

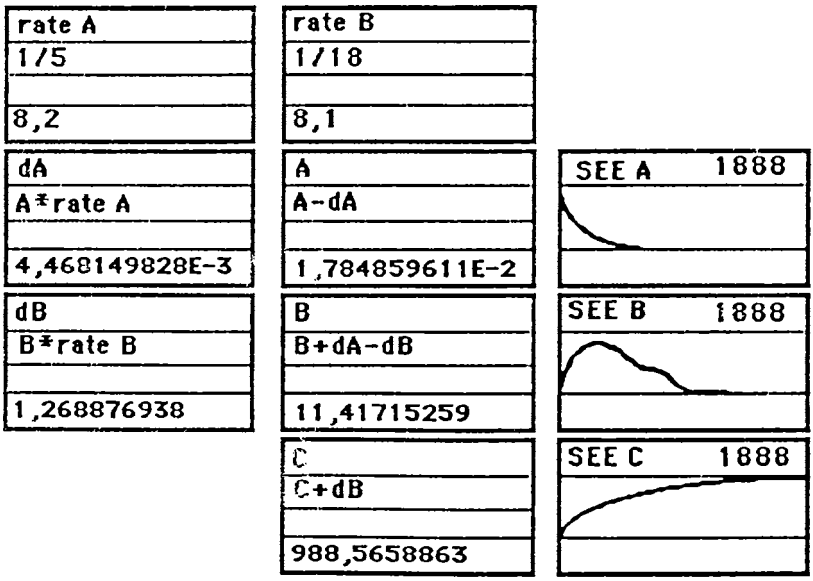


Fig. 5 – Cellular Modelling System

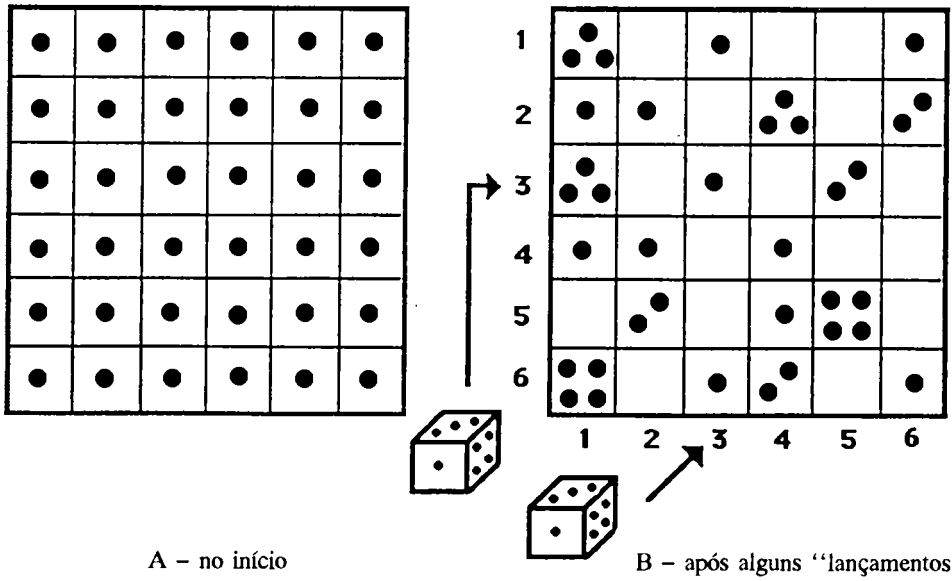
Do ponto de vista educativo é também importante distinguir entre a simples utilização de um modelo criado por outrem e a criação dos próprios modelos.

Veremos, a terminar, a exemplificação dum modelo estatístico ou estocástico implementado em Basic (A) e de um dinâmico ou determinístico em Stella (D, segundo tipo).

7. Dois exemplos

7.1. Um Modelo Estocástico

Este modelo tem por base uma simplificação que leva a considerar um sólido como um conjunto de osciladores harmónicos, independentes entre si que podem estar em estados energéticos correspondentes a 0 ou a um número inteiro dos mesmos quanta de energia. Uma concretização de tal modelo pode ser feita para o caso de 36 "átomos" por um tabuleiro como o da Fig. 6, em que os quanta são simbolizados por marcas e os "átomos" pelas casas.



A – no início

B – após alguns "lançamentos"

Fig. 6 – Modelo do sólido de Einstein para 36 osciladores

As movimentações são feitas de acordo com os lançamentos de dois dados (um referindo-se ao nº da linha e outro à coluna). Começando-se por um estado inicial de um quantum por oscilador (Fig. 6A), ao fim de algum tempo atingem-se configurações mais prováveis (Fig. 6B), verificando-se que a probabilidade das configurações, embora se continuem a processar as transformações a nível atómico, estabiliza razoavelmente (Fig. 7).

Número de lançamentos	n_0	n_1	n_2	n_3	n_4	n_5	n_6	W (Probabilidade)
5	5	26	5					6×10^{10}
10	7	22	7					1×10^{14}
20	13	12	9	2				2×10^{17}
30	17	8	5	6				3×10^{17}
40	16	11	2	7				1×10^{17}
50	17	11	3	2	2	1		1×10^{18}
60	16	13	2	2	2	1		4×10^{17}
70	20	8	2	2	2	2		2×10^{17}
80	18	10	2	3	2	1		7×10^{17}
90	20	8	3	1	1	3		1×10^{17}
100	19	9	3	2	1	1	1	7×10^{17}

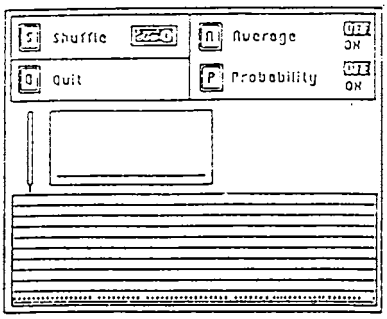
Fig. 7 – Evolução das configurações do sólido de Einstein

Verifica-se também ainda a tendência para a constância da razão –r– dos níveis de ocupação de estados quânticos sucessivos, o que na prática implica uma distribuição

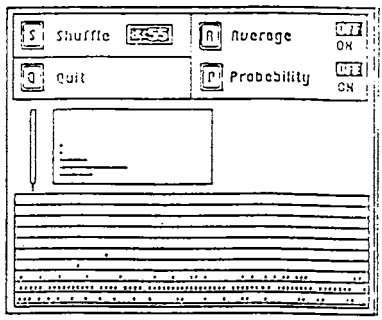
exponencial, podendo a razão da progressão, ou melhor o seu inverso ser tomado como uma medida da temperatura.

Tudo isto pode ser feito muito melhor usando um pequeno programa em computador (Fig. 8), que com a ajuda de alguns gráficos permite não só este exercício de baralhamento de quanta (quantum shuffling) como também a modelagem da mistura de sólidos de Einstein a temperaturas diferentes e a verificação do 2º Princípio da Termodinâmica (Fig. 9).

A – No princípio



B – Depois de algum tempo



C – Mais tarde

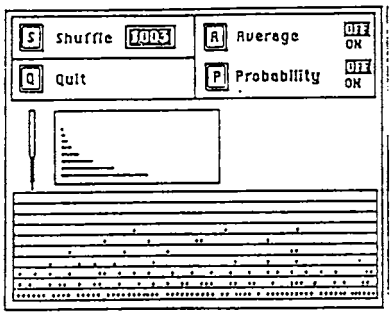
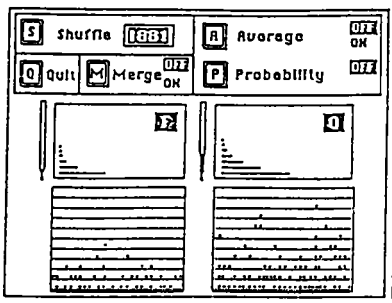


Fig. 8 – Modelo do Sólido de Einstein
Baralhamento de quanta em 200 partículas

A – Os dois corpos isolados



B – Depois de postos em contacto

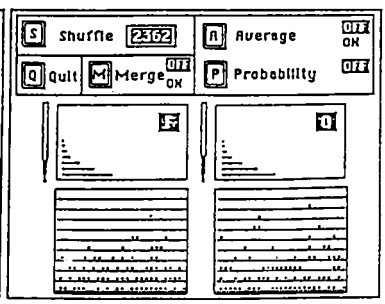


Fig. 9 – Modelo do sólido de Einstein
Equilíbrio térmico

7.2. Um Modelo Determinístico

A aplicação STELLA, para o Macintosh, é um dos melhores e mais versáteis programas de modelagem dinâmica que existe comercialmente.

Baseada numa metodologia recentemente desenvolvida no M.I.T., chamada "System Dynamics" e aproveitando muito bem a interface do Macintosh, assenta no cálculo iterativo de equações incrementais relativas ao tempo, cuja variação é simulada. A ordem dos cálculos é determinada por um diagrama de fluxo (Fig. 10).

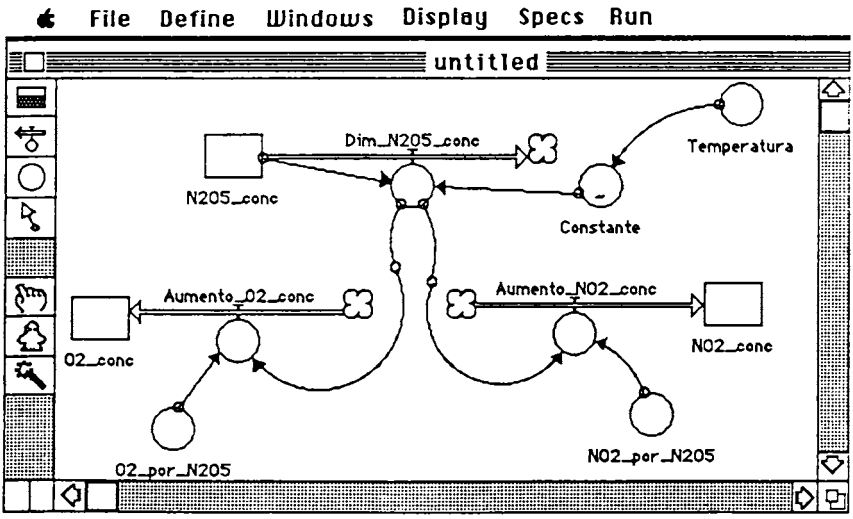


Fig. 10 - Diagrama STELLA do modelo de decomposição do N205

O presente diagrama diz respeito a uma reacção, praticamente completa de 1ª ordem, a decomposição do N205) entre os objectos programáveis, com o auxílio do instrumento representado pelo ícon "mão" ou alienáveis com o auxílio do instrumento representado pelo ícon da "dinamite" e que são (Fig. 11):

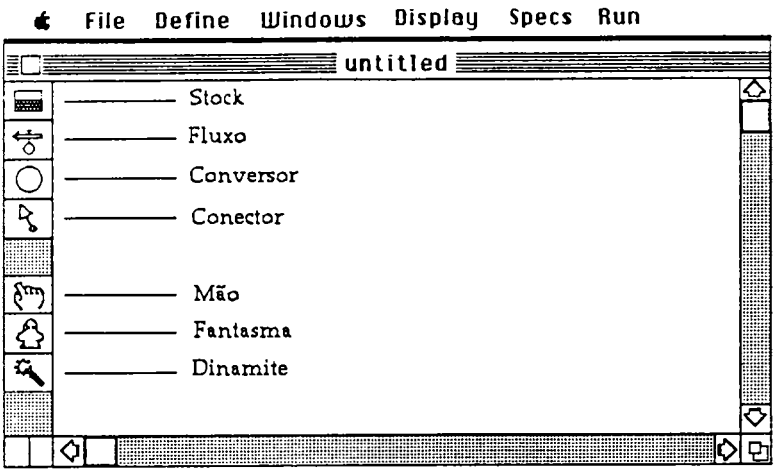


Fig. 11 - Objectos e instrumentos de programação do Stella

- variáveis de "stock" (neste caso as concentrações das espécies químicas em

presença),

- variáveis de fluxo, que são os incrementos (positivos ou negativos) das variáveis de "stock",

- conversores, que estabelecem as relações existentes entre as várias variáveis-conectores, que ligam as variáveis entre as quais existe uma relação de causa e efeito, para além dos fluxos.

- núvens que significam as ligações ao exterior, ou as partes do sistema não explicitadas.

Stocks e conversores, são, de facto, janelas que correspondem a células de uma folha de cálculo, que podem ser abertas e aonde existe espaço para serem escritas as equações do modelo. No caso dos stocks figuram as equações que definem o valor inicial das variáveis. Nos conversores as equações incrementais definindo os fluxos são estabelecidas bem como outras relações, sob uma forma algébrica (Fig. 12) ou gráfica.

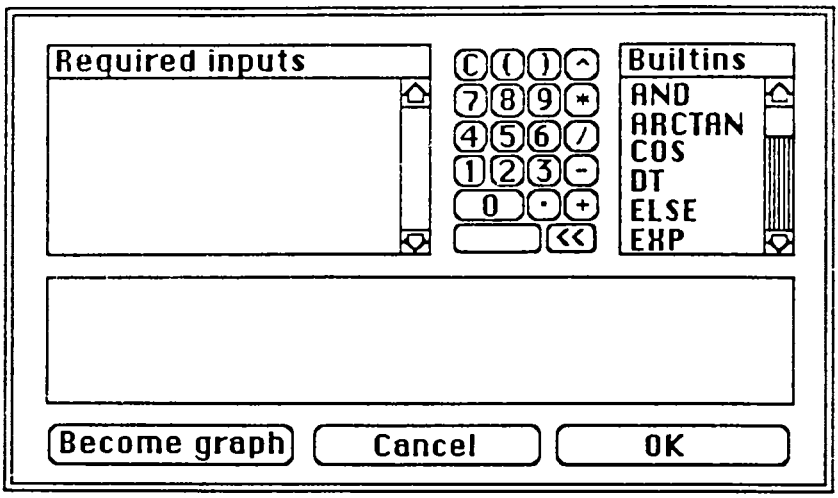


Fig. 12 – Uma janela correspondente a uma variável de stock

A simulação do tempo pode ser desencadeada a partir do menu RUN, notando-se no diagrama o esvaziamento de alguns stocks e o enchimento de outros. Uma melhor maneira de seguir a simulação é através da opção gráficos (Fig. 10) do Menu Windows (Fig. 13).

Note-se por exemplo a influência da temperatura na velocidade da reacção, que pode ser simulada por alteração na janela respectiva do valor da temperatura. Este exemplo é um caso bastante simples, mas basta passar para um caso também simples duma reacção desta vez reversível, para que as coisas se compliquem bastante (Fig. 14).

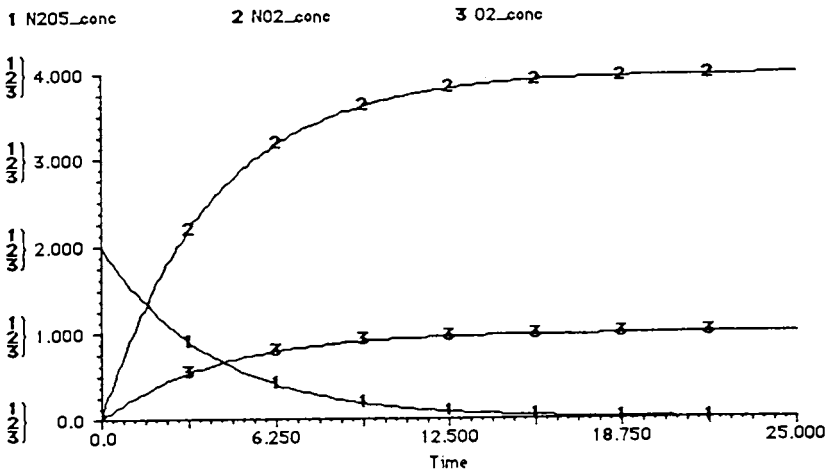


Fig. 13 – Gráfico da variação temporal de algumas variáveis

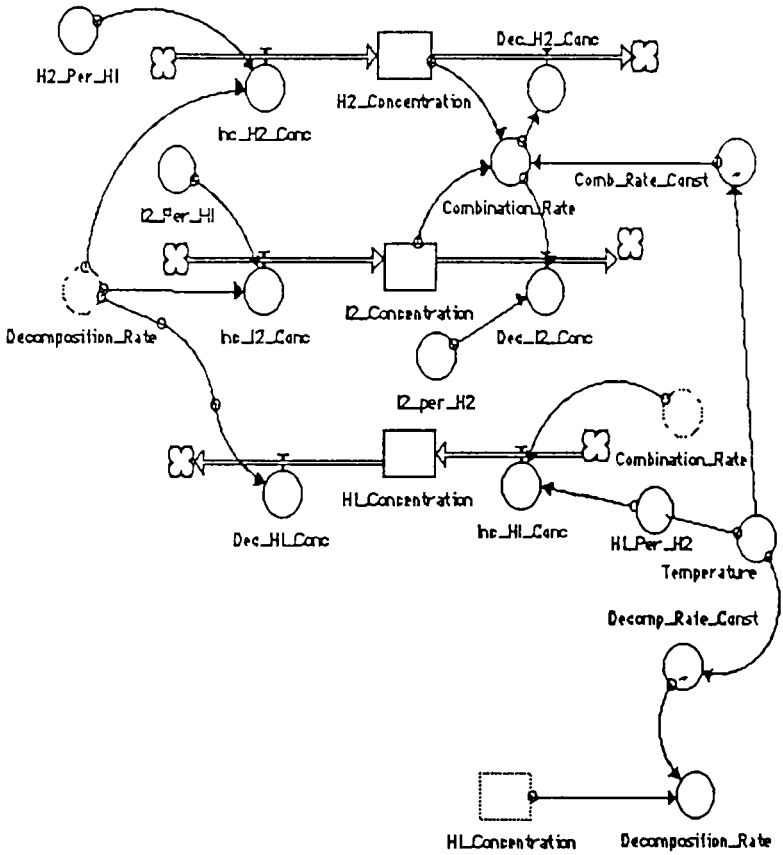


Fig. 14 – Diagrama STELLA relativo à reacção H2+I2=2HI

O valor educativo dos programas de modelagem dinâmica deste tipo, está em não só poderem esclarecer os casos concretos, sem recurso a interpretações microscópicas, como muito principalmente treinarem o pensamento teórico das questões dinâmicas, quer no esclarecimento da estrutura causal dos fenómenos, em unidades que se designam por ciclos causais fechados positivos ou negativos numa perspectiva de prossecução de objectivos, quer ainda no relacionamento destas estruturas (Richmond, Peterson & Vescuso, 1987), com os comportamentos dinâmicos típicos dos sistemas, como por exemplo: crescimento em S (Fig. 15A e B), crescimento rápido e colapso (Fig. 16A e B) e oscilação (Fig. 17A e B).

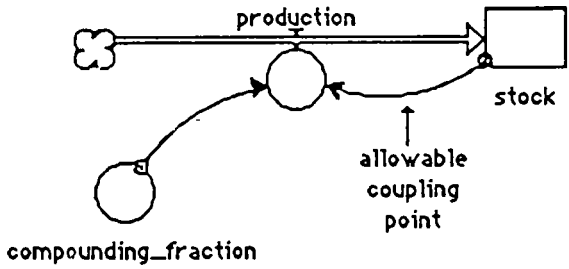


Fig. 15 A

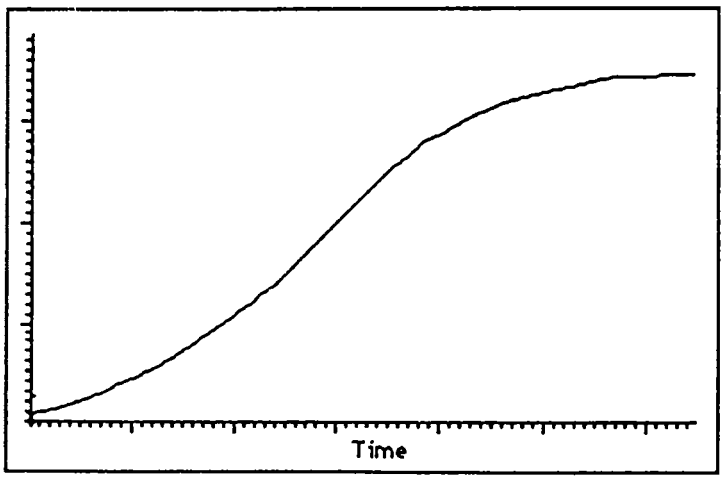


Fig. 15 B

Fig. 15 – Estrutura de composição associada ao comportamento dinâmico de crescimento em S

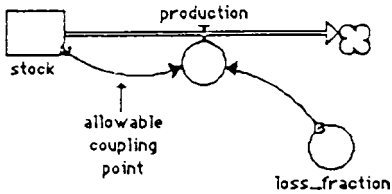


Fig. 16 A

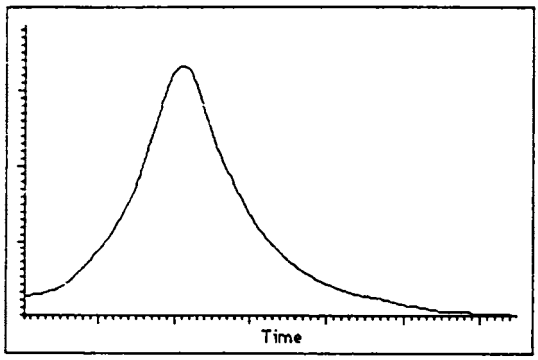


Fig. 16 B

Fig. 16 – Estrutura de drenagem associada ao comportamento dinâmico de crescimento rápido e colapso

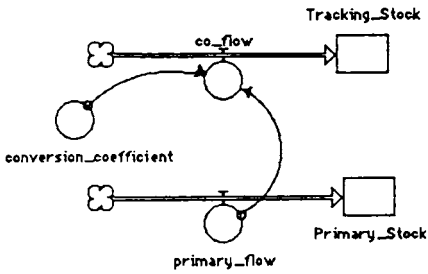


Fig. 17 A

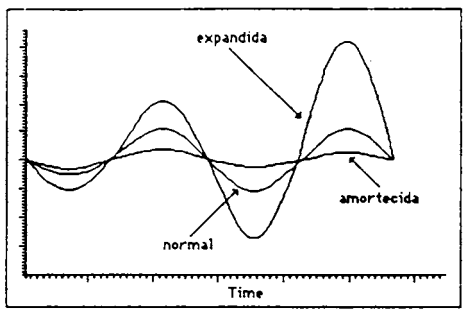


Fig. 17 B

Fig. 17 – Estrutura de co-fluxo associada ao comportamento dinâmico de oscilação.

Para além disso e acima de tudo importa frizar a verdadeira interdisciplinaridade dos comportamentos dinâmicos, que se aplicam indiferentemente às questões das Ciências sociais como das Ciências puras o que contribui, juntamente com a aproximação dos paradigmas (ambos estocásticos) de ambos os tipos de ciências (Vide Introdução), para a pacificação entre cientismos e racionalismos extremos e para que a Ciência desempenhe um papel de relevo na CULTURA, que é inteiramente merecido mas de que anda há longo tempo arredada.

Nota

(1) A configuração corresponderá, no processo ensino/aprendizagem, a uma entidade com estatuto epistemológico intermédio entre os modelos perceptuais (intuitivos) e os conceptuais que são explanatórios.

Bibliografia

Anderson, J. R, *The Architecture of Cognition*, Lawrence Erlbaum Ass, 1983

Bent H. A., *The Second Law*, Oxford University Press, New York, 1965

Brachman R. J. & Levesque H. J. (eds), *Readings in Knowledge Representation*, Morgan Kaufmann Pub, Los Altos, 1985.

Bruner J. S., *Towards a Theory of Instruction*, W.W. Norton & Comp, New York, 1968.

Calado J.C.G., *O Diálogo entre a Ordem e a Desordem*, Boletim da S.P.Q., 20(II), 9–11, 1985

Carey, S., *Cognitive Science and Education*, American Psychologist, 41, 1123–1130, 1986

Collins A., Brown, J.S., Larkin, K.M. *Inference in Text Understanding*. In R.S. Spiro, B.C. Bruce, W.F. Brewer (Eds) *Theoretical Issues in Reading Comprehension. Perspectives from Cognitive Psychology, Linguistics, Artificial Intelligence, and Education*, Hillsdale, J.J.: Lawrence Erlbaum Associates, 1980

Costa Pereira D., *As Novas Tecnologias e o Ensino das Ciências. O Caso da Química*, Actas do I Encontro sobre Educação em Ciências, Braga, 1988

Costa Pereira D., Ribeiro M., *Concepções Erradas sobre Energia, Entropia e Conceitos Afins em Alunos do Ensino Terciário*, a aparecer na Revista Portuguesa de Educação, Braga, 1989

Driver, R., *The Pupil as Scientist*, Milton Keynes: Open University, 1983.

Einstein A., *Relativity: The Special and General Theory. A Popular Exposition*, Londres, 1920.

Gilbert, J.K. & Watts, D.M., *Concepts, Misconceptions and Alternative Conceptions: Changing Perspectives in Science Education*, 10, 61–98, 1983.

Gregory R.L., *Mind in Science, a History of Explanations in Psychology and Physics*, Weidenfeld and Nicolson, Londres, 1981.

Habermas J., *A Ideia da Universidade: Processos de Aprendizagem,* Revista de Educação, 2, 1, 3–9, 1987.

Herbenstreit J., *L'usage des logiciels d'éducation*–Conferência Plenária na Universidade de Verão da CEE–Gent, 1986.

Hinton G.E. & Anderson J.A., *Parallel Models of Associative Memory*, Lawrence Erlbaum, New Jersey, 1981.

Johnson–Laird P. N., *Mental Models*, Cambridge University Press, Cambridge, 1983.

Johnson–Laird P.N., *The Computer and the Mind*, Harvard University Press, Cambridge Massachussets, 1988.

Jurs, C., *Computer Software Applications in Chemistry*, J. Wiley & Sons, 1986.

Kelly G., *A Brief Introduction to Personal Construct Theory* in Bannister D. (ed) *Perspectives in Personal Construct Theory*, Academic Press, London, 1977.

Kintsch W., *Memory and Cognition*, Wiley, New York, 1977.

Kosslyn S. M., *Image and Mind*, Harvey University Press, Cambridge Massachussets, 1980.
Kuhn, T., *The structure of Scientific Revolutions*, Chicago, 1970.
Madureira, C., *Introdução à Informática*, CIUP, Porto, 1984.
Mc Clelland J.L., Rumelhart D. E. and the P.D.P. Research Group, *Parallel Distributed Processing. Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models*, Bradford Books, MIT Press, Cambridge Massachussets, 1986.
Minsky, M., *A Framework for Representing Knowledge*. In P.H. Winston (Ed) *The Psychology of Computer Vision*, New York: McGraw-Hill, 1975.
Newell A., *Production Systems* in Chase W. C. (ed), *Visual Information Processing*, Academic Press, New York, 1973.
Ogborn J., *The Nature of Science and its Implications in Science For All*, Actas do I Encontro sobre Educação em Ciências, 107–124, Braga, 1988.
Pask, G., *Styles and Strategies of Learning*, British Journal of Educational Psychology, 46, 147–148, 1972.
Piaget, J., *The Psychology of Intelligence*, Routledge and Kegan Paul, 1950.
Popper K., *Conjectures and Refutations: the Growth of Scientific Knowledge*, Routledge & Kegan Paul, Londres, 1972.
Quillian R., *Behavioral Science*, 12, 410–430, 1967.
Richmond B., Peterson S. & Vescuso P., *An Academic User's Guide to Stella*, High Performance Systems, New Hampshire, 1987.
Rosado L., *Modelos en los Processos de la Ciencia y su contrastácion*, UNED, Madrid, 1986.
Sanford A.J. & Garrod, S.C., *Towards a Psychological Model of Written Discourse Comprehension*. In J.-F. Le Ny and W. Kintsch (Eds) *Language and Comprehension*. Amsterdam: North-Holland, 1982.
Schank R.C. & Rieger III C.J., *Inference and the Computer Understanding of Natural Language*, Artificial Intelligence, 5(4), 373–412, 1974.
Shepard R. N. & Cooper L.A., *Mental Images and their Transformations*, Bradford Books, MIT Press, Cambridge Massachussets, 1982.
Smith A., Shoben J. & Rips L.J. *Structure and Process in Semantic Memory: a Featural Model for Semantic Decisions*, Psychological Review, 81, 214–41, 1974.
Sousa Santos, B. *Introdução a uma Ciência Pós-Moderna*, Afrontamento, Lisboa, 1989.
Thomaz M.F., *Uma Perspectiva Construtivista para o Ensino da Física I - Psicologia da Construção Pessoal de George Kelly*, Gazeta da Física, 10, 4, 121–128, 1987.
Wilson S., *Chemistry by Computer*, Plenum Press, New York, 1986.
Wittgenstein L., *On Certainty*, Blackwell, Oxford, 1969.

João Pedro Ponte

Dep. de Educação, Fac. de Ciências
Univ. de Lisboa, Lisboa (Portugal)

O computador como ferramenta: resultados de investigação

Resumo

As investigações aqui consideradas traduzem parte da actividade realizada entre 1984 e o presente no DEFCUL (ligado desde 1985 ao projecto Minerva), constituindo teses de mestrado ou provas de aptidão pedagógica e capacidade científica. Embora estas investigações tenham o aspecto comum da utilização do computador como ferramenta de trabalho, referem-se, no entanto, a diferentes objectivos educacionais, níveis de escolaridade, conteúdos de ensino e abordagens pedagógicas.

Em todos os estudos realizados evidencia-se de modo flagrante a importância do papel do professor. Quanto este se interessa de forma genuína pelo computador induz nos alunos um comportamento semelhante e os resultados não se fazem esperar.

Assim, os resultados positivos dos diversos estudos considerados não derivam simplesmente da introdução do computador como ferramenta de trabalho, mas estão ligados igualmente à forma como este foi introduzido e aos ambientes pedagógicos criados.

As dificuldades específicas com a implementação deste tipo de uso do computador são reais (cfr. Ponte, 1989). No entanto, a experiência já adquirida e os resultados alcançados permitem considerar esta via como uma das mais promissoras no que respeita à utilização educativa das novas tecnologias da informação.

Abstract

We consider here some researches which took place within the activities occurred since 1984 in the DEFCUL (connected with the Minerva Project since 1985), related to mastery thesis and tests of pedagogic ability and scientific capability. Having in common the use of the computer as a tool, these researches are referred with educational goals, scholarship levels, learning contents and pedagogic approaches which are rather differentiated.

All the studies show clearly the relevance of the teacher role. When the teacher is truly interested in the computer, he induces the same attitude amongst the students, with fast results.

Thus, the positive results aren't a mere consequence of the introduction of the computer as a tool, but also depend on the way of its introduction and the inherent pedagogic environments.

The specific dificulties with the implementation of this kind of computer use are real (Ponte, 1989). Nevertheless, the already acquired experience and the output results allow us to consider this issue as an extremely promising one amongst the educational uses of new information technologies.

Desde os anos sessenta que se procuram estudar os contributos do computador para a mudança dos processos de ensino. A perspectiva de investigação mais difundida neste

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 431/437
1992, APDC & SPF, Lisboa

domínio refere-se ao chamado "Ensino Assistido por Computador". O termo não será o mais feliz, na medida em que o que se tem procurado desenvolver não são simples meios auxiliares do professor mas aplicações dotadas de plena autonomia (programas tutoriais e de prática) que o possam substituir, pelo menos parcialmente, em algumas das suas tarefas.

Os resultados obtidos são de natureza muito variável, tanto positivos, como neutros e negativos. Verifica-se que podem exisitir ganhos significativos em certas aprendizagens, relativas a conceitos ou habilidades muito específicas, sobretudo quando os alunos se encontram bastante motivados. Estes ganhos tendem contudo a diluir-se com o decorrer do tempo, sobrevindo por vezes fenómenos de rejeição perante utilizações demasiado prolongadas (Clark, 1983; Hativa, 1988).

Esta perspectiva de pesquisa encontra-se longe de esgotada. Tirando partido das possibilidades da Inteligência Artificial, uma nova geração de investigadores procura colmatar as deficiências e limitações emergentes neste tipo de programas, dando corpo ao "Ensino Assistido Inteligente".

A introdução do Ensino da Programação nos currículos escolares de alguns países deu origem a outra corrente de investigação. Procura-se conhecer as dificuldades e as estratégias dos alunos na sua aprendizagem (Rogalski 1986; Rouchier, 1987).

A programação pode ser encarada como objecto de estudo, que vale por si mesmo, no quadro duma disciplina de Informática. Pode, alternativamente, ser vista como um elemento auxiliar para o estudo de outros assuntos ou para o cumprimento de objectivos bem determinados de articulação curricular, como tende a suceder com a linguagem LOGO nos níveis mais elementares da escolaridade. Neste momento, o lugar que o ensino da programação deverá ter nos currículos escolares continua a ser tema de discussão e controvérsia.

O computador pode assim ser visto como um *objecto de estudo* ou como um *substituto do professor*. Uma análise das investigações realizadas tendo em conta estas perspectivas de utilização do computador pode ser encontrada em Ponte (1986a). Neste texto, procura-se discutir, com base na investigação realizada no nosso país, em termos de uma terceira concepção de utilização do computador, em que este é essencialmente encarado como uma *ferramenta de trabalho* nas mãos dos alunos.

Curiosamente, esta forma de utilização do computador, embora seja a que mais se tem desenvolvido na maioria dos países, incluindo no nosso, não assenta nas análises dos grandes teóricos da educação, nem nos resultados dos investigadores, nem nas directrizes dos aparelhos administrativos da educação. Impôs-se por si mesma, no terreno da prática, na sequência da generalização nas pequenas e médias empresas e nas profissões liberais do uso de programas de aplicação como o processamento de texto, a base de dados e a folha de cálculo, permitindo fazer de forma mais rápida, mais eficiente e mais perfeita funções que tradicionalmente eram realizadas por outros processos (Fonseca, 1986; Ponte, 1986b, 1989; Serrazina, 1988).

Enquadramento geral das investigações

As investigações aqui consideradas traduzem parte da actividade realizada entre 1984 e o presente no DEFCUL (ligado desde 1985 ao Projecto MINERVA), constituindo teses de mestrado ou provas de aptidão pedagógica e capacidade científica. Convém notar que, muito embora estas investigações tenham como aspecto comum a utilização do computador como ferramenta de trabalho, referem-se, no entanto, a objectivos educacionais, níveis de escolaridade, conteúdos de ensino e abordagens pedagógicas diferenciadas.

Maria Augusta Neves (1988) desenvolveu um estudo com 24 alunos do 9º ano, duma turma especial com profundo insucesso em Matemática. Utilizou o programa GEM Paint e a linguagem LOGO no ensino da Geometria, em abordagens baseadas em actividades de natureza estruturada. Os alunos trabalharam com o computador em grupos de 2, durante 15 aulas.

Utilizando programas de processamento de texto, Maria dos Anjos Caseiro (1989) implementou e avaliou um programa de supervisão que abrangeu 9 professores de diversas disciplinas, sobretudo da área de letras, tendo envolvido 200 alunos do 7º ao 11º ano de escolaridade. Os alunos usaram o computador na sala de aula, em actividades de grupo, na elaboração de textos ligados aos respectivos conteúdos programáticos, por um período de quatro meses.

Leonor Moreira (1989) desenvolveu uma investigação com 18 alunos do 6º ano de escolaridade, na disciplina de Matemática, utilizando a folha de cálculo Supercalc 3. Os alunos trabalharam em grupos de três, numa primeira fase em ambiente extra-lectivo, numa segunda fase em 8 aulas, e finalmente numa terceira fase, de novo em ambiente extra-lectivo.

Dado o enquadramento das tarefas realizadas pelos alunos, tem interesse considerar igualmente nesta análise um estudo realizado por João Filipe Matos (1987), embora este se baseie no uso duma linguagem de programação e não numa ferramenta informática. Este autor estudou o ambiente de aprendizagem criado com a utilização da linguagem LOGO, numa investigação que envolveu 39 alunos do 1º e 4º ano de escolaridade. Os alunos trabalhavam em grupos de dois, decorrendo as actividades com o computador na sua sala de aula, em articulação com as restantes actividades escolares, sob orientação das suas professoras. As observações realizadas referem-se a um período de 6 meses.

Dificuldades dos alunos no domínio das ferramentas informáticas

Os programas comerciais não foram concebidos para utilização didáctica. São por vezes relativamente complicados no seu funcionamento, requerendo a memorização de um grande número de comandos ou a aprendizagem da "navegação" através de vários níveis de "menús". Além disso, estes programas, na sua maioria, estão em inglês ou têm traduções muito defeituosas. Em que medida constituem estes aspectos obstáculos á sua utilização pelos alunos?

No trabalho desenvolvido por Neves (1988), tanto o programa GEM Paint como a linguagem LOGO foram introduzidos não como objectos de estudo em si mesmos, mas através da abordagem de tópicos de Geometria. Ambos os programas permitiriam uma fácil familiarização dos estudantes com o computador, não se verificando dificuldades particulares no seu domínio.

No seu estudo sobre o processamento de texto, Caseiro (1989) utilizou os programas Tasword 2 (em computadores Sinclair Spectrum e Timex) e Open Access (em computadores Amstrad), que se podem considerar como relativamente primitivos. Embora os alunos não tivessem experiência de trabalho com computadores nem com máquinas de escrever, desconhecendo por isso o teclado, rapidamente conseguiram um domínio satisfatório dos programas, passando a utilizar o computador com um sentimento de desinibição e confiança.

Por sua vez, Moreira (1989) indicou que o diálogo dos alunos com o computador foi inicialmente dificultado pelo facto do programa se encontrar em inglês, mas com a familiarização progressiva o seu domínio pelos estudantes foi-se tornando cada vez mais

fácil e natural. Os alunos só utilizaram um número limitado de comandos e funções, mas a sua estrutura complexa não constituiu obstáculo a um domínio relativamente rápido do programa, possibilitando o seu uso para as tarefas propostas. Trata-se de um resultado particularmente significativo se pensarmos no nível etário dos alunos envolvidos neste estudo.

Finalmente, Matos (1987) verificou que a linguagem LOGO está ao alcance dos alunos do Ensino Primário, considerando no entanto importante a utilização de uma versão adequada aos seus processos cognitivos.

Assim, nas condições em que estes estudos se desenrolaram, os alunos adaptaram-se bem aos programas, que passaram a dominar pelo menos na medida das suas necessidades. Os menus e comandos em inglês apenas constituiram um problema na fase inicial de utilização. No entanto, a relativa facilidade em usar versões em língua estrangeira não nos deve fazer esquecer que a questão da língua a usar, se não se coloca nas empresas, já nas escolas é um problema fundamental, de natureza política e cultural. Queremos preservar a nossa língua, com as suas características próprias, como um dos nossos valores culturais mais essenciais ou aceitamos (seja em nome do desenvolvimento económico ou da facilidade de intercâmbio científico) a sua progressiva diluição, com a introdução de termos e construções estranhas?

Aprendizagem de conteúdos específicos

A aprendizagem de conteúdos específicos de natureza disciplinar é favorecida pela utilização pelos alunos do computador como ferramenta de trabalho?

Neves (1980), mostrou que tanto a realização de actividades fortemente estruturadas com o utilitário de desenho GEM Paint, como com a linguagem LOGO, promoveram progressos significativos, quer em termos do domínio de conceitos de Geometria, quer no que respeita à capacidade de resolução de problemas geométricos, não havendo diferenças significativas entre os dois ambientes de aprendizagem.

Segundo Moreira (1989), a utilização da folha de cálculo, no quadro de propostas de trabalho também significativamente estruturadas, embora muitas delas de natureza relativamente aberta, teve um efeito positivo na construção dos conceitos de proporcionalidade e de percentagem, na resolução de situações problemáticas e na análise de gráficos. Esta melhoria pode dever-se a vários factores. Assim, o facto de o programa se encarregar de fazer os cálculos pode libertar a mente para se concentrar nas tarefas de planeamento e controlo da execução. Igualmente, a folha de cálculo obriga a uma explicitação de relações entre variáveis que ajudam a clarificar o problema e a encontrar estratégias de resolução.

No estudo realizado por Matos (1987) tornou-se evidente como as aprendizagens de determinados conteúdos específicos foram profundamente estimuladas pela utilização da linguagem LOGO. Por exemplo, os alunos tiveram de dominar os conceitos de ângulo e de comprimento, o sistema de numeração e particularmente os números decimais para conseguir levar a bom termo os seus projectos.

No estudo empreendido por Caseiro (1989), verificou-se, segundo esta autora, uma melhoria significativa na produção escrita dos alunos, passando estes a escrever mais, com mais facilidade, resultando isso em textos de melhor qualidade.

Podemos concluir que a utilização do computador como ferramenta de trabalho pode favorecer aprendizagens específicas desde que tal seja explicitamente tido em conta na planificação das actividades a propor e desde que haja o devido acompanhamento por parte do professor.

Estratégias cognitivas gerais

No entanto, a utilização do computador, mais do que promover a aprendizagem de conteúdos específicos, terá interesse na medida em que favorecer aquisição de novas estratégias de raciocínio, de novas capacidades metacognitivas, de novas competências em termos de resolução de problemas.

Colocados perante um problema, os alunos têm normalmente a tendência para começar imediatamente a fazer contas, deixando-se conduzir por aquilo que identificam como "palavras-chave", sem estudar com cuidado o enunciado, sem verificar o que é dado e o que é pedido, e sem reflectir sobre a possível existência de várias estratégias a seguir. No seu estudo, Moreira (1989) indicou que os alunos atenuaram esta tendência, demorando-se mais na análise dos problemas antes de entrar na fase de resolução. Verificou igualmente a tendência para espontaneamente experimentarem por sua iniciativa novos valores, novas relações entre os dados e fazerem a sua visualização gráfica, aspectos que raramente se observam no ensino usual.

Neves (1988) mostrou que os alunos modificaram a sua visão global relativamente à Geometria, evoluindo num sentido positivo no que respeita à sua opinião sobre a utilidade prática do assunto, à facilidade experimentada em estudá-lo e à sua satisfação com o seu estudo e com os conhecimentos adquiridos.

Por seu lado, Caseiro (1989) refere que os alunos perderam o receio de cometer erros e interiorizaram bem o valor, no processo de elaboração de um texto, duma fase de revisão crítica.

Matos (1987) refere que no trabalho com o computador se desenvolvem novas aprendizagens, quer de natureza prática, quer de natureza conceptual. No entanto estas não são muitas vezes reconhecidas pelos alunos como tendo um estatuto de importância comparável ao das aprendizagens curriculares tradicionais.

Destes estudos podemos concluir que a utilização do computador como ferramenta de trabalho, desde que situada num quadro geral de actividades e de relações de trabalho apropriadas, tende de facto a favorecer o desenvolvimento de novas estratégias cognitivas nos alunos.

Domínio socio-afectivo

A introdução do computador na sala de aula é susceptível de criar novas dinâmicas e novas relações entre os intervenientes no processo educativo.

Nos quatro estudos considerados verificou-se uma grande alegria, entusiasmo e empenhamento dos alunos pelas tarefas a realizar no computador. Matos (1987), cujo estudo é o que mais se prolonga pelo tempo, refere no entanto que a partir se certa altura se começou a assistir a uma certa diferenciação, com alguns alunos a envolverem-se cada vez mais e outros a diminuir a frequência das suas actividades.

Neves (1988), que tinha no seu estudo alunos em situação de insucesso profundo, refere uma evolução radical da sua atitude perante a escola e a disciplina de Matemática.

As dinâmicas de grupo colocam sempre problemas delicados de gerir. Moreira (1989) indica que os seus grupos eram no início heterogéneos, verificando-se fenómenos de liderança e obstrução por parte dos alunos com classificação superior em Matemática à iniciativa dos colegas, o que levou a que os grupos fossem reconstituídos de forma tanto possível homogénea relativamente ao aproveitamento nesta disciplina.

Caseiro (1989) indica que no decorrer do seu estudo se desenvolveu entre os alunos

o espírito de colaboração e entreajuda. Esta autora refere uma alteração verificada no clima social da sala de aula, com um estreitamento da relação professor-aluno. Segundo ela, o professor passou a ter uma função facilitadora da aprendizagem dos alunos, que por sua vez deixaram de ver o como detentor e transmissor do conhecimento já construído.

Por seu lado, Matos (1987) refere igualmente a criação de novos papeis dentro da sala de aula, nomeadamente o surgimento dos "alunos especialistas" que em alguns aspectos de ordem técnica chegam a ser mais expeditos que o professor e se prestam a ajudar os seus colegas a ultrapassar dificuldades.

As investigações realizadas por Caseiro (1989), Moreira (1989) e Neves (1988), embora tivessem decorrido, pelo menos parcialmente, em períodos correspondentes às aulas dos alunos, envolveram duma forma geral turmas pequenas ou grupos ainda mais reduzidos. Torna-se por isso difícil de generalizar os seus resultados em termos de dinâmica geral da sala de aula.

Conclusão

Em todos os estudos realizados evidencia-se de modo flagrante a importância do papel do professor. Quando este se interessa de forma genuína pelo computador, induz nos alunos um comportamento semelhante e os resultados não se fazem esperar.

Assim, Matos (1987) indica que o interesse e envolvimento manifestado pelos professores é acompanhado pelos alunos, que revelam um empenhamento mais significativo nas tarefas que sentem ser mais valorizadas. Nesta investigação colaboraram professoras que já anteriormente procuravam proporcionar aos alunos um clima de trabalho virado para o desenvolvimento da sua autonomia e responsabilização pelo processo de aprendizagem. Este autor refere que o ambiente de aprendizagem é particularmente estimulado pela realização de projectos negociados entre os alunos e o professor no âmbito de actividades de natureza mais geral.

No caso de Neves (1988), parece ter sido fundamental para o êxito dos alunos, que à partida se encontravam em situação de insucesso profundo, a relação que com eles foi estabelecida, que os retirou da situação de marginalidade em relação à escola, os dignificou e os valorizou como pessoas.

Na investigação conduzida por Caseiro (1989), terá sido determinante o quadro pedagógico criado, que teve como actividade organizadora a elaboração de jornais de turma com textos de diversos géneros literários, subordinados a temáticas decorrentes dos conteúdos programáticos das disciplinas respectivas.

Assim, os resultados positivos dos diversos estudos considerados não derivam simplesmente da introdução do computador como ferramenta de trabalho, mas estão igualmente ligados à forma como foi este introduzido e aos ambientes pedagógicos criados.

As dificuldades específicas relativas à implementação deste tipo de uso do computador são reais, sendo referidas em pormenor noutro trabalho (Ponte, 1989). No entanto, a experiência já adquirida e os resultados alcançados permitem considerar esta via como uma das mais promissoras no que respeita á utilização educativa das novas tecnologias da informação.

Referências

Caseiro, Maria dos Anjos (1989). *A Utilização do Processamento de Texto na Sala de Aula* (Tese de mestrado). Lisboa: Projecto MINERVA, Pólo DEFUCL.

Clark, Richard E. (1983). *Reconsidering reasearch on learning from media. Review of Educational Research,* Vol. 53, p. 445–459.

Fonseca, Eduarda (1986). *Vamos trabalhar com a folha de cálculo.* Lisboa: Projecto MINERVA, Pólo DEFCL.

Hativa, Nira (1988). *Sigal's Ineffective Computer-Based Practice of Arithmetic: A Case-Study. Journal for Research in Mathematics Education,* Vol. 19, nº 3, p. 195–214.

Matos, João Filipe (1987). *A Natureza do Ambiente de Aprendizagem Criado com a Utilização da Linguagem LOGO no Ensino Primário e as suas Implicações na Construção do Conceito de Variável.* (Provas APCC). Lisboa: Projecto MINERVA, Pólo DEFCUL.

Moreira, Maria Leonor (1989). *A Folha de Cálculo na Educação Matemática.* (Tese de mestrado). Lisboa: Projecto MINERVA, Pólo DEFCUL.

Neves, Maria Augusta (1988). *O Computador na Recuperação em Geometria de Alunos do 9º Ano* (Tese de mestrado). Lisboa: Projecto MINERVA, Pólo DEFCUL.

Ponte, João (1986a). *O Computador--Um Instrumento da Educação*; Lisboa: Texto.

Ponte, João (1986b) *Processamento de Texto: Para Desenvolver o Gosto pela Escrita.* Lisboa: Projecto MINERVA, Pólo DEFCUL.

Ponte, João (1989). *O Computador como Ferramenta: Uma Aposta Bem Sucedida? Inovação,* Vol. 2, Nº 1, p. 41–48.

Rogalski, Janine (1986). *Mental Representations in Programming, by 15-16 Years Old students. Proceeedings of the Tenth International Conference for the Psychology of Mathematics Education,* Londres, p. 283–288.

Rouchier, André (1987). *La Structure Iterativo-Repetitive comme Champ Conceptuel dans l'Enseignement Elementaire et Secondaire en Mathematique et Informatique. Proceeedings of the Eleventh International Conference for the Psychology of Mathematics Education,* Montreal, p. 76–82.

Serrazina, Maria de Lurdes (1988). *Consulta e Classificação como Actividades Educativas: Utilização de Bases de Dados.* Lisboa: Projecto MINERVA, Pólo DEFCUL.

Teresa Levy

Dep. de Educação, Fac. de Ciências
Univ. de Lisboa, Lisboa (Portugal)

Computers in use, communication in process

Abstract

Schools cannot remain immune to the transformations brought about by the computerization of society. Educators must discuss the effects these transformations have on educational settings. Still, for us, it is not enough to claim that "computer literacy" must become an important component of general education. We must examine what this term implies. For us, it must mean more than the ability to be proficient in the use of computers. A functional understanding of how a technology is used must not treat technological devices in isolation by view them within the larger context of human practices. Moreover, we must examine the assumptions upon which work with computers develops. We propose to participate in that discussion by looking, in particular, at the notions of language and communication and by questioning the technomorphic view of human beings as well as the antropomorphic conception of machines.

Resumo

As escolas não podem ficar imunes às transformações acarretadas pela computorização da sociedade. Os educadores têm o dever de discutir os efeitos trazidos por estas transformações às orientações da educação. Para nós, não basta reclamar que a "alfabetização informática" deva tornar-se uma das componentes importantes da educação geral. Impõe-se examinar o que aquela expressão implica. Entendemos que deve significar mais do que a aptidão para um uso proficiente do computador. Uma compreensão funcional do modo como a tecnologia é usada não deve considerar os inventos tecnológicos isoladamente, mas sim vê-los no contexto mais amplo das práticas humanas. Além disso, devemos examinar as assunções que o trabalho com computadores fomenta. Propomo-nos participar no debate considerando em particular as noções de comunicação e linguagem e questionando a visão tecnomórfica dos seres humanos, paralelamente á concepção antropomórfica das máquinas.

Computers are by now everywhere. The computerization of society is well under way. This situation has an important impact on society, on social bonds, on the status of knowledge and the ways of its transmission. The full consequences of such trend are yet unpredictable, but, in no case can schools stay immune to the transformations which are having place both at the level of knowledge and at the level of power. However, or so we are told, there is still, in an important part of the educational sector, a noticeable resistance to a full use of computers and related technology, a frustrating inability to catch up with the knowledge produced by a whole range of areas of research where computers play a major role. With the exceptions of some Faculties or Departments, more technologically

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 438/445
1992, APDC & SPF, Lisboa

oriented and which are usually both producers and consumers of types of knowledge made possible by computers, most schools are bad consumers of technology, incapable of taking advantages of the totality of its potentialities. Inevitably, they are staying behind in the routes paved by progress and, in consequence, they leave themselves open to the criticisms of those who know what is going on that is relevant in our world, of those voiced by experts in leading research areas, and by computer-education experts.

We often hear complaints about the resistance, ignorance and misunderstanding on the part of teachers, shortsightedness of those involved in the management of schools and the school system at large, and of those responsible for the allocation of financial means to the educational systems. It seems as though they have not yet grasp what really is at stake. Worried by this conservative attitude which sometimes is equated with anti-scientific, if not irrational, behavior, those who claim to be more attuned to the transformation of postindustrial societies, in a positive sense, and wish schools to be integrated in the process, have been arguing for the introduction and implementation of new technologies in the educational system at all levels. This claim is often accompanied by a bitter explicit disagreement or by provocative remarks directed against those who are suspicious or skeptical of this implementation and even against those who dare to raise some questions and demand a reflection on the effects that the entrance of the so-called new technologies in schools may have in the processes of learning and in the understanding of culture. For them, skepticism and the raising of questions concerning this domain show nothing but a fear to change or a backwards attitude regarding the function of schools.

For our part, we, lay persons in the area of computing but nevertheless aware of its influence on our lives and on education, on knowledge and learning, while acknowledging that computers are technological devices which are not only here to stay but are part of those radical innovations that shape the potentialities of some networks of human interactions, feel, at the risk of being dismissed as humanists, that we need to gain a functional understanding of how these devices are and can be used, of what they do and how they function in the context of human practices. In fact, for us, computers, like other media, are to be understood in the context of communication and the larger network of practices in which they are inserted. For this, we must find a broad enough framework to think and discuss not only the technical and procedural aspects of computers but also the conceptual, ideological, cultural and educational side of technology. In particular, the manner in which computers influence our views on language and communication, mediate the individual's relationship to the environment, re-encode culture, determining, in some sense, what is to get saved and what, by implication, gets lost in the processes of programming and transmission,leads to important issues that must be considered by educators, if they are to assess what kind of alterations do computers bring into the classroom.

We are here assuming that theories about thought and language, intentionality and meaning, notions concerning the nature of human action have a profound influence on the shape of what we build and how we use it. Society engenders inventions whose existence in turn alters that society. We need to establish a critical stance for looking at what the devices do, not just at how they operate.

In particular, questions about language come to the fore, since we are dealing with a technology that operates in a domain of language. In fact, we are thrown back into questions of language not only because computers are devices for creating, manipulating, and transmitting symbolic objects, but also because computers play a role in mediating, facilitating or repressing linguistic action. We must, therefore, bring to view and examine

the implicit understanding of human thought and language that serves as a background for the development of computer technology. This because the incorporation of technical jargon in everyday discourse, the change in the ways in which we speak about technology alters our language, and our commonsense understanding of ourselves and the world.

In particular, the development of technology has been leading to new uses of terms such as "communication", "information", "knowledge", "language", thereby operating a displacement of their meanings, and this is a movement that requires closer examination, especially because this move often corresponds to a consideration of the the human condition in technomorphic terms.

Ironically, computers which, in principle, only acknowledge literal meanings, have themselves become a generative metaphor for thinking about humans, a metaphor which tends to its naturalization. Now, if there are some computational programs which efficiently simulate some aspects of cognition and decision-making, the transposition of concepts, without qualification from one domain to another can be quite misleading, and reductive. Thinking is not to be reduced to formal processes of cognition, and cannot be fully independent of experience, knowledge is not to be reduced to explicit, procedural forms but encompasses non-formalized, contextualized modalities which require other ways of access than those used in the techno-sciences, and which from the dominant approach privileged today are in danger of seeing themselves disregarded or confined to the domain of irrationality. Lyotard (1979), for example, predicts that any constituted form of knowledge which cannot be translated in computer languages will be abandoned and new knowledge will have to be invented having in mind this constraint. It remains to be explained if that it has to be so or even if they can be so.

In particular, we may also see what is entailed by the comparison of memory with a data bank, recalling with retrieval of information, communication with information transmission. We must think what this use of the computer as a paradigm which frames our way of understanding not only some of man's capabilities but his way of being, while illuminating certain aspects of man's ways of thinking, leaves a whole host of aspects of man's existence and experience unaccounted for. For example, viewed from such perspective, it no longer seems appropriate to link memory with certain aspects of personal lives, with the personal responsibility to testify, to a commitment to fight forgetfulness. Data bases are out there in the world, so that experiences and knowledge to be remember, recalled and retold, which has by then become objectified and neutralized, can be retrieved at will, with little effort and little commitment. In the process, we repress the power of the past, live a present engulfed in a future of technological consumerism. What we retrieve when using a data bank remains always, in a sense, exterior to ourselves, there to be consumed but with no compelling force to commit ourselves to act, except in terms of new productions, to orient ourselves intentionally, aware of our roots and projected toward the future viewed as a personal project.

This becomes more complicated because, sometimes at the same time and sometimes in alternation, the technology of computers encourages an antropomorphic vision of machines. Computers are, for example, often treated as brains, that is, endowed with intelligence, capable of thought, and of using language in the same way as humans do. This is not an innocent analogy,but entails an implicit claim which rests on a more fundamental discourse about this technology, and about human thinking, language and being.

Also, when we say, that a word processor, from the user's perspective, must be understood by virtue of the role it plays in "communication", the distribution of "information", and the accumulation of "knowledge", all these terms tend to be taken for

granted. However, we must examine the ways in which they are applied for, ultimately, they have to do with our conception of being and rationality.

Advocates of the implementation of microcomputers in schools often view such introduction as an educational priority, even as an imperative and justify such claim in terms of the absolute necessity of students, and for that matter of teachers, to become computer literate if they are to suceed or even survive in a world where technological developments rampantly flourish, following a criterion of efficiency.

Although we are quite ready to agree with the importance of computer literacy we contend, against the grain, that this literacy involves more than a technical, procedural expertise or proficiency in the use of the calculating and information-process powers of the machine; moving at a slower pace than the experts, we revindicate that such a literacy must involve an understanding of the taken-for-granted background against which these technologies develop, an understanding concerning what computers do and how they fit in the fabric of society, a sense of the gains but also of the losses implied by the adventure which the technological enterprise constitutes.

For education and educators too, the capability of judging the educational implications of computers, their role in the classrooms having in view the achievement of broad educational purposes which, ultimately, have to do with norms and values concerning human beings and their world, transcend technical and procedural problems concerning the use of the technological devices, and should imply a reflection on their role and place as mediators and part of the much more complex symbolic world that makes up culture, on the marks they leave on us and our surroundings.

Thus, the first question we must raise concerns the frame within which those issues can be addressed, a frame that allows a dialog between computer experts, educational computing experts and those who are not experts, but nevertheless active members of our society, and of the educational community in particular, with a serious interest in understanding what computers are and how they fit into our lives, in general, in grasping the implications of computer use in learning processes and broader educational purposes. This question must be raised because we find that the interpretative context selected to orient our thought about the educational potential of this new technology is almost exclusively dictated by computer experts, by those who understand the operational characteristics of the new technology. To put it bluntly, the only discourse that is considered legitimate not only to deal with computers but also to speak about the insertion of computers in society at large and in schools in particular, is the one held by those who control this technology. Only them, it is implied, are entitled to speak about them, determining, at the same time, the public meanings of terms such as those mentioned above. Now, since their primary professional interests are technical and the language they use mainly derived from the technical characteristics of the machine we are supposed to speak this language when discussing computers and their use in education. As Lyotard says in *Postmodern condition*, "along with the hegemony of computers comes a certain logic, and therefore a certain set of prescripions determining which statements are accepted as "knowledge" statements" (p. 4). This means to say, that questions of legitimation of purposes become framed as if they were questions of internal legitimation of scientific statements, and are being coached in the same language. However, what we must realize is that if we are to discuss the larger questions mentioned above, we must break through this technicist vision, which determines to what aspects we must attend in computer education and what, by omission, is to be left out of focus. We must carefully uncover what is implicitly accepted by such vision, what goes unquestioned as long as we keep within

that frame, so that we can perceive what facets of the social order are thus strenghtened and, by the same token, what are those which are weakened, disregarded or even put away as faulty, on the grounds that they can no longer be integrated in the modern order.

We must realize that one cannot raise questions about technology that are relevant for its and our insertion in the world from a neutral, objective standpoint or from the perspective where technical experts would have us. Every questioning grows out of a background, a way of understanding within which we interpret and act. Computers too are anchored in a dimension of our tradition, a dimension which has undoubtedly become predominant in our society so that its assumptions are taken-for-granted and almost never questioned since they are viewed as givens. This dimension, the so-called 20th century technocratic mind set which privileges instrumental forms of rationality with its emphasis on particular styles of consciously rationalized thought and action, functions as a screen, lenses for the interpretation of man and world. We can see this vision at work in textbooks and teacher guides on computers which deal almost exclusively with technical aspects of their applications. What is not addressed is, for one thing, the influence of computers on the interplay among culture, language and thought processes. For this to be confronted we must not stay intimidated by the hubris developed by some and which leads to viewing that the solution of our most pressing problems is more tecnology, and that problems must be formulated in a abstract-theoretical decontextualized manner so that decisions can be made on the basis of explicit forms of knowledge and formalized rules. Explicit knowledge in this context is that knowledge which can be observed, treated as a fact or as a quantifiable, calculable measure. Knowledge not put in a propositional form, contextual meanings which are nevertheless very much part of every form of cultural activity are not taken into account.

As we said, we are especially interested in considering how computer programs mediate our use of language (what they select, amplify and reduce) in the communicative process. For this we must look at the notions of communication and language incorporated in the works of computer experts. To complicate matters, we must examine these concepts by means of language, natural language, that is, since language represents a way of putting in focus our shared symbolic world, a way of interpreting human beings, as they inscribe themselves in discursive practices.

When we turn our attention to the views of language and communication used by those who write software programs and, in particular educational programs, we find a certain number of assumptions that if applied to the discursive practices as they occur among persons are biased and reductive. In effect, what most of the literature of the field will reveal is an intrumentalist view of both language and communication which appear as neutral, non-problematic, taken-for-granted if proficiency is present at both ends of the communication circuit.

Our point is not that mathematical theories of information and communication are incorrect as such, but that they cannot claim to fully encompass what is involved in human communication. This point seems of particular consequences if instruction is to be based, as some advocate, on the use of software programs, that will promote, it is said, a more individualized but at the same time more interactive type of learning.

According to the dominant view, the mechanisms that are supposed to be at work in communication have to do with the transmission of a sequence of symbols from one place to another, from sender to receiver such that this propagation opens the passage for the apprehension of encoded meanings, thanks to the mediation of a common code, perceived as given and static. Still, in human communication we must not think of the code in an

objectivistic fashion, reduced to a repertoire of conventional signs preceeding all use. The code can not be viewed as an homogeneous absolute, stable and invariant during the whole of communication, independent of the conditions of its utilization by situated interlocutors. The code does not impose a totalitarian demand upon us. It is true that this "mediating third" trancends to some extent the intervining parts that make use of it, but these must enter in relation in order to activate it and even regenerate it. The code needs to be reactualized in context, given life in speech, and eventually be transformed in interlocution. At the same time, language is not constituted by pre-established, univocal signs ready to be selected, regardless of the interlocutors and their contexts. It is constituted by an ensemble of virtual conditions for discursive practices, slowly forged by an interlucotory community which is able to sustain their potentialities for the production of such practices and their production of understanding. If we ignore the dynamics of communication, the process of interlocution which puts language in motion, what then passes for communication is greatly reduced to the transmission of information, viewed as brute data, constructed be means of a fixed, univocal code, shared by sender and receiver, uniting them but somehow suspended and external to them. Interpretation on the part of the receiver is reduced to a process of decoding. This decoding is supposed to be made with the help of syntactical and other rules of application fixed since the beginning by codification. The transference of information is, of course, a way of being in relation, but corresponds to a particular poor type of relation. In this case, it is a passive sense of communication that is emphasized, since what it is privileged is the notion of information, mono-coded, measurable, decipherable. Those using the programs easily forget that an act of communication cannot be viewed apart from its generation, from what invests it with a signifying function, and from its consequences. What is concealed is that between a speaker and his interlocutor we usually have more than a transmission of information. More generally, we have a meaning effect and the establishment of a co-reference, achieved during the process of communicative interaction. The problem of the subjectivity and intersubjectivity inscribed in language as well as considerations of the institutional sites of language are kept out of focus. The use of language in computation implicitly admits that language functions at the level of meaning in the same way that at the syntactic and phonological levels, so that syntactic and discursive distinctions coincide. Such an assumption has also something to do with the approaches of semantics that are privileged. The formalistic approaches that are favoured assume that sentences in a natural language can be set into correspondence with one or more possible interpretations in a formal language for which the rules of reasoning are well defined. The study of meaning is viewed as including both the translation of ordinary language sentences into the corresponding formulas of a formal language, so that the essence of their meaning is preserved, and the establishment of logical rules that account for the interrelation of truth conditions for different formulas. The meaning of formal structures is determined on the basis of the meaning of their parts, by a system of rules which must also take into account the modes in which those parts are combined. In addition, it is implicitly acknowledged that for the compositional rules to be pertinent, the meanings of the items being composed should be fixed without reference to the context in which they appear. These theories of language are to some extent related to theories of mind which adopt some sort of a representational hypothesis by which it is assumed that thought is the manipulation of representational structures in the mind which are treated as sentences in a "internal language".

The approaches we have sketched thus far are in accordance with a view of language as a conduit for the communication of information and ideas, thoughts and feelings (see

Reddy, 1979). This view, which the manner in which microcomputers are used reinforces and which is taken-for granted, given its extreme pervasiveness, are of great educational consequence. According to this view the communication process is seen as some kind of "transportation" of a certain fixed meaning or "content" from sender to receiver, as if this transfer of information, thoughts and feelings, can be viewed independent or prior to the various "coding" processes involved at different stages. What is at stake is the way in which the connection between thought and language is construed and this the educators must understand more fully. The explicit account of this notion is usually concealed in the literature written on the uses of computers in education but can be uncovered by means of a closer look at how they present the technological capabilities of the computer, specially those linked with the tasks of receiving information, storing it, manipulating it, and providing output.

According to this conduit view of language, concepts are put into words, words are expected to carry meaning to others (meaning that is contained in the transmitted message), who then extract the concepts and information once again from words. Language is the conduit through which objective information flows. Again, this shows that instrumentalist views of language and communication emphasize the latter's passive aspect. Listening, reading and learning appear as passive tasks, since the reader, listener or learner have only to take in ideas and information conveyed by language. Moreover, the complex dynamics of the interplay between thought and language is completely erase. By the same token, the social dimension of language and knowledge, the interpretative-existential frameworks acquired throughout a life time through processes of interaction are left out of the picture.

The interconnected formulations of language we have been mentioning are consistent with cognitivists theories of thinking whichs treat thought as information processing, asserting that cognitive systems can be best understood by analogy to programmed computers. We can mention here some of the important tenets of the information processing approach:

1. Cognitive systems are symbol systems. They achieve their intelligence by symbolizing external and internal situations and events. Thinking, for example, is a composite representation based on a series of elementary or primitive process, combined serially according to explicit and predetermined rules.

2. All cognitive systems share a basic underlying set of symbol manipulating processes. In thought processes, information is manipulated in the form of discrete and contextless bits. Thus, thinking is seen as processing that is avowedly mechanistic and elementistic.

3. A theory of cognition can be couched as a program in an appropriate symbolic formalism such that the program when run in the appropriate environment will produce the observed behavior.

The emphasis on the representational character of the psychological was the information processing approach's reaction to the previously predominant behavioristic orientation in psychology which aimed at conceptualizing psychological life without reference to cognition at all. However,the "paradigmatic revolution" which this cognitivism claimed to have achieved is not a whole heartedly revolution, for it conserved some basic assumptions of traditional psychology, namely the conceptual foundation of the experimental methodology, that is, objectivism in the sense that is believed that psychology's subject matter must be treated as objects, mechanisms for causal analysis. Contemporary cognitivism, as information processing, has been able to introduce "cognitive processes" into psychology by retaining the fundamental presupposition that these processes are

mechanistic – as demonstrated by computer models. The question is whether this new paradigm is any more faithful to the essential nature of psychological life, and what light is shed on the "thinking man" when we derive form the computer model the calculative representational concept of thinking and then posit in onto the experiencing person.

Conclusions

The consideration of speech and discursive practices as interlocution gives us not only a criterion which will help us to ascertain the limits of communication with a computer but also a frame within which we want our dialogue with computer experts to develop. We must be given a voice in the process of how computers are going to be introduced in schools and for this we cannot simple be asked to appropriate and assent to an already established discourse by those who are in position to legitimate their own discourse. We need to force the conditions for a real dialogue, to revindicate to be partners in the discursive endeavor of constituting communicable meanings, so that reference is too viewed as co-reference established in a process of transaction between interlocutors in verbal interaction. We must evaluate the possibility if not of a shared code that of the confluence of mutually intelligible language games, of something that might bind us as members of a community. Mutual understanding of the issues involved does not have to lead an inevitable consensus but it also prevents the take over of the public place of discussion by experts who would then artificially promote homogeneity, leading to the supression of the heterogeneity of perspectives.

Only with the reinforcement of that public space can we discuss at a meta-level of analysis what is implied by the use of computers, having at the basis the underlying question of what it means to be human which includes, no doubt, an ethical dimension.

Let it be emphasized in the end that for us what is in question is not technology, as such, and still less rational modes of thinking, but, most especially, a certain presentation of technology by means of which the human condition is seen in technomorphic terms, a position that, we think, leads to questionable claims as well as the complementary version according to which characteristics of the machine are identified with that of humans. The use of analogic thinking which tends to be literalized must continuously be under constant surveillance. As far as information technologies are concerned, we must be aware of the types of knowledge that can be encoded by computers – mainly procedural knowledge – and therefore what, by nature, is left out of the domain of computerization, even in a computerized society.

Asher Idan

The Open Univ. and Tel Aviv Univ., Tel Aviv (Israel)

Educational Ideologies in the Informational Age

Abstract

Informational societies are societies in which the largest economic sector is the informational one and/or in which the largest number of workers is of the informational workers.

Ideologies are collective systems of ideas, values, goals and patterns of behavior, which force individuals and subgroups to be integrated in the social system.

Educational ideologies are applications of ideologies of the above kind in educational thougts and pratices, mainly in contexts of learning, teaching, management and building educational institutions, etc.

Most of the texts which deal with ideologies refer mainly to the reality of the ***industrial societies****. The central problem of this paper is: What happen to these ideologies in the* ***new reality*** *of the* ***informational societies****?*

Our central thesis is the following one: The reality of the informational society "deconstruct" the basic concepts of marxism and liberalism, since neither "freedom" or "mode of production" can be adequate concepts for this new reality. In this new reality the informational workers are the ruling group (the "elite" in liberal terminology, the "ruling class" in marxist terminology). At the same time, the economic and political power is "moving" from the West towards the East (Japan, China, etc.) and thus, new nations are moving to the center of the world system (Wallerstein, 1979). Thus, we live in a dynamic period in which new groups (in a given nation) and new nations (in the international context) are taking advantages from the "informational revolution". At this dynamic period, ***pluralism*** *will be a strong educational ideology.*

Resumo

As sociedades informacionais são sociedades em que o sector económico maior é o informacional e/ou o maior número de trabalhadores é constituído por trabalhadores da informação.

Ideologias são sistemas colectivos de idéias, de valores, de objectivos e de padrões de conduta que forçam os indivíduos e os subgrupos a integrarem-se no sistema social.

Ideologias educacionais são aplicações ideológicas da referida natureza que actuam ao nível dos pensamentos e das práticas educacionais, principalmente em contextos de aprendizagem, ensino, gestão e construção de instituições educativas, etc.

A maior parte dos textos que se referem a ideologias tratam principalmente com a realidade das ***sociedades industriais****. O problema central do nosso estudo é: O que acontece a essas ideologias no âmbito da* ***nova realidade*** *das* ***sociedades informacionais****?*

A nossa tese central é a seguinte: A realidade da sociedade informacional "desconstrói" os conceitos básicos do marxismo e do liberalismo, uma vez que "liberdade" e "modo de produção" não podem ser conceitos adequados a esta nova realidade. Na nova realidade, os trabalhadores

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 446/450
1992, APDC & SPF, Lisboa

informacionais estão a tornar-se o grupo dirigente (a "élite" na terminologia liberal, a "classe dirigente" na terminologia marxista). Ao mesmo tempo, o poder económico e político está a "mover-se" do Ocidente para o Oriente (Japão, China, etc.) e, deste modo, há novos países que se movem para o centro do sistema mundial (Wallerstein, 1979). Assim, vivemos um período dinâmico em que novos grupos (num dado país) e novos países (no contexto internacional) ganham vantagens com a "revolução informacional". Neste período dinâmico, o ***pluralismo*** *torna-se uma poderosa ideologia educacional.*

1. Introduction

The central aim of this paper is to analyse the educational ideologies of the second half of our century in the light of the educational ideologies of the last three centuries. In the following section, I explicate the concept of ideology in the light of five concrete ideologies. In section 3, I derived educational implications from the discussion in section 2. In sections 4 and 5, I tried to explore the new ideologies and the new educational ideologies which are still emerging in these days.

2. Ideology and ideologies in the Industrial Age

There are many ideologies and many definitions of the term "ideology". In my book (forthcoming, 1990) I made a critical analysis of many definitions of the term. From that analysis I developed a synthesis between Marx's and Durkheim's definitions of the term. According to what I call "the Marxian-Durkheim definition", ideology is: *a collective representation of the physical and social world, which serve as an instrument of power in conflicts between collectives.* From an epistomological point of view ideology is a paradigm, a language-game, or an epistemic "iron cage" which determines the thoughts and practices of the members of a given collective. It is determined in three ways:

A. The economic-political way (Marx): The member can learn, read and publish only what fits the interests of the dominant group in his collective.

B. The psycho-motivational way (Marx): The member is biased by his private interests to learn, read and publish only what fits the interest of his group.

C. The psycho-cognitive way (Durkheim): The member is limited in his ability to grasp and conceptualize complex systems such as nations, tribes, states, classes, etc. Only the collective conceptualization of many individuals can achieve a conceptualization of such complex entities.

The concrete ideologies which fit the above general analysis of ideology are: *socialism, fascism, liberalism, scientism,* and *theism*. All the five are collective representations which serve as instruments of power. What kind of representations each ideology has? Who are the collectives which use the power of these ideologies? In the following table I will answer these questions for all the five ideologies:

	explicit theory (cosmology)	explicit theory (sociology)	implicit messages and uses	the collective user
liberalism	atomism	the power of the collective arise from the power of its members	legitimation for exploitation in the name of the «freedom of the individual»	capitalists
fascism	organic holism	the power of members arise from the power of the nation	legitimation for the monopolization of the state's power	the state apparatus
scientism	mechanism (natural forces)	all kinds of power can be described through the method of the physical sciences	legitimation for the monopolization of the scientific method	scientific communities and universities
socialism	mechanical holism	the power of members arise from the power of the class	legitimation for the monopolization of the party's power	workers parties and unions
theism	animism (holy forces)	all kinks of power arise from the power of God	legitimation for the monopolization of the church's power	religious leaders and churches

The first four ideologies are ideologies of the *industrial age* (or capitalism), while the last ideology (theism) is the ideology of the *agricultural age* (or feudalism). Till now we have analysed the concept of ideology and the kinds of ideologies in the industrial age. Before we will check what happened in the informational age we will check the implications for the educational ideologies in the industrial age.

3. Educational ideologies in the Industrial Age

Educational ideologies are *the ideologies which are presupposed by teachers and students.* The liberal explicit educational ideology emphasizes the competition between individuals and tends to ignore systematic and collective constraints which limit the possibilities of competition. Thus, its implicit educational ideology (or what is called in educational theory "a hidden curriculum") is the need for the reproduction of the differences between the working class and the capitalists from one generation to another. The Marxist explicit educational ideology emphasizes the competition between classes and systems (modes of production). Its implicit educational ideology is the reproduction of the relations of power between the party's elite and the working class, from one generation to the next generation.

Both the liberal and the Marxist explicit educational ideologies emphasize progress and the belief in a better future. The implicit educational ideology which is common to both of them is social *evolutionism* and the belief in the supremacy of *science and of the Western culture*. Thus the educational ideologies of both tend to reproduce the existing relations between north and south, between the West and the non-West.

The two ideologies and the educational practices which are based on them are the children of the nineteenth century and the first half of the twentieth century. Most of the educational ideological disputes in our age are still going round these two ideologies (like

in the second half of the twentieth century; the reality of "the informational age", of post-colonialism, post-capitalism, post-socialism, post-liberalism and post-industrialism.

4. Ideologies in the Informational Age?

What is "informational age?" Do the central characterization of the end of the twentieth century is its technology? There are other labels for this period, such as: the post-modernic age, the neo-imperialistic age, the post-capitalistic age, etc.

Bell (1980) and Toffler (1980) think that the central characteristic of the second half of the twentieth century is its information technologies (computers, new means of communications). Critics such as Foucault (1977), Amin (1977) and Schiller (1981) think that the central characterization is *not technological* but *social*. They think that the new informational technologies are the strongest means of power that have ever been known. The new technologies have widened the global gaps between the north (U.S.A., Europe, Japan, U.S.S.R.) and the south (Africa, South America, India), and the local gaps between the rulers and the ruled. The world became not only a global communicational village (McLuhan) but also it became global industrial village (Amin, Wallerstein). In this "industrial-informational global village" arise new elites of power/knowledge (Foucault). Thus, terms such as "the computer age" are not neutral terms. They are ideological terms which cover the negative social results of the new technologies.

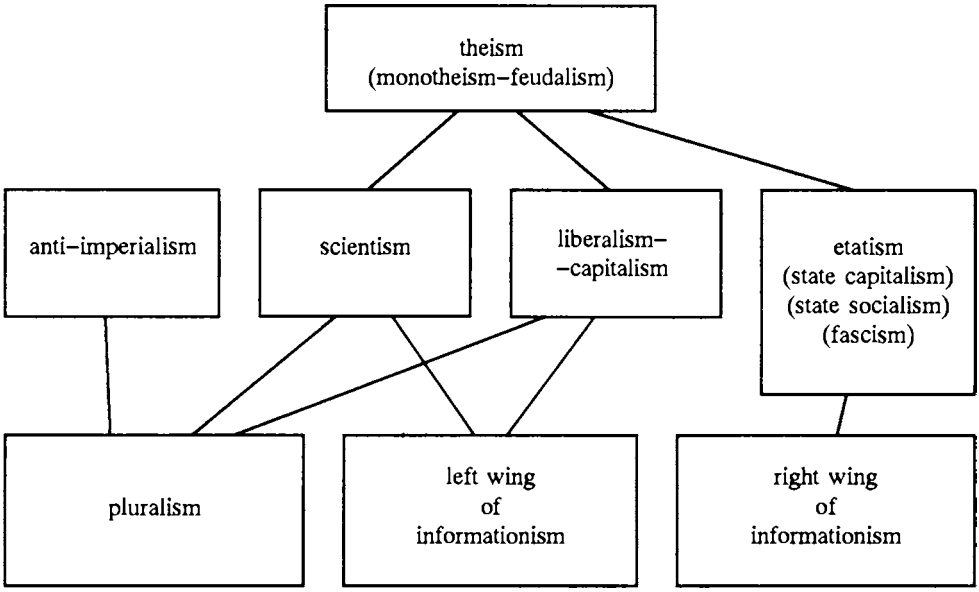
Let us call the end of the twentieth century "our age". What are the ideologies of our age?

We saw that there are two important ideologies of our age: 1) Informationism (Bell, Toffler); 2) "Postism" (post-modernism, post-imperialism, etc.).

Informationism is the ideology which on the *explicit level* (see the table at section 2) assert that the central moving force of reality is not *matter* neither *energy* but *information*. Redner (1986) sees the founders of the computer sciences (Von Neumann, Wiener) and of the genetic sciences (Watson and Crick, Monod) as the prophets of this ideology. On the *implicit level* informationism presupposes that the information workers are the most important group of society. It also presupposes that there are "informational forces" (of the computer, of the D.N.A., of the brain, of the language, etc.) which are not apparent to everyone but only to information experts. These *informational forces of informationism* are similar in some aspects to the *natural forces of scientism* (see also the table at the end of section 2). Thus scientism can be regarded as an initial stage or a special case of informationism.

The second important ideology of the second half of this century is "postism". Postism is the ideology which on the *explicit level* asserts that there is not only one moving force, but that there are many (not one scientific method (Feyerabend), or one social force (Foucault against the Marxist concept of the proletariat), or one culture (Amin and others against the pseudosuperiority of Western culture), etc. On the *implicit level* "postism" presupposes the death of ideology, its narrowness and its Europocentric biases. This is because *pluralism* is against any kind of monism (in religion, in society, in culture, in science, etc.), and monism is the presupposition of any claim to monopol (see also the table at the end of section 2).

We can summarize our discussion of ideologies in the "agricultural age", in the "industrial age" and in the "informational age" through the following chart:



5. Educational ideologies in our age

What are the educational ideologies which can be derived from the "informational ideology" and from the "postist ideology"? We already saw that the information ideology has pluralistic tendencies and centralistic tendencies. On the *explicit level*, the centralistic educational attitude of informationism tries to use the telecomputational means in order to achieve *a global school or a global class* (like the communicational global village of McLuhan). On the *implicit level* it presupposes that the global school should be governed by the information workers.On the *explicit level* the pluralistic educational ideology negate any kind of one central teaching authority, or of one central cultural authority, or of one central scientific method. It emphasizes the central role of the *local and communal committees* (parents committees, student councils, etc.). It negates modernization as a linearistic Europocentric narrow educational attitude.

References

Amin, S., 1977, *Imperialism and Unequal Development,* Monthly Review Press, New York.

Arrow, K.J., 1980, *"The Economics of Information"* in: Dertouzos and Moses (eds.) *The Computer Age.*

Bell, D., 1980, *"The Social Framework of the Information Society"* in: Dertouzos and Moses (eds.) *The Computer Age.*

Dertouzos M. and Moses J. (eds.), 1980, *The Computer Age,* M.I.T. Press, Cambridge.

Foucault, M. 1977, *Power/Knowledge,* Random House, New York.

Idan, A., 1990, *What is Ideology,* Dvier, Tel Aviv.

Redner, D., 1987, *The Ends of Science,* Westview, London.

Schiller, V.I. (eds.), 1981, *Who Knows: Information in the Age of Fortune 500* Ablex, New Jersey.

Toffler, A., 1980, *The Third Wave,* Gordon and Breach, New York.

Wallerstein, I., 1979, *The Capitalist World Economy*, Cambridge University Press, Cambridge.

3

LINGUAGENS & HERMENÊUTICAS

LANGUAGES & HERMENEUTICS

André Robinet

Centre d'Histoire des Sciences et des Doctrines
CNRS, Paris (France)

Applications de l'Informatique à l'étude des textes

Résumé

Les techniques récentes de l'informatique appliquée à la statistique lexicographique ont conduit à obtenir deux types de résultats: documentaires et interprétatifs. La question documentaire est bien connue; la question herméneutique l'est beaucoup moins.

On prendra pour base des résultats portant sur La profession de foi du Vicaire savoyard. *Les dépouillements montrent comment on peut affiner la connaissance génétique et la connaissance structurale des discours philosophiques ou littéraires ou historiques. Ce texte de Rousseau est pris comme exemple inter-disciplinaire.*

Resumo

As recentes técnicas informáticas aplicadas á estatística lexicográfica conduziram à obtenção de dois tipos de resultados: os documentais e os interpretativos. A questão documental é bem conhecida; a questão hermenêutica, menos.

Toma-se por base resultados respeitantes ao La Profession de Foi du Vicaire savoyard. *Os apuramentos mostram como se pode afinar o conhecimento genético e o conhecimento estrutural dos discursos filosóficos, literários ou históricos. Este texto de Rousseau é tomado como exemplo interdisciplinar.*

La recherche en histoire de la philosophie s'appuie aujourd'hui sur les applications de l'informatique à l'étude des textes. Dans la mesure où l'on a tenu compte de l'importance du système des signes pour toute émission de "pensée" il convient d'interroger ce système sémiotique, de considérer les discours déposés dans l'historie de la philosophie, de la littérature ou de la connaissance du passé des sociétés, comme un ensemble signifiant qui a, en lui-même, sa propre portée. On voit donc d'entrée que ces problématiques reposent sur une certaine conception de la relation entre "pensée" et "langage", et qu'elle postule que le système des signes n'est pas indifférent à l'analyse de la pensée.

Un gros progrès a été accompli dans la méthodologie de la dernière période quand on s'est mis à publier les oeuvres considérées, non plus selon le principe très controversable, du "bon texte", mais selon toute l'épaisseur du discours que la reconstitution du fait textuel permet de produire. Ainsi, pour ma part, j'ai conduit les opérations relatives aux oeuvres complètes de Malebranche, de Leibniz, de Cournot, ou d'autres, en recherchant tout ce qui pouvait nous renseigner sur la genèse de l'écriture considérée dans sa simultanéité rédactionnelle aussi bien que dans ses apparences successives. Ainsi, il n'y a pas "un texte", mais une pluie de variantes, quand la reconstitution en est possible, qui

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 453/461
1992, APDC & SPF, Lisboa

renseignent sur la composition du discours se faisant. Une oeuvre, armée de son apparat critique composé à partir des brouillons, des corrections d'épreuves, des révisions d'éditions successives, apprend comment l'oeuvre s'est comportée dans sa propre graphie. On peut, à partir de cette observation, effectuer une multitude de constats de la plus haute importance pour l'approche des architectoniques. Car si l'approche du texte en épaisseur remplace l'établissement du bon texte, la recherche architectonique supplante la redite du "système". Il fut un temps où parler d'un auteur consistait à paraphraser, avec plus ou moins de précision et d'élégance, les thèses qu'il avait énoncées et l'ordre selon lequel il en avait logiquement décousu avec la disposition de ses "pensées". Il s'agit aujourd'hui de sonder les relations langagières sur lesquelles repose l'énoncé du système, qui n'est plus considéré que comme un produit superficiel, une sorte d'épiphénomène d'un travail concret sur le langage.

Ainsi la triple considération du système des signes, au lieu de l'intuition de la pensée, des architectoniques plutôt que des systémes, des discours plutôt que du bon texte, conduit-elle à mobiliser de nouvelles techniques bien rodées par les linguistes, notamment par les lexicographes.

Les procédures dont je veux vous entretenir reposent en effet sur les tentatives qui ont été effectuées dans le domaine de la lexicologie statistique, qui nous a apporté sur le langage des lois considérables et des traitements innovants. Cette convergence de l'historien de la philosophie avec certaines procédures cybernétiques n'allait pas sans soulever de problèmes: équipement, saisie des textes, procédures d'analyse, centre de traitement. Rien de plus opposé, selon certaines traditions, que l'informatique et le philosophique, que le déferlement du "nombre" face à la solitude de la "pensée". Comment pouvions nous aborder la "décomposition du discours", comme on avait tenté, au moment de l'idéologie du début du XIXème siècle, la "décomposition de la pensée".? L'herméneutique armée devait composer ses propres méthodes. Ces méthodes reposaient sur la convergence de la philosophie et de la linguistique, par le biais de la lexicographie automatisée. Ces recherches, considérées aujourd'hui comme relevant l'intelligence artificielle, entrent donc dans un contexte épistémologique d'avant-garde, qui rejoint à la limite les questions de la "simulation" que j'avais abordées vers 1974 dans *Le Défi cybernétique*. La question de l'automate parlant-pensant fait ressurgir la mythologie du Golem, et il convenait de s'interroger sur cette renaissance.

C'est donc sur la fonction du langage que l'on se tourne dés que l'on pose la question de la composition du discours, avec pour horizon la simulation de l'automate parlant-pensant, et, de manière plus réaliste, la question de la traduction automatique. Et il est vraisemblable que si j'ai été intéressé par cette approche, c'est en raison d'une part leibnizienne très importante dans mes propres travaux, qui me mettait à même d'aborder ces problèmes dans un autre contexte qu'augustinien ou cartésien.

* * *

Pratiquement nous partons de resultats que je voudrais commenter en divisant nos procédures en procédures de documentation et procédures d'investigation.

Les procédures de documentation portent sur le discours considéré comme un assemblage de parties noires et blanches, les mots séparés par un espace dans nos langues occidentales. Cette mise en miette du discours est opérée après un travail de codage qui peut être très simple et aisé à porter sur le texte utilisé. Texte et codage sont enregistrés: les programmes utilisés permettent de jouer sur les parties noires du discours, les mots, considérés dans leur simple graphie. Cette graphie peut d'ailleurs être enregistrée avec sa diversité ordinaire, italique, majuscule, types de caractère, et même avec ses différences,

comme c'est le cas dans l'état archaïque du français ou de l'italien. Une fois ces formes saisies, on profitera de cet émiettement maximum pour observer quel matériau a servi à la constitution de l'oeuvre.

Une première procédure permet d'obtenir la table des formes qui sont présents dans le corpus considéré. Il ne s'agit plus de la manipuler de manière quelconque mais, par exemple regrouper ces formes alphabétiquement: on obtient ainsi *un vocabulaire d'auteur* qui permet d'apprécier exactement les mots qu'utilise le scripteur. Il est évident que la comparaison entre différentes oeuvres permet de dégager de premiers enseignements: notamment sur la spécificité du vocabulaire philosophique par rapport à d'autres corpus dans d'autres disciplines, et de l'individualité de chaque vocabulaire d'auteur, par comparaison entre eux des corpus philosophiques.

La constitution d'*index des occurrences* s'ensuit: il s'agit de faire accompagner chaque forme par un codage de l'endroit où elle se trouve dans l'oeuvre. Car on aura pris le soin lors de la saisie du texte de la faire accompagner d'un indice du tome considéré, de la page et de la ligne. Si bien que l'on entre en possession de ce qu'un bénédictin n'aurait jamais imaginé: la série exhaustive des formes d'un discours avec leurs emplacements successifs. Les possibilités de repérage et de comptage qui s'ensuivent sont considérables, puisque rien n'échappe à la désignation de ce large *situs*. Ne serait-ce que comme index statique on se procure ainsi un index complet et absolument rigoureux. L'automatisation de l'enregistrement et des procédures d'appoint supprime toute approximation, procure un résultat global exact et fiable, ce dont est incapable la méthode naturelle, même la plus attentive et la mieux armée mémoriellement. Voici par exemple les index des occurrences d'ORDRE dans *La Profession de foi du discours savoyard* de Rousseau, que je prends comme exemple pour cet exposé.

OCCURRENCES D'ORDRE

FAG	F1	F2	F3	FN	ITEM	OCCURRENCES:		PARTIE,	PAGE,	LIGNE
1	0	0	1	0	ORDONNAT		3.45.09			
5	2	2	1	0	ORDONNE		1.12.12	1.12.39	2.28.11	2.28.11
					3.35.26					
6	4	2	0	0	ORDONNE		1.01.18	1.12.35	1.16.33	1.21.17
					2.28.14	2.28.31				
1	1	0	0	0	ORDONNÉE		1.01.27			
3	3	3	0	0	ORDONNÉS		1.11.08	1.12.05	1.12.15	
41	21	10	9	1	ORDRE		1.01.25	1.10.27	1.10.42	1.10.43
					1.11.10	1.11.50	1.12.37	1.13.12	1.13.34	1.13.45
					1.14.12	1.16.43	1.17.40	1.18.07	1.18.08	1.18.43
					1.19.14	1.19.37	1.20.07	1.21.37	1.21.38	2.28.04
					2.28.08	2.28.09	2.28.29	2.28.30	2.28.46	2.29.04
					2.30.04	2.30.11	2.30.13	3.32.30	3.32.36	3.34.45
					3.35.20	3.35.44	3.38.17	3.38.21	3.38.30	3.48.41
					N.26.54					

Tout le lexique étant ainsi décomposé, un certain nombre d'observations peuvent être dégagées, concernant par exemple *les tables de fréquence*. On peut se demander, pour un discours considéré, quelle est la quantité des formes employées, au total et dans le détail, par parties ou individuellement. On obtient alors une sortie qui classe les formes selon leur plus ou moins grand emploi dans le texte. On aura évidemment de très basses fréquences, mots peu utilisés, de moyennes fréquences, et de hautes fréquences. Les lexicographes ont tiré un certain nombre de lois de ces fréquences, dont nous observons la validité pour le

discours philosophique. Ces tables de fréquence ont une importance: mais il importe de les commenter avec précision. Il est manifeste que des mots sont très employés par certains philosophes, qui ne le sont pas par d'autres. On pourra donc établir une comparative des fréquences, effectuer des rapprochements entre des utilisations similaires ou des distinctions entre des usages dissemblables. D'autre part, dans l'oeuvre d'un auteur, on pourra observer l'apparition de formes nouvelles et la désuétude de formes auparavant utilisées. Bref un criblage du lexique d'auteur selon les fréquences conduit à de nouveaux résultats sur les parentés lexicales ou les étrangetés observées. Ainsi, on observe pour le *Vicaire savoyard*, une familiarité du lexique de Rousseau dans ce passage de son oeuvre, avec les tables de fréquence des auteurs philosophes comme Malebranche ou Bernard Lamy: le vocabulaire de l'ordre notamment, en témoigne. On peut alors évidemment se poser la question de lectures et d'influences sur laquelle l'histoire donne une réponse. Par exemple, COEUR intervient moyennement dans le *Vicaire*: c'est plutôt alors à un apparentement avec les fréquences pascaliennes que l'on peut la comparer.

TABLE DE FREQUENCE DU VICAIRE

ITEM	FREQ. GLOBALE	Ventilation	1	2	3	Notes
DIEU	107		25	5	71	6
HOMME	97		36	19	35	7
HOMMES	69		8	6	45	10
NATURE	59		30	18	8	3
NATURES	2		2	0	0	0
ORDRE	41	etc...				

S'il est bien évident qu'à ce niveau, il ne faut manipuler les observations qu'avec la plus grand précaution, c'est parce que l'émiettement des formes les a rendues purement sémiotiques, sans aucune relation au sémantique. Force est d'en revenir à des procédures plus globalisantes, qui vont réinvestir les formes dans leur contexte, non pas qu'il faille passer au sémantique, mais qu'il est nécessaire de constituer une sémiotique de proximité: ce que l'on amorce avec les *tables de concordance*. Cette fois on prend l'un des termes du lexique comme pôle et on s'interroge sur l'environnement de ce pôle. On fait sortir le contexte du mot considéré, avec l'amplitude que l'on veut: deux ou trois mots, un paragraphe, une page etc... Voici un type de sortie très opérationnel et maniable: la sortie d'*une concordance linéaire*, accompagnée de son site. On peut obtenir différent type de sortie. Les concordances successives ne font que reprendre l'ordre d'apparition des formes dans le discours: elles procurent donc ces formes dans leur contexte selon leur ordre d'apparition dans le texte. Sous cette apparence, elles sont purement documentaires et n'apportent pas d'autres enseignements. Or on peut également classer les concordances chronologiquement; par ordre d'apparition dans l'ensemble d'une oeuvre. Mais on peut aussi, et c'est là le plus payant pour l'analyse philosophique, demander une sortie telle que les signes qui succèdent à la forme (ou qui la précèdent) soient classés alphabétiquement aussi bien horizontalement que verticalement. On assiste alors au regroupement de tout ce qui est disposé autour du terme considéré. Par exemple, avec le terme ORDRE dans le *Vicaire*, on obtiendra des paquets de syntagmes qui, par eux-mêmes, font problème: "ORDRE DE L'UNIVERS, DE LA NATURE, DES CHOSES, DES ETRES etc... Dans les discours philosophiques, on observe la présence de très lourds syntagmes qui occupent

toute la ligne et qui expriment le caractère logique et architectonique répétitif du discours: dans nos dépouillements de Malebranche ORDRE DE LA NATURE ET DE LA GRACE etc...

CONCORDANCE DU MOT ORDRE

TOUT NE FINIT PAS POUR NOUS AVEC LA VIE, TOUT RENTRE DANS L	ORDRE	A LA MORT-. J AURAIS A LA VERITE L EMBARRAS DE ME 1.18.43
LES ETRES ET LEURS RAPPORTS, JE PUIS- SENTIR CE QUE C EST OU	ORDRE,	BEAUTE, VERTY, JE PUIS- CONTEMPLER L UNIVERS, M ELEVER A 1.13.34
QUANT AU CULTE EXTERIEUR, S IL DOIT ETRE UNIFORME POUR LE BON	ORDRE,	C EST PUREMENT UNE AFFAIRE DE POLICE, IL NE FAUT POINT 3.32.36
DE SES SEMBLABLES-, ET LA BONTE DE <DIEU> EST L AMOUR DE L	ORDRE:	CAR C EST PAR L ORDRE OU IL MAINTIENT CE QUI EXISTE ET 1.21.37
AGENT COMMUN QUI LES DIRIGE. LOIN DE POUVOIR IMAGINER AUCUN	ORDRE	DANS LE CONCOURS FORTUIT DES ELEMENTS, JE N EN PUIS- PAS 1.10.27
OBJET ET DE NOUS MEMES QUI EN SOMMES AFFECTES: IL-N--A QUE L	ORDRE	OF CETTE AFFECTION QUI DETERMINE LE NOM QUI LUI CONVIENT. N.26.54
JE M EXERCE AUX SUBLIMES CONTEMPLATIONS. JE MEDITE SUR L	ORDRE	DE L UNIVERS, NON POUR L'EXPLIQUER PAR DE VAINS SYSTEMES, 2.30.04
JE SAIS PAR MON EXPERIENCE OU ELLE S OBSTINE A SUIVRE L	ORDRE	DE LA NATURE, CONTRE TOUTES LES LOIS DES HOMMES. ON A BEAU 1.01.25
JE NE L ENTENDS PAS. <L INSPIRE->. DES CHANGEMENTS DANS L	ORDRE	DE LA NATURE, DES PROPHETIES, DES MIRACLES, DES PRODIGES 3.38.21
ET CE QUI N EST QUE FIGURE- ORATOIRE; QUELS FAITS-SONT DANS L	ORDRE	DE LA NATURE, ET QUELS AUTRES FAITS- N Y SONT PAS, POUR 3.34.45
JE REVIENS A MOI, ET JE CHERCHE QUEL RANG J OCCUPE DANS L	ORDRE	DES CHOSES OU ELLE GOUVERNE, ET QUE JE PUIS- EXAMINER. JE 1.13.12
QUE MON ESPECE, ET SI J'AVAIS A CHOISIR MA PLACE- DANS L	ORDRE	DES ETRES, QUE POURRAIS JE CHOISIR DE PLUS QUE D ETRE 1.13.45
UN SI GRAND INTERET A LA FORME- DE L'HABIT DU PRETRE, A L	ORDRE	DES MOTS OU IL PRONONCE, AUX GESTES QU IL FAIT A L AUTEL, 3.32.30
QUI TOMBE, DANS LA FEUILLE OU EMPORTE LE VENT. JE JUGE DE L	ORDRE	DU MONDE, QUOIQUE J EN IGNORE LA FIN-, PARCE-QUE POUR 1.10.42
LA DERNIERE EVIDENCE. ELLE NE S EST PAS CONTENTEE D ETABLIR L	ORDRE	ELLE A PRIS DES MESURES- CERTAINES POUR QUE RIEN NE PUT 1.1.50
A L ORDRE OU IL ETABLIT, SUR DE JOUIR MOI MEME UN JOUR DE CET	ORDRE,	ET D Y TROUVER MA FELICITE: CAR QUELLE FELICITE PLUS 2.28.30
A LA MEME FIN-, SAVOIR LA CONSERVATION OU TOUT- DANS L	ORDRE	ETABLI. CET < ETRE- > QUI VEUT ET QUI PEUT, CET < ETRE- > 1.12.37
MIRACLES EN MA FAVEURE MOI QUI DOIS AIMER PAR DESSUS TOUT L	ORDRE	ETABLI PAR SA SAGESSE ET MAINTENU PAR SA PROVIDENCE, 2.30.11
MEME DE <MOISE> FAIRE LES MEMES SIGNES- OU IL FAISAIT PAR L	ORDRE	EXPRES DE < DIEU >, POURQUOI DANS SON ABSENCE, N EUSSENT 3.35.44
VERITES ETERNELLES DONT IL EST LA SOURCE, QUAND LA BEAUTE DE L	ORDRE	FRAPPERA TOUTES LES PUISSANCES DE NOTRE AME, ET QUE NOUS 1.19.37
SA SAGESSE ET MAINTENU PAR SA PROVIDENCE, VOUDRAIS JE QUE CET	ORDRE	FUT TROUBLE POUR MOIE NON, CE VOEU TEMERAIRE MERITERAIT O 2.30.13
QUE L ABUS DE LA LIBERTE OU ELLE LUI LAISSE NE PEUT TROUBLER L	ORDRE	GENERAL. LE MAL- QUE L HOMME FAIT RETOMBE SUR LUI SANS 1.16.43
A RAPPORTER TOUT A LUI, ET LUI DONNE UN INTERET CONTRAIRE A L	ORDRE	GENERAL. OU ELLE EST POURTANT CAPABLE DE VOIR ET D AIMER; 2.29.04
MONDE, QUOIQUE J EN IGNORE LA FIN-, PARCE-QUE POUR JUGER DE CET	ORDRE	IL ME SUFFIT DE COMPARER LES PARTIES- ENTRE ELLES, D 1.10.43
FANATIQUES PERSECUTES IL-N-Y-EUT POINT DE MIRACLES. C EST L	ORDRE	INALTERABLE DE LA NATURE QUI MONTRE LE MIEUX LA SAGE MAIN 3.35.20
PEUT ON LUI DONNERE LA VERTU, DISENT ILS, EST L AMOUR DE L	ORDRE	MAIS CET AMOUR PEUT IL DONC ET DOIT IL L EMPORTER EN MOI 2.28.04
AMOUR DE L ORDRE, PRIS DANS UN SENS- DIFFERENT. IL-Y-A QUELQUE	ORDRE	MORAL PARTOUT OU IL-Y-A SENTIMENT ET INTELLIGENCE. LA 2.28.09
DE L ESSENCE DIVINE. JE NE FAIS QUE SUPPOSER LES LOIS DE L	ORDRE	OBSERVEES ET < DIEU > CONSTANT A LUI MEME. NE ME 1.20.07
DE MOTS: CAR JE DIS AUSSI, MOI, QUE LE VICE EST L AMOUR DE L	ORDRE	PRIS DANS UN SENS- DIFFERENT. IL-Y-A QUELQUE ORDRE MORAL 2.28.08
MAL SOI MEME. EN ATTENDANT DE PLUS GRANDES LUMIERES, GARDONS L	ORDRE	PUBLIC; DANS TOUS PAYS RESPECTONS LES LOIS, NE TROUBLONS 3.48.41
SIENNES, ET PAR LE BON USAGE DE MA LIBERTE ; J ACQUISCE A L	ORDRE	OU IL ETABLIT, SUR DE JOUIR MOI MEME UN JOUR DE CET ORDRE, 2.28.29
LA BONTE DE <DIEU> EST L AMOUR DE L'ORDRE; CAR C EST PAR L	ORDRE	OU IL MAINTIENT CE QUI EXISTE ET LIE CHAQUE PARTIE- AVEC 1.21.38
RESTE LIBRE ET PUR, QUEL MERITE- AURAIT IL D AIMER ET SUIVRE L	ORDRE	OU IL VERRAIT ETABLI ET OU IL N AURAIT NUL INTERET A 2.28.46
QUI LES REMPLISSENT, QUE DEVIENS JEC QUEL SPECTACLE. OUU EST L	ORDRE	QUE J AVAIS OBSERVEE LE TABLEAU DE LA NATURE NE M OFFRAIT 1.14.12
L ORDRE QUI LE PRODUIT S APPELLE <BONTE>, ET L AMOUR DE L	ORDRE	QUE LE CONSERVE S APPELLE <JUSTICE>. <DIEU>, DIT ON, 1.18.08
JUSTE: AUTREMENT IL SE CONTREDIRAIT LUI MEME; CAR L AMOUR DE L	ORDRE	QUE LE PRODUIT S APPELLE <BONTE>, ET L AMOUR DE L 1.18.07
QUE DANS LE DESORDRE, ET JE VOIS DANS LE SYSTEME DU MONDE UN	ORDRE	QUI NE SE DEMENT POINT. LE MAL- PARTICULIER N EST QUE 1.17.40
JE CROIS QUE L AME SURVIT AU CORPS ASSEZ POUR LE MAINTIEN DE L	ORDRE	, QUI SAIT SI C EST ASSEZ POUR DURER TOUJOURSE TOUTEFOIS 1.19.14
PEUT SE REFUSER A SON TEMOIGNAGE, A QUELS YEUX NON PREVENUS L	ORDRE	SENSIBLE DE L UNIVERS N ANNONCE T IL PAS UNE SUPREME 1.11.10
LE RAISONNEUR- > . LE TEMOIGNAGE DES PEUPLES EST IL D UN	ORDRE	SURNATURELE <L INSPIRE-> . NON; MAIS QUAND IL EST 3.38.30
DIFFERENCE. MES PREUVES SONT SANS REPLIQUE;; ELLES SONT D UN	ORDRE	SURNATUREL. <LE RAISONNEUR->. SURNATUREL. QUE SIGNIFIE 3.38.17

_ MOT < ORDRE > COMPORTE 41 OCCURRENCES

On peut dire qu'avec cette procédure on remédie à l'émiettemment des sorties précédentes, et que l'on procure des observations intéressantes pour l'analyse. En effet ces paquets de syntagmes ont des répétitions plus ou moins fréquentes, varient dans leur composition interne, font jouer des termes vicariants, procurent des remarques sur le bougé des constellations lexicographiques ainsi rendues évidentes. Comment cela évolue-t-il? Telle est une première question que concerne directment l'architectonique, l'auteur ne s'étant jamais rendu compte du fonctionnement de son propre langage ni de ses propres aberrations.

Disons enfin quelques mots d'une autre sortie, dite pyramidale, par laquelle nous obtenons un nouvel éclatement reclassé des termes qui entourent le mot pôle. En voici l'exemple (arbitraire et pouvant être tout autre) avec NATURE et ORDRE dans le *Vicaire*. On définit les mots présents dans le discours dix colonnes avant le pôle et dix colonnes après. Le principe du classement fait intervenir deux facteurs. Le premier concerne la plus ou moins grande proximité du mot-pôle le signe est alors repéré à l'éloignement **-1, -2,**

-3 etc... ou **+1, +2, +3** etc... Ce mot peut exister dans cette colonne un nombre variable de fois, définissant ainsi une fréquence à tel éloignement du pôle. Le classement automatique s'effectue en mettant au sommet de la pyramide les mots qui environnent le pôle avec la plus haute fréquence pour la plus grande proximité. L'avantage de cette disposition consiste à donner une spectrographie de l'environnement d'un terme considéré, jusque dans ses plus grands éloignements. Mais il confirme surtout par le détail numérique ce que la concordance fait voir plus perceptiblement que numériquement. On retrouve ainsi, décomposés et classés, les paquets de syntagmes dont on peut alors entreprendre l'étude du détail.

LE POLE CHOISI NATURE APPARAIT 59 FOIS

CLASSEMENT DES CO-OCCURRENTS EN FONCTION DE LEUR DISTANCE ET DE LEUR CO-FREQUENCE LOCALE

0	0	0	0	0	0	0	0	1	3	ORDRE	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	FAIT	3	0	1	0	0	0	0	0	0	0
0	0	2	0	2	1	0	0	2	0	NORME	2	0	0	2	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	2	VOIX	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	1	TOUTE	1	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	INDIVIDUS	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	PRINCIPE	2	0	0	0	0	0	0	0	0	0
0	0	1	0	1	0	0	1	1	0	DIEU	1	0	0	0	0	0	0	0	0	0
0	1	0	0	2	0	0	0	1	1	JAMAIS	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	1	0	0	1	1	AME	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	LOIS	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1	EXPLIQUER	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	FAIRE	1	0	0	0	0	0	0	0	0	0

Telles sont quelques-unes des procédures par lesquelles on soumet le texte à une décomposition électronique qui équivaut à une observation au microscope de ses élements composants, atomiques et moléculaires. Jusque-là, il ne s'agit guère que de données lexicographiques raffinées. Le philosophe à la recherche d'une interprétation du texte, en quête de sémantique commentée, ne saurait s'en satisfaire. Aussi convient-il d'entrer dans la recherche de procédures interprétatives qui donneront satisfaction au déploiement herméneutique armé.

J'en donnerai trois exemples tirés de cette étude du *Vicaire savoyard*.

Une première investigation rejoint le fond même de l'étude conceptuelle, en s'appuyant uniquement sur les résultats obtenus par la distanciation précedement créée, sans tenir compte d'aucune donnée interprétative externe. Soit les termes métaphysiques présents dans le corpus du *Vicaire* et leur fréquence générale observée selon les trois parties traditonnelles de l'oeuvre:

	FREQ. GEN.	1	2	3	N
NATURE	59	30	18	8	3
ETRE-	58	41	7	4	6
ETRES	18	17	0	0	1
TOUT-	16	11	2	3	0
MONDE	33	19	3	9	2
UNIVERS	19	16	1	2	0
TERRE	30	10	5	13	2

Rien que la table des fréquences montre que les emplacements sont nettement distribués, essentiellement dans la première partie, celle où il s'agit de ''savoir'', avant "d'agir'' et de ''croire''. Si l'on se réfère en même tempos à la table des cooccurrences ou à notre tableau pyramidal, on peut tenter de discriminer ces vocables métaphysiques d'ordinaire confondus, et d'amorcer une réponse à l'observation portant sur l'apparition des syntagmes autour de ORDRE DE...DU MONDE, DE LA NATURE, DE L'UNIVERS. La méthode consiste à analyser chaque emplis de ces syntagmes, en s'appuyant sur les voisinages de termes environnants. Or on s'aperçoit que loin d'être simplement vicariants, ces syntagmes que l'on pourrait estimer seulement supplétifs et synonymes, comportent variations sémiotiques assurées qui les dotent d'une sémantique diversifiée.

ORDRE DES ETRES ou ORDRE DES CHOSES n'est qu'un collectif pluralisé désigne ce qu'il y a de plus indéterminé dans l'individuation de chaque être. La considération de l'ETRE s'accompagne d'une réflexion sur son RANG, sa PLACE, son RAPPORT aux autres ETRES.

ORDRE DU MONDE vise la multitude des parties d'un TOUT, ainsi que la relation qui unit ces parties entre elles, leur CONCOURS, leur CONCERT, leur HARMONIE et MONDE vise un système juxtapositionnel intégral dont les MAXIMES traduisent l'INSTITUTION DE LA NATURE. MONDE n'a de positif que sa composition élémentaire, ORDRE DE L'UNIVERS fait apparaître un vecteur de sens situé en position de valeur face aux réalités partielles, élémentaires ou conflictuelles qui composent le MONDE. L'essentiel pour UNIVERS est d'appeler la considération des LOIS et de l'HARMONIE face au chaos. Mais pas plus que MONDE, UNIVERS ne suscite l'ORDRE qui lui est affecté: il le reçoit d'une suprême intelligence. UNIVERS insiste donc plus que MONDE sur la nécessité de l'ORDRE, mais n'en a pas, n'en est pas la clé. ORDRE DE L'UNIVERS exclut toute explication par les fins, se réfère à un système VRAI, et entraîne une ouverture vers l'auteur de cet ORDRE.

ORDRE DE LA NATURE comporte 3 syntagmes binaires et un ternaire ORDRE INALTÉRABLE DE LA NATURE. C'est vers NATURE que se reportent les observations concernant l'animation de la NATURE et son ordonnancement. Aussi une puissante antithèse dresse-t-elle cet ORDRE DE LA NATURE contre la LOI DES HOMMES, contre les faits allégués hors d'ordre (miracles, prophéties, exceptions etc...) Cet ORDRE DE LA NATURE totalise ce que les précédents syntagmes détaillaient et en produit la raison existentielle et classificatrice. ORDRE DE LA NATURE implique l'intelligence de l'Etre suprême: c'est effectivement cet ordre qui sera transmis par la VOIX qui PERMET et qui PRESCRIT, qui est essentielle à tout ce qui est.

Ainsi les collectifs métaphysiques trouvent leur sens singulier que la simple lecture aura tendance à sous-estimer en la globalisant. Il y a un fonctionnement interne au langage philosophique de Rousseau que cette analyse met en évidence. On peut donc dire que *l'analyse sémantique reçoit un nouvel éclairage du simple décryptement sémiotique constitué pendant cette phase méthodique de distanciation.*

Un second résultat concerne le dépistage des métaphores dans le langage naturel. On sait avec quelle concision et quelles ellipses fonctionne le langage courant, qu'il soit littéraire, historique ou philosophique. La question est de savoir si ces procédures peuvent saisir les intentions de la métaphore, en l'absence de sa complétude. Il est bien évident que, pour que nous estimions qu'il y a métaphore, même quand l'un des éléments est défaillant, il nous faut au moins une amorce de proportionnalité pour pouvoir la repérer et la comprendre. Si l'édifice métaphorique est présent dans le discours sémiotiquement par l'un de ses éléments, cette amorce est nécessaire, mais elle est suffisante pour tirer du langage

les autres éléments de sa complétude. L'exemple est manifeste avec cette VOIX DE LA NATURE. Nous sommes là en pleine métaphore et cette métaphore est longuement filée et variée dans la troisième partie du *Vicaire.*

NATURE entre alors dans un champ métaphorique fortement aimanté en fonction du langage humain pour aboutir au syntagme LIVRE DE LA NATURE. C'est le seul LIVRE qui subsiste, tous les autres étant discrédités. Cette opposition NATURE-CULTURE constitue, dans la troisième partie du *Vicaire*, le refuge des idées les plus radicales des deux premiers discours. Des 33 occurrences de LIVRE(S) 30 figurent dans cette dernière partie, dont 18 occurrences transitent en une page et demie! Cette rafale s'ouvre avec la question: "voulez-vous vous instruire dans les LIVRES?''. Mais laissons de côté la partie négative du discours pour examiner comment on doit aborder le LIVRE DE LA NATURE.

Si la nature du livre, c'est d'abolir le LIVRE DE LA NATURE, ce dernier aura pour effet de supplanter tous les livres. La LANGUE DE LA NATURE qui y est inscrite nous parvient par deux voies que la métaphore laisse confondues si on ne passe pas à cette analyse microscopique.

En effet le LIVRE est donné à l'oeil, la NATURE est un LIVRE ouvert à tous le yeux. La direction externe du regard fait graviter la métaphore dans le sens visuel; selon un système de lecture réciproque, je lis dans le LIVRE, Dieu lit en moi.

Or ce n'est pas seulement sur l'arc sensoriel visuel que se déploie la métaphore: elle est également canalisée dans le conduit auditif. C'est un LIVRE parlant autant qu'à lire, qui parle à tous les hommes, une VOIX de la nature. Si l'on veut alors, en méthodologie complétive, passer à l'examen des occurrences de VOIX, on s'aperçoit que les 17 occurrences soulignent cette sensorialité, non plus de la lecture visuelle, mais de l'écoute auditive. La prolation renforce la visualisation. Et déferlent les adjectifis qui authentifient cette VOIX comme SAINTE, IMMORTELLE, CELESTE, INNOCENTE, et qui fait l'EXCELLENCE DE LA NATURE DE L'HOMME. La boucle se referme ainsi puisque l'étude de la VOIX DE LA NATURE aboutit à fonder la valeur de la NATURE DE L'HOMME.

On peut même raffiner en remarquant que la partie 3 du *Vicaire* se différence de la partie 2 par ce passage du DIRE au LIRE, de la VOIX au SPECTACLE, de la PAROLE au LIVRE. Cette architectonique lexicographique soutend la différence entre le plan de la conscience morale, aux prises avec un sentiment d'écoute, et de la conscience religieuse, aux prises avec les injonctions lisibles. Ainsi non seulement la métaphore reste dépistable, mais encore on découvre, ce faisant, ce que je n'ai jamais rencontré dans aucun commentaire de Rousseau: cette double coenésthésie du visible et de l'audible dans la révélation du LIVRE.

Soit une autre sortie spectaculaire, qui permet d'illustrer la découverte dans le domaine des répercussions lexicographiques sur la paléographie. L'analyse des positions d'ORDRE dans le Vicaire fait apparaître six zones de rafale où, sur les 41 occurrences du mot, on en trouve 38 regroupées sur trois petits espaces. Le tableau suivant est référé au découpage en deux parties opéré par Rousseau, aux trois parties également proposées par ses propres commentaires, et à la distribution des occurrences d'ORDRE dans le texte.

Découpage *Rousseau*	Découpage VICAIRE 76	Occurrences d'ORDRE	(41)	Zone *définie*	Origine *paléographique*
1.V.1.01	1.V.1.01	V.1.01			(F) pour l'Intr.
(Pl. 566)	(566)	(566)	(1)	Z^1	(P)
		V. 10.27			
		(570)	(20)	Z^2	(F) via *Cahiers*
		V. 22.09			
		(594)			
	2.V.22.10	V. 22.10			
	(594)	(594)	(1)	Z^3	(Z) via *Lettres* 5 e 6 à S.
		V. 28.04			
		(602)	(10)	Z^4	(F)
2.V.30.40	3.V.30.40	V. 30.40			
(Pl. 606)	(606)	(606)	(0)	Z^5	(F) transition
		V. 32.30			(P) pour le début
		(608)	(8)	Z^6	(F) pour le dialogue, via x
		V. 38.30			
		(616)	(1)	Z^7	(F)
fin	fin	V. 51.21			

Dans un zone 1 d'une dizaine de pages, on ne trouve qu'une seule occurrence. Une zone 2 d'une dizaine de pages également comporte 20 occurrences. Puis vient un vide de 8 pages avec une seule occurrence. Une zone 4 comporte en rafale dix occurrences sur deux pages. Puis suivent deux pages sans aucune occurrence. Huit occurrences définissent une zone 6 comprise sur six pages. Enfin la zone 7 terminale sur 13 pages ne voit apparaître qu'une seule occurrence.

Or si l'on réfère ces trois zones de rafale à l'origine paléographique du discours, on s'aperçoit du résultat suivant. Les six sources manuscrites qui façonnent la version terminale du *Vicaire* se comportent de la manière suivante 1º Les trois dernières interventions de Rousseau (édition princeps Duchesne, addition marginale de Rousseau sur son édition, copie Moultou) n'apportent plus aucune occurrence d'ORDRE. Le manuscrit de Genève, qui vient en troisième rédaction en ajoute une. Le manuscrit de l'Assemblée Nationale de Paris en fait intervenir 10. Quant au manuscrit Fabre, origine de cette pièce, il en comporte immédiatement 30. Or l'occurrence observable en zone 1 vient de Fabre cette zone étant surtout récupérée à partir des *Lettres à Sophie,* qui ne comportent pas de mention d'ORDRE. La zone 2 est rédigée pour sa première moitié dans le manuscrit Fabre et pour la seconde dans le manuscrit de Paris. La zone qui ne comporte qu'une occurrence d'ORDRE est en provenance des *Lettres V et VI à Sophie*. La zone 4 est très intéressante car elle provient du manuscrit Fabre lequel comporte à cet endroit deux occurrences supplémentaires d'ORDRE qui ne passent pas dans le *Vicaire*; de plus ces occurrences sont retravaillées dans le manuscrit de Paris et s'enrichissent qualitativement. La zone 6 importe ses huit occurrences à partir du manuscrit de Paris où elles apparaissent sans avoir été dans Fabre. Enfin la zone 7 est fonction de *La Nouvelle Héloise*, VI, XI.

Il semble ainsi que les apports sur les plans sémiotiques, conceptuels et sur les diverses applications que nous avons mentionnées, valent la peine d'appliquer avec précision les faits construits para l'analyse lexicographique automatisée.

Ephraim Nissan

Dep. of Mathematics & Computer Science,
Ben Gurion Univ. of the Negev, Beer-Sheva (Israel)

Derivational knowledge and the common sense of coping with the incompleteness of lexical knowledge

Abstract

Thanks to word-formation skills, humans are able to cope with words they never faced with before. ONOMATURGE is an expert system that embodies skills in word-formation. Beside proposing candidate neologisms, it also evaluates how clear they would seem to native or fluent speakers of the language considered. In this paper, we expose the definition of the basic attributes that, in the lexical database of ONOMATURGE, embody information of the diffusion of terms and acceptations, and that, once retrieved, are employed by the scoring functions of ONOMATURGE.

Resumo

Graças a aptidões de formação de palavras, os seres humanos são capazes de lidar com palavras que nunca encontraram antes. ONOMATURGE é um sistema perito que incorpora aptidões de formação de palavras. Além de propôr candidaturas de neologismos, avalia a sua clareza para nativos ou falantes, com fluência, da língua em causa. Expômos neste estudo a definição dos atributos básicos que, na base de dados lexical de ONOMATURGE, incorporam a informação de difusão de termos e de aceitações e que, uma vez recuperados, são usados nas funções de avaliação.

1. The context

Humans are able to cope with words they never met before, owing to skills in analyzing word-formation according to kown lexical bases and morphological derivational rules. On derivational morphology, see, e.g., Aronoff (1976). Nir (1981) has conducted statistical studies of neologism acceptability. Modeling computationally the estimation of how *clear* new terms are, could make NLP more robust at the lexical level. Recovery in general, in NLP, has been discussed by Carbonell and Hayes (1984). KLAUS, developed by Haas and Hendrix (1984), classifies a new term in a conceptual taxonomy, by asking the user questions, but it does not handle word-formation. ONOMATURGE is an expert system that generates candidate Hebrew neologisms (Nissan 1987 a), and scores them according to their estimated clarity. ONOMATURGE embodies lexical and morphological knowledge of a certain language (in the present implementation: Hebrew), and is intended

. to *generate* a set of *candidate neologisms* able to convey the input concept,

. to *evaluate* how well would each coined candidate neologism result *clear* and *unambiguous* for a native or fluent speaker: how easy it would be for him or her to guess the intended meaning of the proposed new term. Such an estimate is based on *morphological, lexical, semantic* and *sociolinguistc* information that ONOMATURGE

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 462/477
1992, APDC & SPF, Lisboa

• *retrieves* from its database (and with whose definition we are concerned in this paper), and
• *composes* to account for the way the new term being coined is formed.

Lexical innovation as being the application of the program is reflected in the name, *ONOMARTURGE*, of the system: it was derived from a Greek term, 'ονοματουργος', used by Plato (*Cratylus, 388e7–389a3)*, to indicate a *word inventor.*

In the knowledge-base of ONOMATURGE, *lexical entries* (terms or lexical roots) and, possibly, *formation rules* as well, have deeply nested relations associated, used as passive frames. Polysemy, near-synonymy, etc., are described there; please observe that not only LEs, but FRs as well (e.g., suffixes), are often capable of conveying different meanings. Unlike inflection, that is rarely ambiguous, deriving new words introduces ambiguity, owing to the polysemy of the derivation rule as well as of the base. We discuss numeric scores, found in frames, according to which a scoring component in the control of ONOMATURGE estimates the clarity (morphosemantic transparency) of terms it coins. The approach has been ad hoc so far, but fuzziness (through a single mark) or ultrafuzziness (through a numeric range) – cf. (Zadeh 1984) – are suited, too, and contribute in formal organicity. We claim that the approach can be extended from the socio-linguistic dimension of NLP, to semantics – cf. (Nissan 1987b) – and to a "socio-epistemic" dimension, concerning the diffusion of concepts rather than of terms, and potentially useful to tailor automatically generated text to the expected audience.

2. Word-Formation

The current implementation of ONOMATURGE has focused on Hebrew, but versions can be developed for other languages as well, with potential benefits for those languages as well as for the science and technology of artificial intelligence, as explained in the following. In general, knowledge involved in *word-formation* is often much more complex than the kind resorted to when forming Latinate international ('*Pan-European*') terms out of *morphemes*: both *affixes* (i.e., *prefixes* and *suffixes*) and *lexical roots*. Indeed, morphemes of Latin or Greek origin as used to form international technical terminology, are most often used there almost unambiguously. Such an assumption is no longer realistic when we consider the formation of terms in daily use. Indeed, first of all, terms usual outside strictly technical speech or text, and morphemes involved in their formation, tend to be ambiguous, *polysemous*: the most frequent terms are also those most ambiguous, as they can convey any out of several wordsenses (*acceptations*); people who, even subconsciously, are in the process of trying to understand a term, select the correct acceptation according to the context. That relation between *frequency* in use and *ambiguity* constitutes a phenomenon, in the economy of conveying information, that has been described in what is known by the name of *Zipf Law* (Zipf 1949).

While it is desirable that neologisms be as unambiguous (*morpho-semantically*) as possible, linguistic resources employed in the process of coining them tend to be either ambiguous, or in rare use. *Lexical roots* are even more ambiguous than terms are, as each root corresponds to several derivatives. Affixes, too, are often ambiguous. Cues for understanding, a listener or reader is offered by a term he or she never met with before, are just roots and affixes, and ambiguity is involved

. in finding out the uninflected form (that is, the form listed in dictionaries) of an inflected word,
. in decomposing the uninflected term into its elements of *derivation* (roots and affixes),
. in figuring out the correct meaning of roots and affixes as involved in the formation of the terms considered, and in finding a plausible composition of those meanings as corresponding to the concept that the term is intended to convey.

Even illiterate persons have enough knowledge in grammar and semantic concepts to allow

therm to cope with their own language even when a new term is met with, provided that the term was formed with elements of that language.

Inflection, in modern English, is very simple, as compared to the one of other European and extra-European languages. Although considerably more complex than English *inflection*, the morphology of *deriving terms* is also relatively simple, in English, as compared to the derivation means proper of many languages spoken all over the world. For exemple, terms can be conjured up not just be *concatenating* morphemes (as usual in English or even Neo-Latin, Romance languages), but even, and most often – in languages such as Arabic, Hebrew, Aramaic, or Amharic (Ethiopia) – by "plugging" the *radicals*, that is, the separate consonants of the root, into "holes" interspersed in a string of vowels and sometimes also consonants, that constitute a grammatical pattern (a free-place formula). Different families of languages exhibit different grammatical means for forming words, and those means are often very intricate, as compared to those usual in English, and sometimes are such, that they represent a tight constraint on the adaptation of foreign words. It is easy, in English or in other languages with a *concatenative* model of word-formation, to take a foreign word and naturalize it with just some refurbishing in the ending, or in pronunciation. Naturalization in Semitic languages, instead, requires that only the consonants be kept, of a foreign term, and then they should not be more than three or sometimes four (the latter, limitedly to the case where the middle two consonants are consecutive – with no vowel between – in the original, foreign term). Otherwise, when trying to "plug" those consonants in some native free-place formula, some would be left over. That means that, most often, international terms, usually long, just cannot be naturalized in such languages: if adapted as *is*, they would be "barren", that is, no verb or further derivative could come forth out of them, while, say, in English or even the more conservative Romance languages, *loanwords are productive*, fertile, because derivatives are formed by just attaching some suitable suffix or prefix at the end or the beginning, respectively.

3. Score attributes: 'Auro' and 'Aurc'

In ONOMATURGE, each frame associated with a lexical entry is a tree of properties describing the entry, and that is coded by means of nested attributes that constitute a hierarchy. See an example in the **Appendix**.

Such frames include, inside the subtree of properties associated with each single acceptation (wordsense), the following numeric attributes:

.*AURO*, standing for *Actual Use Relevance Ordinal*, a property whose numeric positive value expresses a subjective estimate of the relative frequency, in actual use in the present stratum of the language considered, of the term that owns the frame, with respect to the AURO value stated for the other acceptations of the same lexical entry. This way, a *poset* (partially ordered set) of AURO values is found in a single frame; for example, [1, 2, 2, 7]. The acceptation whose AURO value is *1* is, conventionally, the most frequent one, while a score such as *7* is associated, instead, with rare acceptations.

. *AURC*, standing for *Actual Use Relevance Cardinal*, a property whose numeric positive value expresses a subjective estimate of how frequent the lexical entry is, as expressing the given acceptation, with respect to the whole lexicon. Practically, I found it convenient to assume that, say, *AURC = 800* indicates a very high frequency, while *200* indicates rarity, while very rare instances (possibly, terms from past historical strata of the languages that are no longer in normal use but are still understood), could have, for exemple, a value such as *70* associated.

Let us think of AURO values of acceptations of the same lexical entry (and, therefore, indicated in the same frame) as being aligned along an axis: the fact that two consecutive values

are close rather than far away, makes ambiguity even bigger, and is exploited by the scoring component of ONOMATURGE – that composes retrieved numeric values according to certain functions – and this fact alone makes the contribution of AURO to the output scores (as synthesized by the scoring component for a candidate neologism are not), nonlinear with respect to the AURO axis.

Figures 1 and 2 express somewhat more formally these notions about lexical entries and the AURO and AURC scores.

```
Lexicon:   { L1 , ... , Li , ... , Lm}



Lexical
entry:

 Li  ---------/-------->   { A1 , ... , Aj , ... , An}
              n
```

Figure 1:
Lexical entries (either terms or roots), such as L_i, have frames associated. A frame in ONOMATURGE is represented as a tree of properties (a nested relation), coded by means of nested parentheses. In particular, the tree of properties associated with a lexical entry can have several lexemes or acceptations, each represented in a particular subtree. In this figure, the name of the lexical entry is the root of the tree, while A_j is the root of the subtree of a particular acceptation, whose schema is shown, instead, in Figure 2.

```
                 -----  Descriptions of denotation and connotation.
                 |
                 |
                 |
Acceptation      |----  RELATED_TO ----- RULES:  { R1 , ... , Rz} × Aj
                 |
        Aj ----|
                 |
                 |                          Li  ∧  Aj
                 |----  AURO:  ------------------------------------
                 |                 Li × { A1 , ... , Aj , ... , An}
                 |
                 |
                 |
                 |                          Li  ∧  Aj
                 -----  AURC:  ------------------------------------
                                 { L1 , ... , Lm} × { A1 , ... , An}
```

Figure 2
A simplified schema of the subtree of properties rooted in a certain acceptation, that can be considered to be the intersection, $L_i \wedge A_j$, between the semantic concept symbolized by A_j and the lexical entry L_i that

conveys it. Arrows in the figure indicate the way properties are nested. Upper case strings (RELATED-TO, RULES, AURO, AURC) are names of attributes. Under RULES, there is a list of pointers to derivational rules able to convey the semantic concept A_j. We consider the subtree rooted in RULES to describe a set that is the Cartesian product of all derivational morphemes in the language, and the given semantic concepts (with zeroes considered in those cases that, in the virtual R/A table, correspond to an inexistent association). We may say, formally, that AURO expresses the ratio between the given acceptation (the intersection between the lexical entry and the semantic concept), and the set of all acceptations of L_i. The fraction line stands for "with respect to". The Cartesian product at the denominator is the set of all of the semantic concepts that L_i is capable of conveying, that is, the vector resulting from the product of that lexical entry for the universe of all semantic concepts represented in the lexicon. As to AURC, it expresses the ratio between the given acceptation and the set of all the acceptations of all entries of the extant lexicon.

4. Quantifying diffusion: a discussion

In a multilingual project, the schema of a machine-dictionary for the lexical roots of the whole family of Semitic languages (Nissan, *to appear*), *frames of lexical roots shared by these kinned languages include a subjective estimate of how diffused* acceptations the lexical root is, in terms of *how many languages and derivatives* include them. This is *not* the same as *AURO* and *AURC*, that, in turn, should not be mistaken for the exact frequency of occurrences in a given textual *corpus* (calculated by *leximetric* tools), albeit they try to approximate if for a virtual "corpus" that actually is the collection of lifelong verbal and written communication that yielded the linguistic competence and experience of the human expert in his native language.

Subjective numeric scores are commonly associated with knowledge in the construction of *expert systems*, and are an important device in knowledge engineering. I noticed that linguists are often reluctant to accept the use of such heuristic values; yet, linguistics has no peculiar problems, under this respect, as compared to other knowledge domains: the epistemological discussions is general, in expert systems, concerning the quantification of imprecision, and several formal approaches have been developed in order to make subjective quantification more reliable. There is no easy way to quantify, and the ambition of automating certain kinds of common-sense knowledge calls for such representations. It is the *untested* knowledge representation that is unreliable. Scores estimated subjectively are likely to cause perturbations. *Tuning* is a most important phase of knowledge engineering. In the case of ONOMATURGE, tuning can be performed according to the feedback of extensive testing of the performance of the expert system. In the knowledge-acquisition, the results of leximetric analysis on significant corpora could be used to support the determination of scores encoded, but for the time being we are not using leximetry; for this discipline, see Muller 1973, 1985.

Another important thing to observed about estimates of diffusion, is that actually, the topic of the context is extremely important for predicting diffusion. However, ONOMATURGE is meant to provide the user with proposals of neologisms that – because of a choice to that effect – should be as unambiguous as possible even outside a specific technical domain.

5. Ambiguous word-formation patterns

Not only lexical entries (and, in particular, roots) are ambiguous. The morphological formation patterns, too, are *polysemous*, as they, too, are able to convey one out of several meanings.

Let us see an example from English. The *Oxford English Dictionary* lists *elephantry* as derived from the noun *elephant* and the suffix *-ry* just as English *cavalry* was derived from the (Italian) noun *caval[lo]*, i.e., "*horse*", and the suffix *-ry*. However, "*a military unit riding on <noun>s*"

is not the only possible meaning of the suffix *–ry*. Let us examine the problem in further detail. *Camelry* means "*a force of men mounted on camels*". However, the suffix *–ry* does not mean "a set of riders" as applied to the formation of the noun *poultry*. Instead, *–ry* has there a collective meaning directly applied to the broader meaning of *poult*: "*poults, chickens, ducks, geese, turkeys, etc.*".

In *apery*(1), that is, in the first English *lexeme* (lexical entry) associated with the string *apery*, the suffix *–ry* forms an abstract noun, with a negative connotation, and meaning a characteristic of *apes*: "*foolish mimicking*".

In *apery*(2), instead, the suffix *–ry* indicates "*a place*": "*a monkey houses in a zoo*". The locative meaning (that is, "*a place*"), of the suffix *–ry* is documented also for another entry in the same dictionary *–camelry–* but *camelry* as meaning "*a place where camels are held*" is much rarer than the one referring to troops (see above).

In the example of *cavalry*, phonetic adaptation occurred – at a certain point in the history of English – of a word borrowed from French, *cavalerie*. With another meaning, *chivalry* is instantiated in English. However, let us ignore *phonetics*: English examples such as *camelry* or *elephantry*, or even *apery* or *poultry*, allow *morphological* word–formation to be considered alone, in relation with *semantics*. What kinds of different phenomena can occur? Even if we considered only one particular meaning of the suffix *–ry*, some other phenomena of word–formation are involved: *cavalry* was not formed directly in English, but it arrived into an ancient historical stratum of English from a stratum of French, where an equivalent and related suffix was resorted to. Beside that *cross–language* phenomenon, we may meet *variants*, and extensions to broader – "neighboring" – meanings (i.e., *metonymy*) of the term that was formed by means of the suffix. For exemple, now *cavalry* is extended to motorized military units that once mounted horses (that is: they kept the organization and the name of ancient units). Moreover, as we have already seen, the variant *chivalry* exists in English; it is used in figurative speech. *Chivalry* was derived from the French word *cheval*, instead of Italian *cavallo*, even if both mean "*horse*", and both *cavalerie* (whence *cavalry* was derived) and *chevalerie* (whence *chivalry* stemmed) exist or existed in French, where those words were formed. This way, we find also kinship of suffixes in different languages that were related historically; beside kinship, the suffixes, or particular meanings of them, may exhibit some degree of equivalence.

Please observe that we limited the class of the examples to nouns (*camelry, elephantry, cavalry, apery, poultry*) that were formed out of a noun meaning some "*animal*", and of the suffix *–ry*.

Yet, ambiguity occurs: how would you understand *ostrichry?* Well, "*a certain negative quality of ostrichs*" is most likely. Moreover, the best known ("morally" negative) quality of ostrichs is the one related of in the story about ostrichs hiding their head in the sand, in order to avoid seeing the danger. This originated the expression *ostrich policy*. As to the sense of *ostrich policy*, it is perhaps the likeliest interpretation of *ostrichry*, even though the story about ostrichs hiding their head in the sand, is only a myth. (However, what about "*an ostrich stomach*"?) On the other hand, according to another possible meaning of *–ry*, the interpretation "*a farm where ostrichs are bred*" is also likely, for *ostrichry*: *–ry* may have, indeed, a locative meaning, and actually there are specific places for breeding ostrichs. The preference for understanding *ostrichry* either as meaning a particular case of "*a certain kind of place*" or as meaning a particular case of "*a certain bad quality*", would depend on context. Less likely to be guessed is the meaning "*a group of persons mounted on ostrichs*", even if some Roman matrons used to ride on ostrichs out of snobbery, and even if perhaps some other people in the world happen to mount those birds, that indeed *are* good runners. A possible objection is that snobbish ladies (and dandies: let us avoid gender–bias) usually dislike wearing or featuring something other ladies (or dandies, respectively)

that are present wear, etc. The pretension of rarity or even uniqueness may be taken, then, as being an indicator against the *collective* meaning. However, one may counter-object that there may be a collection of any particular thing which happens or is meant to be rare or unique in a given point in spatio-temporal coordinates. If that is the case, and the occurrence of the phenomenon is only sporadic in a given culture, would the phenomenon be percepted in that given culture as being a concept deserving a word of its own? (Sporadicity would justify a *lexical gap*, that is, the absence of a specific term for a given concept.) Or maybe affectation can be expected to lead an author – sometimes – to coin an awkward term to indicate an awkward phenomenon?

One may complicate more and more knowledge needed for disambiguation, and resort – for example, by adding facts and rules to a database resorted to by non-monotonic reasoning, say, to some modal logic of likelihood – by detecting suitable examples requiring the given complication. We do not consider, in ONOMATURGE, such far-fetched and, after all, sterile subtleties. Yet, we are interested in a tentative evaluation of how easy it would be to guess the intended meaning for candidate neologisms; certain cases offer few hints: how would you understand, say, *elkry*? Probably, its meaning would be identified as if the term were *elkness*, that is, "*the quality of being an elk*". If we mean *elk* as being a species of big, deer then we have nothing likelier, to hold on (but maybe, an elk is to a deer, as a big, coarse version of anything is to that thing? Metaphor would be involved).

Then, one could have ambiguity because the base itself is polisemic: *elk* is either a species of animal related to deers, or even a member of the *Elks*, a Rotary-like association, and then, according to the U.S. slang, *elk* may mean the same as "*a square head*". Hence, we found a quality, and its connotation is bad relatively: nonconformists – and hippies all the more – are known to dislike "*square heads*".

In this section, among other things, we have developed a discussion of word-formation in English, concerning only one rule – *-ry* – as applied to a limited class of stems: those indicating an animal. As one can see, determining the clarity of word-formation is not an easy task. From the viewpoint of speculative research, the topic certainly deserves a computational model. We implemented such a model in an expert system.

6. Scores for word-formation patterns

Inside the subtree of properties (that is, the *subframe*) associated with acceptations of lexical entries, several kinds of attributes, concerning different kinds of knowledge, are associated. They pertain either to the wordsense as specifically expressed by given lexical entry, or, instead, to the semantic concept on its own: unlike our choice in the design of the multilingual Semitic machine-dictionary (see in Sec. 4), in ONOMATURGE we chose not to separate the representation of semantics from the representation of the lexicon.

Among the attributes nested inside the subframe of single acceptations, there may be RULES, under which formation-patterns are stated that are capable of conveying the acceptation considered as a component in the sense of terms coined by employing those patterns.

Sometimes, some pattern (usually, a composition pattern) may be shared by all acceptations (and therefore, a RULES property may be shared, correspondingly). This rarely happens for patterns of formation by morphological derivation, instead; derivation patterns are typically specific of the various acceptations of the lexical entry, as they are related not to the lexical entry owning the frame, but, instead, to the semantic concept of the acceptation, a concept that is practically independent of the lexicon.

With each single formation-pattern listed, a *RO property is associated. The RO* attribute – *hat* may also be stated in full: *RELEVANCE-ORDINAL* – has, as value, a subjective numeric estimate

of the relative frequency of the considered formation pattern, when expressing the given meaning, with respect to the frequency of the other formation–patterns stated in the same RULES subtree (because they are also capable of expressing the same semantic concept). Figure 3 represents the RO score according to the same formalism introduced in *Figures 1* and *2* for AURO and AURC.

```
RULES  -----------/--------->   ( (IS  <rule> )  (RO  <value> ) )
                  n
```

```
                 ------- IS:  identifier of the derivational rule.
                 |
                 |
IS/RO pair ----|
                 |
                 |                    Rp ∧ Aj
                 ------- RO:  ------------------------------
                                 Aj × { R1 , ... , Rp , ... , Rz }
```

Figure 3

Under the RULES attribute, there is a list of derivational rules (with RO scores associated). Among the semantic concepts that each rule, R_p, is capable of conveying, A_j, is included. That is, the rule indicated as value of IS, in an IS/RO pair, is the intersection $R_p \wedge A_j$.
The RO numeric score of this intersection indicates how diffused is the use of R_p to convey A_j, with respect to the set of all rules also capable of expressing the same semantic concept, that is, with respect to the vector resulting from the Cartesian product of the given semantic concept, and the set of all derivational morphemes of the language.

Moreover, each derivation pattern is associated with a program that

. selects the *allomorph*, that is, the particular form of the derivational morpheme, according to phonological properties of the root, and

. outputs the term obtained by applying that allomorph to the lexical root.

The other kind of formation patterns – composition patterns – do not need each a specific program, but one program handles them.

Not only: every derivation pattern may have several acceptations, and deserves a frame, whose structure is very similar to the structure of lexical frames. Each acceptation of the derivation pattern has a subtree of property, that includes *AURO* and *AURC*, that, to avoid confusion, can be referred to as *rule–AURO* and *rule–AURC*.

. *Rule–AURO* embodies an estimate of the diffusion of the acceptation of the pattern with respect to the diffusion of the other acceptations of the same pattern.

. *Rule–AURC* expresses an estimate of the diffusion of the acceptation of the pattern with respect to the whole lexicon, considered as being the whole set of instantiations of application of the collection of all formation–patterns of the considered language. That is, the *rule–AURC* expresses an estimate of how diffused is the formation pattern with respect to the whole *morphological word–formation ruleset*, that is, to the collection of all formation patterns, in whatever semantic sense they are instantiated in the lexicon.

7. Remarks on subjectiveness

The value of numeric attributes discussed so far is a single integer or real numbers, but *pairs* of such numbers are arguably more suitable, if interpreted as being the extremes of ranges, either heuristically, or in the framework of a formal *ultrafuzzy* model (cf. Zadeh 1984). Two reasons support scoring by pairs (and, then, by ranges), instead of just by single numeric values.

. *First*, the determination of AURO, etc., does not stem from exact leximetric measures (cf. Sec. 4), and should be regarded as subjective, and possibly varying according to the person that provides the estimate, even when persons involved have a comparable competence in the language considered (and even ignoring the effects of bilingual interference, that are still important in the specific case of Hebrew: interference may involve associations of ideas with terms from outside the native lexicon). Let us express AURO by an interval, in order to have such variation accounted for. Even if we suppose that native speakers – and, thus, experts, given the domain we are considering – X and Y have the same linguistic background and a comparable experience of life (to exclude possible effects of different acquaintance with technical terminologies due to different professions, hobbies, etc.), they may happen to provide different AURO values for a set of two acceptations, P and Q, of the same term. This may happen both for a pair of acceptations that are more or less diffused comparably, or in estimating how rare a lexical entry (either a term, or a root) should be considered as in a given acceptation. Figure 4 associates straight lines with each single acceptation, intending by this to indicate the AURO axis for each acceptation. On the straight line of an acceptation, we indicate values provided by the two persons, labeling abscissae accordingly with the label associated with the given person that stated the value. Now, the figure illustrates a situation where person X considers the diffusion of acceptation Q to be prevalent with respect to acceptation P (that is, X thinks that the root or the term is likelier to express Q than P), while person Y believes the opposite is the case.

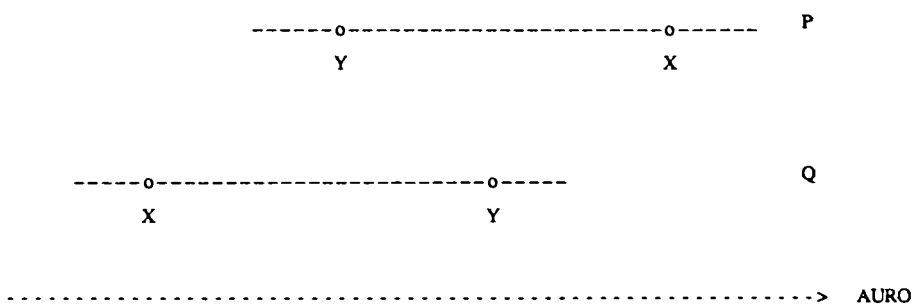


Figure 4
AURO values as differently stated for two acceptations, P and Q, by two persons, X and Y, competent in the language.

. Second, the same person may find it more convenient and plausible to indicate a *range* for AURO (as opposed to a single value).

Thus, we have a duplicity of ranges:

. ranges whose extremes correspond to the most divergent estimates provided by different persons,

. and ranges as stated by the same person: the most divergent opinions (at the extremes of ranges as of the former kind) are ranges themselves.

However, it is known, from the literature of fuzzy sets and systems, that *higher-order ultrafuzziness* may be reduced to simple ultrafuzziness. This way, the metaphorical *wooden Dad* case shown at the rightmost bottom corner in Figure 5 does not require further complicating the format of representation, as simple ranges, corresponding to *ultrafuzziness,* are enough, with no need for "*ultra²-fuzziness*".

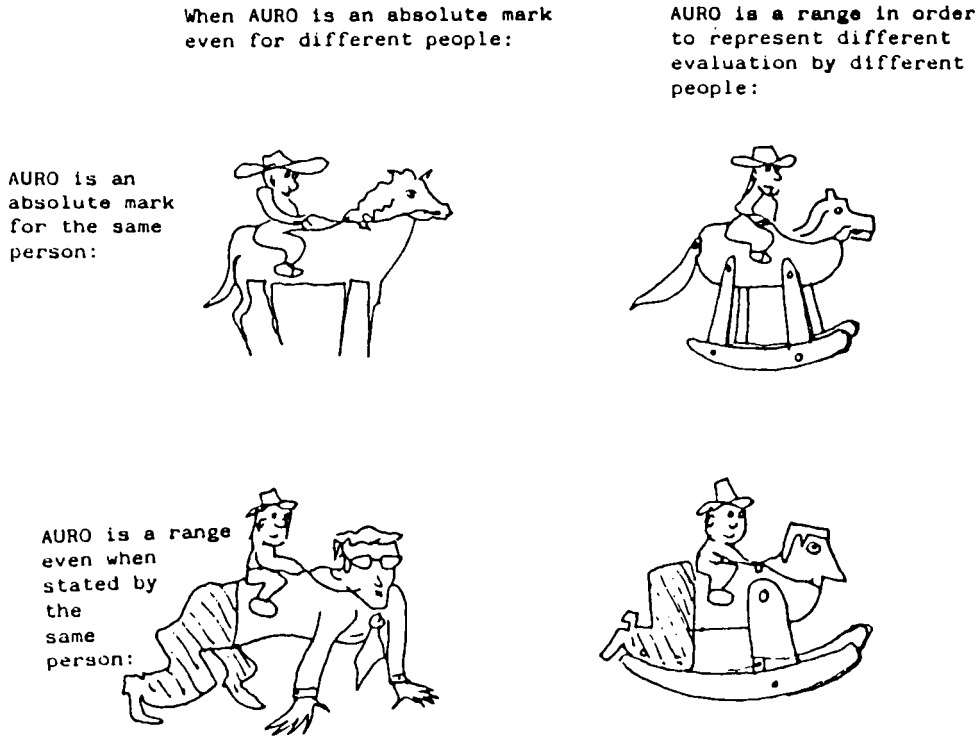


Figure 5
Possible combinations of representational choices for AURO.

In a collection of frames intended to represent a considerable portion of the lexicon and of the derivational ruleset of a given language, one should care about *consistency*. It would not be realistic to expect that the process of coding frames, including numeric scores, could prove always consistent, because of the limited knowledge that different encoders have of each other's specific estimates, and because of the limited recollection a person may expect to have of his or her own past estimates; trying to reconstruct a landscape of the situation of scores, to be of practical use when scores are stated, should be made often, and this would be prohibitive, ergonomically. On

the other hand, one could even try to have the expert system *exploit inconsistencies*: automatic interpolation between values retrieved and found to be contradictory because of the redundancy of information, is legitimate, if we assume that inconsistency reflects two valid estimates, that correct each other. Therefore, it is not necessarily wrong to have a *overconstrained* system of scores.

8. Static versus dynamic aspects

The various numeric values discussed so far are composed by the scoring component of ONOMATURGE by means of a set of heuristic functions.

. First, a function specific for each scoring attribute transforms the value retrieved for that attribute into an intermediate value, according to a segmentwise or continuous, generally nonlinear behavior. For example, the contribution of RO is "filtered" through a parabolic function, shown in Figure 6.

. Then, joint effects of the various contributions are accounted for by composition functions that take as arguments the various "filtered" contributions.

Legend
is ro2
RO–contribution
C is ----------
4
where Ro is the value of Ro as retrieved from the frame, and Ro–contribution is the filtered value output by the heuristic function which specifically handles Ro.

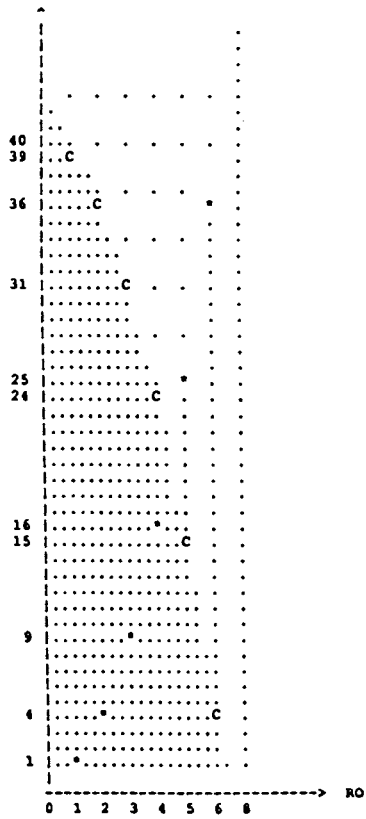


Figure 6

*A sample function drawn out of the heuristic scoring component of the control of ONOMATURGE. The function concerned here is the intermediate filter that transforms the retrieved RO value. Its output is C*4. The C curve delimits the dotted area.*

We are not going to discuss further, here, the functions of the scoring component of ONOMARTURGE (that is, the *dynamic* aspects of scoring, concerning the control of the expert system at run time), both because of scope and space limitations, and because those functions are rather intuitionistic and arbitrary, and have been tuned and refined during a testing loop.

In this paper, we have focused on the *static* aspect of scoring in ONOMATURGE, that is, on the definition of the numeric attributes that convey information on lexical diffusion as stated in the lexical database of the expert system, and that, once retrieved, are used by the dynamics of scoring in the estimation of the clarity of word-formation.

9. Concluding remarks

For a computational model of word-formation with the ambition of providing support for native language planning and fostering – as the case is for ONOMATURGE – it is essential that it be capable of evaluating how "good" the terms it coins are. Of more general an interest, a word-formation capability, for generation or analysis, is an enhancement for natural-language processing systems, as it would make them more robust when faced lexical material they did not previously have knowledge explicitly encoded about.

We discussed some numeric attributes that, stated in frames of the knowledge-base of ONOMATURGE, allows it to estimate how clear terms it forms are going to appear, to native speakers of Hebrew. Morphological knowledge, some knowledge in semantics (just the dichotomy of wordsenses), as well as some knowledge in sociolinguistics is involved.

Numeric scores could refer to the diffusion not of terms, but of knowledge of concepts or of semantic features, in age-classes or professional groups, in order to select an apt presentation in text-generators. For exemple, whenever a concept is considered to be unknown to the expected audience (e.g., children), a path of explanation should be sought that conveys the message plainly, through semantic concepts that either the audience is expected to be familiar with, or can be explained without having to exceed a given threshold of contrivance. In Nissan 1987 b, I have discussed the representation of knowledge of semantic features, of their typicality and exception-admissibility, and of the sociocultural diffusion of awareness thereof. This adds a *socio-epistemic* dimension to the *sociolinguistic* dimension we have been concerned with in the present paper. The socioepistemic dimension is capable of providing support for the representation of knowledge fitting in the sociolinguistic dimension itself, to the extent that the representation of knowledge in semantics and on the interrelationships of topics lends itself for estimating the likelihood of occurrence of given acceptations in given semantic or cultural contexts.

Acknowledgements

I am grateful to the Information Processing Association of Israel, that has prized my ONOMATURGE project with the *1988 IPA Award of Computer Science*.

References

Aronoff, M. 1976 *Word Formation in Generative Grammar*. The MIT Press (Cambridge, Mass.)

Carbonell, J.G. and P.J. Hayes 1984 *Recovery Strategies for Parsing Extragrammatical Language*. Technical Report CMU-CS-84-107, Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pennsylvania.

Czap, H. and C. Galinski, Eds. 1987 *Terminology and Knowledge Engineering*. Indeks Verlag (Frankfurt/M, FRG). Proceedings of the International Congress on Terminology &

Knowledge Engineering, Trier, FRG, 29 September - 1 October 1987, sponsored by INFOTERM (UNESCO, Vienna).

Haas, N. and G.G. Hendrix 1984 *Learning by Being Told: Acquiring Knowledge for Information Management*. In: Michalski, R., Carbonell, J.G. and T.M. Mitchell, Eds., *Machine Learning: An Artificial Intelligence Approach*. Vol. 1. Tioga Publ. (now: Morgan Kaufmann Publ.), Palo Alto, California.

Muller, C. 1973 *La statistique linguistique*. In: Pottier, B., Ed., *Le langage*. La Bibliothèque du CEPL - Centre d'Étude et de Promotion de la Lecture (Paris): pp. 288-303.

Muller, C. 1985 *Langue française, linguistique quantitative, informatique*. Slatkine (Paris & Genève).

Nir, R. 1981 *Acceptability and Normativism in the Use of a Revived Language*. In: Hartig, M., Ed., *Angewandte Soziolinguistik*. Tübinger Beiträge zur Linguistik 170. Narr (Tübingen): pp. 89-98.

Nissan, E. 1987a ONOMATURGE: *An Expert System for Word-Formation and Morpho-Semantic Clarity Evaluation*. Part I: The Task in Perspective, and a Model of the Dynamics of the System (pp. 167-176). Part II: The Statics of the System: The Representation from the General Viewpoint of Knowledge-Bases for Terminology (pp. 177-189). In Czap and Galinski 1987.

Nissan, E. 1987b *Exception-Admissibility and Typicality in Proto-Representations*. In Czap and Galinski 1987: pp. 253-167.

Nissan, E. (to appear) *Structure and Representation for a Multilingual Semitic Machine-Dictionary: Issues in Engineering a Lexicon Meant for Symbolic Manipulation*. In: Schmidt, K.M., Ed., *Collected Articles and Papers on Conceptual and Content Analysis by Computer* Two thematic volumes in *Advances in Computing and the Humanities*. JAI Press (Greenwich, Massachusetts).

Zadeh, L.A. 1984 (August) *Making Computers Think Like People. IEEE Spectrum* 21,8: pp. 26-32.

Zipf, H.P. 1949 *Human Behavior and the Principle of Least Effort*. Addison-Wesley (Reading, Massachusetts).

Appendix

A lexical frame from the database of ONOMATURGE

```
(assigne-frame  mak$ir; pronounce: ''makhshir''

(  (MORPHO-CAT   (  (SEE-ACCEPTATION)          ;  For the Hebrew
                    (noun masculine singular)  ;  term considered
                    (participle masculine singular) ;  the morphological
                 )                             ;  category depends
   )                                           ;  on the wordsense.
```

```
(GENERAL-INFORMATION.................................)


(ACCEPTATIONS: The terms may express four different acceptations.

  (ARE
      (  (MEANING  (KEYWORDS   (tool) )  )  ;  "tool" is one out
                                            ;  of the wordsenses.

         (MORPHO-CAT   (  (noun masculine singular)   )      )

         (RELATED-TO

            (RULES  ;  Word-formation rules, capable of
                    ;  denoting a "tool" (possible out of
                    ;  a set of acceptations of the rule),
                    ;  and subdivided by morphological kinds:

               (NONCONCATENATIVE-PATTERNS

                  (  (IS (maP^eL) )  (RELEVANCE-ORDINAL (1) )  )
                  (  (IS (Pa^yan)  )  (RELEVANCE-ORDINAL (3) )  )
                  (  (IS (maP^an) )  (RELEVANCE-ORDINAL (6) )  )
               )
               (SUFFIXES.........................)

         )     ;  end of RULES

         (RELATIVE-FREQUENCY-AMONG-SYNONYMS  (1) )

         (SYNONYMS

                  (ARE
                       (  (IS (kli)  )
                          ..............
                       )                  ;  end synonym
                  )                       ;  end ARE
         )                                ;  end SYNONYMS
      )                                   ;  end RELATED-TO
```

```
    (HISTORICAL-PRECEDENCE-ORDINAL (2) )
    (ACTUAL-USE-RELEVANCE-ORDINAL (1) )  ;  AURO
)  ;  end acceptation.

    ;  One more acceptation follows:
(  (MEANING (KEYWORDS (enabling) )  )
    (MORPHO-CAT  (  (participle masculine singular)  )  )
    (RELATED-TO.............................)
    (HISTORICAL-PRECEDENCE-ORDINAL (1) )
    (ACTUAL-USE-RELEVANCE-ORDINAL (2) )
)  ;  end acceptation

    ;  One more acceptation follows:
(  (MEANING (KEYWORDS (that-makes-become-kosher)  )  )
    (MORPHO-CAT  (  (participle masculine singular)  )  )
    (RELATED-TO.............................)
    (HISTORICAL-PRECEDENCE-ORDINAL (3) )
    (ACTUAL-USE-RELEVANCE-ORDINAL (3) )
)  ;  end acceptation

    ;  One more acceptation follows:
```

```
        (   (MEANING   (KEYWORDS    (declaring-kosher)    )    )

            (MORPHO-CAT     (    (participle masculine singular)    )    )

            (RELATED-TO.............................)

            (HISTORICAL-PRECEDENCE-ORDINAL  (3) )

            (ACTUAL-USE-RELEVANCE-ORDINAL  (3) )

        )   ;   end acceptation

    )       ;   end ARE

)           ;   end ACCEPTATIONS

[           ;   end all open parentheses.
```

José António Barreiros

Advogado, Lisboa, (Portugal).

A arte combinatória jurídica no século das tecnologias da informação

Resumo

As tecnologias de informação no Direito, que a força persuasória do mercado tenta tornar uma finalidade autónoma, são preparatórias da resolução de problemas jurídicos heurísticos por uma via que se reclama de inteligência artificial.

Com a presente comunicação pretende-se demonstrar a eventualidade de essas tecnologias potenciarem novos padrões cognitivos no âmbito do sistema jurídico e racionalizarem, enquanto associadas a modelos formais de lógica deôntica, a auto-regulação global do sistema social.

Abstract

The Information Technologies for Law, despite the attempts of the market persuading force to make them an autonomous purpose, are preparatories to solve heuristic legal problems by means related to AI.

The aim of this paper is to show how these technologies could potentiate new cognitive patterns within the legal system and, being associated with formal models of deontic logic, rationalize the self-regulation of the social system.

1. Assumamos o Direito tal como ele resulta dos modelos cognitivos elaborados pela teoria geral dos sistemas e pela teoria geral da informação cujos trabalhos fundamentais se devem a von Bertalanffy, Shannon, Weaver e Colin Cherry:

(1) o Direito configura-se como um sistema, dotado de uma estrutura:

(1.1.) sistema, enquanto complexo de variáveis fácticas ou normativas relacionadas através de uma rede causal interactuante;

(1.2.) estrutura, dado o carácter relativamente estável de algumas dessas interrelações;

(2) as relações sistémicas do sistema jurídico não se configuram como um paradigma absoluto, mas de acordo com graus variáveis de sistematicidade;

(3) o sistema jurídico está integrado, numa perspectiva holística, através de acoplamentos de tipo hierárquico ou pelo menos de relações cinemáticas de tipo arborescente, numa rede de sistemas de tipo sociocultural, as quais constituem o meio ambiente daquele, sucedendo que:

(3.1.) por meio ambiente deve entender-se um conjunto de estados, elementos ou acontecimentos relativamente distinguíveis ou autonomizáveis, que podem constituir eles próprios um sistema;

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 1992, APDC & SPF, Lisboa

(3.2.) sendo variedade do sistema o número desses estados autónomos que o observador, em função dos seus padrões de discriminação, consiga individualizar;

(4) trata-se de um caso de complexidade organizada dado o elevado número e distinta natureza da variedade possível dos estados possíveis do sistema e o carácter não rígido da estrutura, susceptível de comportamentos alternativos para além dos que pressupostamente resultaram da programação do legislador;

(5) é um sistema aberto, de tipo morfogénico, dado que o intercâmbio com o meio é factor essencial para a respectiva estruturação e viabilidade, capacidade reprodutiva e de mudança, bem como para a elaboração de níveis diferenciados de estrutura a níveis progressivamente mais complexos de organização, os quais significam graus sucessivamente mais refinados da regulação jurídica em face da complexidade crescente da vida comunitária;

(6) é um sistema adaptativo, dotado de capacidade de aprendizagem em função do seu relacionamento com o meio ambiente;

(7) neguentrópico, por fazer corresponder o aumento da sua ordem interna e da respectiva organização aos distúrbios emergentes do seu envolvimento com o meio, sendo o desrespeito e a fraude à lei causa do aperfeiçoamento deste e da sua melhor organização;

(8) cuja trajectória está dirigida para metas, sendo certo que:

(8.1.) por trajectória entende-se a sequência dos vários estados das variávies constitutivas do sistema;

(8.2.) e por direcção para metas a orientação finalística cujos objectiovs não resultam necessariamente de pré-programação, antes podem derivar da reorientação estratégica do sistema no decurso da sua trajectória.

(9) o processo de organização do sistema constitui uma relação mapeada deste com o meio exterior, através da elaboração de um modelo tendencialmente isomórfico (a) dos outros sistemas integrantes da rede sistémica; (b) da sua própria trajectória, sendo:

(9.1.) modelo a estrutura relativamente à qual são válidos os axiomas cuja validade relativamente ao sistema modelado esteja certificada;

(9.2.) isomórfico a particularidade de um modelo relativamente ao qual (Ashby) a aplicação no modelo de uma transformação um-um dos estados de input e output gera a representação canónica desta no sistema modelado.

(10) o sistema está dotado de meios de sensibilidade e percepção às variações ambientais que possam significar mudanças relativamente às suas variáveis de critério, bem como aos estados da sua própria trajectória;

(11) o sistema está realimentado pela informação obtida pelos seus mecanismos de percepção;

(12) a realimentação informativa pode gerar, dentro do funcionamento auto-adaptativo, uma morfogénese caracterizada pela equifinalidade - estados diversos geram objectos idênticos - ou de multifinalidade - estados iniciais similares conduzem a estados finais díspares;

(13) o sistema jurídico é, assim, um sistema regulador auto-dirigido do sistema sociocultural;

(14) o intercâmbio de informação factual e normativa entre os emissores e os sistemas receptores de comunicação do sistema jogam um papel decisivo no processo de autoregulação;

(15) a transmissão de informação implica que tanto a fonte como o receptor tenham modelado a variedade ambiental segundo certos termos de referência comuns;

(16) a quantidade de informação fornecida por um sistema é inversa à sua organização,

sendo o resultado da previsibilidade comportamental daquele ou, o mesmo é dizer, a medida de redundância do sistema;

(17) assumindo a entropia como medida da incerteza verifica-se que a crescente produção normativa se correlaciona com a desorganização social sobre a qual tal produção incide.

2. Assumamos agora que o processo de funcionamento dessa macro-estrutura que denominámos como sistema jurídico passa por um modelo de raciocínio susceptível de ser formalizado de acordo com os paradigmas da lógica.

2.1. Autonomizemos as propostas decorrentes da aplicação da lógica alética, simbólica e de dois valores, que há quem tome como a única interessante para o processamento eletrónico do Direito (Mario G. Losano):

(1) alética enquanto reportada a um universo em que os estados eventuais do sistema são de (a) necessidade (b) possibilidade (c) contingência (d) impossibilidade; e no qual as conclusões lógicas são de verdade ou falsidade, ou dizendo de modo a acentuar o ponto nodal da questão, uma lógica que pressupõe que os enunciados normativos são redutíveis a valores de verdade;

(2) simbólica enquanto referente a um raciocínio que, tentando evitar as aporias da linguagem natural, se reporte à redução das normas jurídicas a sinais e proposições cujas regras de formação e de transformação obedecem a uma sintaxe lógica dada;

(3) de dois valores, respectivamente os mencionados de verdade e falsidade, o que exclui a existência de outros valores do tipo verdadeiro/falso/possível (Tarski) ou verdadeiro/falso/p começa a ser ou não p começa a ser (Rogowski).

(2.2.) Mas centremo-nos na probalidade operativa sugerida pela lógica deôntica cuja estruturação se ficou a dever ao finlandês von Wright relativamente à qual se pode dizer que:

(1) estabelece um sistema integrado de compreensão holística do universo ao aprofundar as analogias estruturais de tipo isomórfico entre as modalidades aléticas (relacionadas com o campo da verdade), epistémicas (referentes ao conhecimento), deônticas (reportadas ao modo axiológico do ser) e existenciais (reportadas à circunstância da existência do ente), que o quadro seguinte (von Wright e Sanchez-Mazas) tenta ilustrar:

Tipologia de modalidades lógicas

Aléticas	Epistémicas	Deônticas	Existenciais
necessidade	verificado	obrigatório	universal
possibilidade	incerto	permitido	existente
contingência	falseado	indiferente	
impossibilidade		proibido	vazio

Incerto por tradução de "undecided" (von Wright), propondo Mazas em castelhano "no-decidido".
Falseado por "falsified" segundo von Wright.
Indiferente no sentido infra de contingência bilateral (A e não A são permitidos).

(2) extrai das analogias estruturais entre algumas das modalidades aléticas e deônticas, pois que mais intensas (nomeadamente entre *o deve, o pode* e *o tem que* aléticos e a

necessidade, possibilidade e impossibilidade deônticas) as regras e instrumentos para um cálculo normativo;

(3) permite a elaboração de um algoritmo orientado para a operacionalização de um sistema experto cujo funcionamento assente na lógica dita modal porque intencionalizada para a resolução dos problemas emergentes dos vários modos de apresentação do sistema normativo;

(4) supõe, enfim, uma lógica cujo objecto se estende para além do problema da verdade.

2.3. Concentremo-nos como ponto de referência relativamente à lógica modal no trabalho do matemático Sánchez-Mazas, sem esquecer o contributo dos argentinos Alchourrón e Bulygin e do americano Layman Allen num campo cujos fundamentos foram estruturados decisivamente com o polaco Kalinowski.

2.4. Assentemos nalgumas linhas de orientação metodológica que assumam especial valor heurístico ou criem uma relativa sinergia intelectual instigadora do ulterior aprofundamente da matéria.

Sistematizando algumas ideias no sector:

(1) a tipologia das relações deônticas fundamentais pode reconduzir-se ao seguinte esquema:

Tipologia	**Operadores**
implicação deôntica	
- positiva ou afirmativa de necessidade deôntica	O
- negativa de proibição ou de impossibilidade ou incompatibilidade deôntica	I
negação de uma implicação	
- negação de O, contingência unilateral deôntica positiva (faculdade unilateral)	F
- negação de I, possibilidade deôntica unilateral, permissão unilateral	P
permissão ou faculdade bilateral	F2 ou P2
regulamentação ou prescrição positiva ou de proibição	U

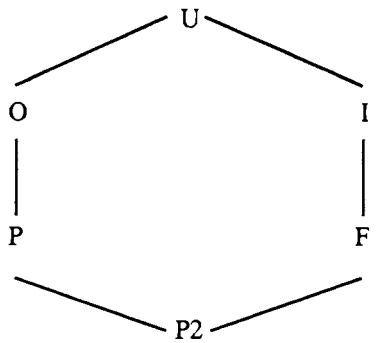
(2) tomando como antecedente deôntico uma condição elementar denominada como g e como consequente deôntico o facto ou acção a que chamaremos h e assumindo que as condições deônticas elementares se podem estruturar como conjunções deônticas tipo Kc1...cn, então será este o quadro da tipologia das proposições normativas:

Tipologia das proposições normativas

Necessidade	O (g, h)
Impossibilidade	I (g, h)
Faculdade unilateral	F (g, h)
Permissão unilateral	P (g, h)
Permissão bilateral	F2 (g, h) ou P2 (g, h)
Prescrição	U (g, h)

o que se lê como: satisfeita a condição g, então h será obrigatório, ou proibido ou permitido, etc.

(3) o que de acordo com o hexágono deôntico do tipo dos de Blanché e Kalinowski dá:



(4) o universo sobre o qual incide o cálculo deôntico e relativamente ao qual se constituem as relações deônticas é o sistema normativo constituído por um conjunto de normas que determinam o estatuto deôntico de uma determinada acção ou conduta em determinados casos pois que;

(4.1.) o sistema normativo é integrado por um conjunto de enunciados com algumas consequências jurídicas, formado por normas mas também por enunciados não normativos;

(4.2.) as normas integrantes do sistema normativo são enunciados de correlações entre certos casos e determinadas soluções;

(4.3.) o estatuto deôntico de determinadas acções ou condutas é caracterizado por aquilo que sejam as relações deônticas que as integram (v.g. conforme o estatuto deôntico de permissão, obrigação, etc.);

(5) um sistema normativo dado oferece um número finito de soluções a certos casos, variável em função das propriedades relevantes constitutivas do universo de propriedades daquele sistema, uma vez que:

(5.1.) as propriedades relevantes são aquelas que estando presentes em relação a um determinado caso geram quanto a ele um estatuto deôntico distinto, inexistente no caso da sua ausência;

(5.2.) a rede de relações lógicas entre as propriedades relevantes pode assumir vários comportamentos, expressivos dos vários compostos veritativo-funcionais de tal universo;

(6) sendo o sistema normativo uma rede deôntica de condições e factos/acções, poderemos considerar tal conjunção como:

(6.1.) satura- uma conjunção de condições (KKc1...cm) é satura numa rede deôntica se e apenas se uma e só uma entre as relações a seguir expressas é verdadeira, sendo a outra falsa: O (Kc1,,,cm,h) I (Kc1...cm,h);

(6.2.) hipersatura quando logicamente contraditória;

(6.3.) hiposatura quando permissiva do tipo P2 (Kc1...cm,h).

3. Assumamos finalmente como dados da experiência empírica que:

(1) ocorre actualmente uma produção legislativa cuja frequência significa um crescimento exponencial da massa documental respectiva relativamente às médias ponderadas das iniciativas legisferantes antecedentes;

(2) perante essa circunstância, amplificam-se os processos de incoerência e de inconsistência do sistema jurídico, verificando-se inclusivamente uma disrupção dos nexos formais da estrutura respectiva;

4. Chegados a este estádio de crise do sistema jurídico, que novos padrões cognitivos ou operacionais poderão emergir do contributo proposto pela inteligência artificial?

4.1. O primeiro é puramente pragmático, pois tem a vêr com a inevitabilidade do processamento automático das bases de dados jurídicos se efectuar no quadro de uma programação lógica das proposições normativas, que permita ultrapassar as aporias do tratamento em pura linguagem natural.

4.2. O segundo é construtivista quando implica, ao estilo leibniziano, uma arte combinatória entre símbolos representativos dos vários conceitos jurídicos relativamente aos quais possa operar uma cálculo proposicional cujas conclusões funcionam - integráveis com o bom senso da tópica - como meras hipóteses de trabalho de uma solução jurídica que possa encontrar outras vias de confirmação e legitimação.

4.3. O terceiro, finalmente, é ideológico, quando - ante o nó górdio das aporias lógicas verificadas pelo processamento automático do Direito - pretende a irredutibilidade do sistema jurídico à lógica, a inexistência, no que ao sistema jurídico respeita, de um sistema como tal.

Tratar-se-ia de proclamar como Claus-Wilhelm Canaris que "a tentativa de conceber o sistema de determinada ordem jurídica (...) como lógico-formal ou axiomático-dedutivo está, de antemão, votado ao insucesso (...)", pois que, e isso demonstraria segundo este professor de filosofia do Direito o carácter estrangeiro dos valores no território da lógica: "quem poderia seriamente pretender que a regra de tratar o igual por igual e o diferente de modo diferente, de acordo com a medida da diferença, possa ser acertada com os meios da lógica?".

Seria convocar o testamento filosófico de Hans Kelsen para — sabendo-o autor de uma teoria pura do Direito — dele ouvir a afirmação da inaplicabilidade das regras da inferência e do princípio da não contradição ao sistema jurídico.

Lembrar, em suma, que o próprio pai fundador da lógica deôntica, o finlandês von Wright, revendo em 1983 a obra da sua vida, havia excluído a existência de relações lógicas entre as normas, admitindo que aquela lógica modal se limitasse a fornecer os critérios de uma legislação racional.

Sem sistema, nem lógica, o Direito oferecer-se-ia assim na metáfora de Karl Popper aqui parafraseada, como uma ciência sobre palafitas.

Mas não disse também esse anarquista metodológico que dá pelo nome de Paul Feyerabend que a ciência só progride quando transgride as regras impostas pelos lógicos?

Ephraim Nissan Gilad Puni Tsvi Kuflik

Dep. of Mathematics & Computer Science,
Ben Gurion Univ. of the Negev, Beer-Sheva (Israel)

Finding Excuses With ALIBI: Alternative Plans That Are Deontically More Defensible

Abstract

ALIBI is a program that analyzes actions listed in an accusation, and then proposes alternative explanations of the behavior imputed. Actions are recursively decomposed into constitutive elements, and objectively ascertained events are separated from deontic (i.e., legal or moral) interpretations; then, objective elements are recomposed and integrated into a legitimate plan of action, that, proposed by the defense as an interpretation of ascertained facts, is meant to exculpate the person indicted.

Resumo

ALIBI é um programa que analisa as acções arroladas numa acusação e, seguidamente, propõe explanações alternativas para o comportamento inculpado. As acções são decompostas recursivamente em elementos constitutivos e os eventos objectivamente certificados são separados das interpretações deônticas (i.e. legais ou morais). Seguidamente, os elementos objectivos são recompostos e integrados num plano de acção legítimo que é alegado pela defesa como interpretação dos factos certificados, com vista à ilibação do acusado.

1. A Definition and Description of 'Alibi'

ALIBI a program coded in *Prolog,* is a planner that generates and proposes *pretexts,* alternative interpretations (in the following, improperly termed "*alibis*" by generalization), for behavior reprehended and ascribed to a human being that *ALIBI* tries to defend.

A session with the program involves the following operations, in the order:

- A list *imputed actions* are fed into *ALIBI* as specified in a simplified "*police report*". For the time being, there is no component, in the system, for analyzing sentences in natural language, so the input is formulated by the user as a set of logic predicates.
- *ALIBI* decomposes the actions listed in the input into *constituent actions,* recursively, until "*atomi*" *terminal actions* in the decomposition *tree* are reached.
- The program spoils them of connotational interpretations pertaining to *deontic* modalities (*legality or morality*). For example, *stealing* is reduced to *taking* as in given objective circumstances.
- Then, starting from the atomic constituent actions, another tree is constructed,

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 484/510
1992, APDC & SPF, Lisboa

bottom up: constituent actions are recomposed to form an interpretation of the behavior imputed that is deontically legitimate, or, at least,less reprehensible.

- For each interpretation generated, a simplified set of English-like utterances, expressing the interpretation synthesized, is displayed: this phase is unsophisticated in the implemented prototype of *ALIBI*, because we have used just a composition of *canned text* (ready-made fragments of sentences, possibly interspersed with variables), but future versions could exploit the representation of semantic knowledge that subserves the previous phases, in the framework of the generation of more natural output text.

In this paper, we discuss both the present architecture, and some shortcomings, by outlining also a broader, more sophisticated model.

2. Sample Sessions

Let us consider an input accusation as follows, first formulating it in English:

> *"The defendant robbed a pack of diamonds at a jeweller's shop: he broke the display window, by throwing himself against it, and sneaked inside the shop through the broken glass; on seeing the jeweller, the defendant fired with a sub-machine gun and wounded the jeweller, and then took away the pack of diamonds."*

For the prototype of *ALIBI*, we have to formulate this accusation by means of *Prolog* predicates (and rather primitively so):

```
done(rob, diamonds_pack,jeweller-s_shop).
done(injure,sub-machine_gun,jeweller).
done(break,body,display_window).
done(sneak,display_window,jeweller-s_shop).
done(take,diamonds_pack,jeweller-s_shop).
```

Several excuses are alternatively proposed by the program:

1) *I was pushed on the display_window and it broke.*
I fell and accidentally the sub-machine_gun shot the jeweller.
The diamonds_pack should be watched, so I took it
in order to keep it and bring it back.

2) *I did not mean to do it,*
the sub-machine_gun fired accidentally
and the jeweller was injured.
The diamonds_pack should be watched, so I took it
in order to keep it and bring it back.

An injured shopkeeper is helpless, and cannot watch his property by himself...

3) *I entered the jeweller-s_shop to leave a note*
telling I broke the display_window accidentally.
I fell and accidentally the sub-machine_gun shot the jeweller.
The diamonds_pack should be watched, so I took it
in order to keep it and bring it back.

Having taken valuables is certainly incriminating, more than just having been caught while ransacking a drawer, which could have been motivated by the need of finding paper and a pen to write the note (unless it is ascertained that the defendant already had material for writing), but by now, *ALIBI* does not carry out such reasoning.

4) *I heard strange voices, so I entered to check,*
through the display_window.
I fell and accidentally the sub-machine_gun shot the jeweller.
The diamonds_pack should be watched, so I took it
in order to keep it and bring it back.

In the latter case, the defendant has chosen to play the hero: he claims he intended indeed to break the glass and enter the shop, because he believed he could conclude that there was somebody inside that either was committing mischief, or needed immediate help.

5) *I committed a mistake, I took a bag from the*
jeweller-s_shop that looked like my bag,
and the diamonds_pack was inside.

Now, let us consider another accusation:

"*The defendant robbed a bank: he threatened an employee with a rifle, and snatched the money.*"

In predicates, we formulate the input as follows:

done(rob,money,bank).
done(threaten,rifle,employee).
done(take,money,bank).

This time, the excuses proposed are:

I did not aim at the employee with the rifle.
I had the rifle
because I was on vacation from the army.

The employee panicked and gave me the money.

An alternative excuse for going around with a gun is:

I had the rifle
because I went hunting.

The present version of *ALIBI* does not try to check whether the defendant is in the army (which the police could easily check), or whether he is a licensed hunter and the season is the hunting season (anyway, going hunting without a license, or when hunting is forbidden, is a lesser offense with respect to robbing a bank). Besides, the setting may be important, as in the United States, citizens can carry weapons legally; generally, hunters are not allowed to detain semiautomatic rifles, except in the United States. Besides, in certain countries (Israel and Switzerland) citizens are often called into reserve duty, for years after they completed their military service. Such knowledge can be added into the program, as we are going to see in the following sections.

Even if it were true that the employee panicked, misinterpreting the intentions of the defendant, this would not justify, by itself, the defendant for having accepted the money. The defendant could claim he did not mind the rifle and had misinterpreted the feelings of

of the employee. Receiving the money – he could claim – did not arouse his suspicions, as he was asking for a loan. *ALIBI* does not go to such lengths, but future versions could ponder even such considerations.

Let us consider, now, a slightly modified accusation, replacing the *rifle* with a *knife*:

done(rob,money,bank).
done(threaten,knife,employee).
done(take,money,bank).

This time, excuses proposed are:

I did not aim at the employee with the knife.
I held the knife
because I was carrying it to a grinder
to have is sharpened

The employee panicked and gave me the money.

The generation of excuses the way they are displayed, is performed, by the present version of the program, as follows: once the elements of the excuse have been determined and put together, the program invokes predicates that, for each element, print *canned text,* that is, short fragments of text, possibly with variables interspersed that are replaced by a suitable value (for example, *the...shot the...* is completed with the kind of gun, and with the identifier of the victim). For the time being, the inclusion in the canned text itself of some explanatory elements has not been completely avoided, but ideally, natural–language text should be generated by an unobtrusive linguistic component from a complete semantic internal representation: future versions of *ALIBI* should include a natural–language processing component capable of performing both analysis and generation.

3. Some Remarks on Justification

In the input, to state an act of *sneaking* occurred, is represented by an instance of the predicate schema

done(sneak, THROUGH,INTO)/

such as

done(sneak,display_window,shop).

Instead, an event involving the act of *breaking to gain access* is formulated according to the format

done(break,THROUGH,INTO).

Stealing involves, as arguments, *OBJECT* and *LOCATION*. Further arguments would be appropriate: the number of specified arguments for different actions may vary, while the value for some arguments may be unspecified; we interpret arguments in a *deep–case* optics (cf. Harris 1985: Chs. 6 to 8), making arguments in given ordinal positions stand for the object, the instrument, the location, etc.

Let the input accusation be:

done(hit,stone,glass).
done(break,stone,display_window).

done(get_in,display_window,shop).

Please observe that the *display_window* is told to have been broken, while a particular component, the *glass*, has been hit. An excuse that is proposed, then, is:

I hit the glass with the stone
because I wanted to kill an insect that was standing there.

A similar accusation is:

done(throw,stone,display_window).
done(break,stone,glass).
done(get_in,display_window,shop).

Where the act of *throwing* is aimed at the whole *display_window*, while it is stated that a particular component, the *glass,* was broken. An excuse is:

I threw the stone on the display_window
because I wanted to kill an insect that was standing there.

Another excuse that can be thought of is:

I threw the stone on the display_window
because I wanted to attract the attention
of a person who was standing by.

The latter excuse is even less apt than the previous one; both of them are not very credible. In the last excuse, the defendant, to explain why he hit a fragile object – the glass – that was an obstacle to entering the shop, claims that his goal was to draw attention. This is *futile*, as the plan adopted for the goal of attracting the attention of another person was costly and risky: instead of calling that person, he allegedly hit the glass, that was fragile, costly, and necessary for the protection of somebody else's property. The risks were of breaking the glass, and of being accused of evil intentions concerning property the glass guarded. Hitting the glass was instrumental, but on the other hand, the glass was the object of that instrumental action, and the object for hitting, too, should be considered. Hitting by means of a stone is less costly than hitting with a watch, that is, in turn, less costly than hitting with the head of a child. Even more futile than the goal of attracting attention, the goal would have been, for the defendant, of relieving himself of his anger after looking at himself in the glass and noticing his dress was soiled; on the other hand, affective situations, such as anger, may lessen the ability of a person to behave rationally.

Having employed a stone is considered to be an ascertained fact, but it would have been helpful, for the defendant, to be able to claim the window was broken by an object that is more normal to carry in the street: carrying a stone is a plausible indicator of premeditation. The defendant has a relative interest of acting the fool, by claiming he tried to crush an insect on the glass (which may justify, to a limited extent, the fact he procured a stone: but was he motivated by anger, or by compulsion?), or that he was carrying a stone (for some lawful purpose that he should invent) and then he slid. Presently, there is a representation, in *ALIBI*, of stones being hard, and of insects being rather soft; more properly, the chitin coating of insects is not a soft material, but the coating is thin enough to prove fragile when subjected even to a very moderate effort by humans. Another claim could have been that the glass was already broken, but witnesses could be expected to be easily found to testify to the contrary.

4. Representation and Processing

Once the *Prolog* interpreter is entered, the user loads the file with the program *(alibi)* and the file with the accusation:

[alibi, accusation1]

then, to obtain just one explanation, the user has to type *alibi([ ])*. While to obtain a list of excuses, the command to be typed is *alibi(List)*.

Actions are subdivided in *elementary actions and compound actions*. This subdivision is conventional; for example, for the prototype of *ALIBI*, the actions of *taking aim at a target*, of *throwing*, or of *shooting with a gun*, are elementary actions:

elementary_action(aim).
elementary_action(throw).
elementary_action(shot).
etc.,

even though we may conceive them as being composed of more elementary actions. Instead, actions that are represented in *ALIBI* as being compound, constitute a hierarchy. The root of the tree, that is, the most aggregated action, that is relevant for the examples of **Sec. 2**, is *obtain*, while the most aggregated action that is relevant for the examples of **Sec. 3**, is *get_in_unlawfully* . Therefore, the top predicate in *ALIBI*, named *alibi*, has as argument the identifier of that most aggregated action. The present version of *ALIBI* accepts accusations that include detailed, elementary actions, as well as an action at an intermediate level of aggregation (such as *steal* or *rob*).

Intuitively, to obtain the physical control of an object, some lawful means could be employed, as opposed to *robbing*, or *stealing*, or *swindling*. Getting into proximity of the object is necessary along with *taking*; in turn, *getting in* (a private place) is analyzed into alternatives, that may include just *entering*, as opposed to *sneaking* or *breaking in*, either in presence or in absence of the owner or anybody that is delegated to guard the property (a private watchman, a policeman, a watch dog, or even just honest passers-by or neighbors). *Breaking in* is furtherly analyzed, by indicating alternative plans that may accomplish that action; to the extent that a given choice occurred and was ascertained and included in the accusation, that is the path that is determined, while whenever circumstances were not ascertained, the defendant can choose among appropriate plans for constructing an excuse. *Robbing* requires *threatening* or *overpowering* or *injuring* or *killing* the owner (or a watchman), actions that can be accomplished by *shooting*, or *stabbing*, etc. Beside the hierarchy of actions, also their *effects* are indicated.

Compound actions are represented as a list of more elementary actions: it is a logical conjunction *(AND)* of those constituent actions. On the other hand, each compound or constituent action may correspond even to several predicates, that each state a different way that action could be accomplished: an action represented by predicates specifying alternative species of its, is a logical disjunction *(OR)* of those alternatives. Therefore, the set of actions represented in the database constitutes an *AND/OR graph* (cf. Nilsson 1982); in such directed graphs, those vertices that are the successors of a given vertex – let us term it *V* – have to be selected (that is, reached from *V*) either jointly *(AND)*, or alternatively *(OR)*. The top predicate of *ALIBI*, that is, *alibi*, is executed through three phases: *how*, that visits the composition tree of actions, *collect_effects*, that takes care of the effects of actions, and *why*, that recursively invokes *explain* predicates for the various actions.

Explanations are developed separately for the various actions imputed, but, if this is relevant, with knowledge of other actions having occurred or having to be assumed.

In *Prolog* rules, first the *consequent* of the rule is given, on the left of : - (the implication symbol). On the right of this symbol, the *antecedent* of the rule is coded; it includes conditions, as well as operations that must be executed by the *Prolog* interpreter when conditions are verified. The name of a predicate is followed by a parenthesis containing the arguments of the predicate, unless the predicate has no arguments. Brackets enclose lists.

Here is an example of a *Prolog* rule that describes the compound action on *threatening by means of a gun:*

```
compound_action(threaten,[aim,hold]) :-
        done(threaten,X,Threatened_Victim),
        is_a(X,gun),
        asserta(has_as_effect(frightened),
        etc.
```

that is rendered, in English, by the statement:

> *"threaten is a compound action that has to be assumed as constituted by the sequence of actions aim and hold, if the accusation imputes, among actions performed by the defendant, having threatened the victim by using a gun as instrument; if this is the case indeed, then define, as being an effect, a predicate* has_as_effect(frightened)., *where* frightened *is assumed to refer to the victim. (etc.)"*

Then, the predicate *has_as_effect* will appear among the conditions belonging to the antecedent of at least one rule whose left part is the predicate *explain* with a suitable values as argument.

For example, stemming from an ascertained or claimed situation where threatening led to the victim being frightened, one possible heuristic for devising excuses in such a situation is expressed by:

```
asserta(goal(gave)in_panick,_,Threatened,_Victim)).
```

that means

> *"Define, as a goal to be currently pursued, the generation of an excuse claiming that the victim, in panick, gave the defendant property he is accused of having obtained unlawfully (robbed, as the situation requires the presence of the possessor)."*

and is invoked in the antecedent of a rule, after the condition *has_as_effect(frightened).*, included in the same antecedent. Please observe that in *Prolog*, an underline character (that is, _) stands for an unspecified argument.

Besides, an *effect* predicate states the effects - *help_needed* (for the victim) and *guard_needed* (for property) - that follow the effect *frightened:*

```
effect(frightened,[help_needed,guard_needed])
```

The antecedent of *explain* predicates includes especially predicates of three kinds: *done* (a condition requiring that a certain kind of action be included in the the accusation), *has_an_effect* (that selects one of the effects of such an action), and *goal* (that invokes the

generation of an excuse according to a given heuristic that is suited for the particular situation).

For example, the following invokes, for predicate *explain(aim)*, the generation of two alternative excuses, claiming that the defendant aimed at the victim accidentally, because he had fallen through the way of access (say, a broken display-window), or denying that the defendant ever aimed at the victim:

```
explain(aim):-done(threaten,X,Y),
        has_as_effect(fall_through),
        goal(accidentally_aim,X,Y,).
explain(aim):-done(threaten,X,Y),
        goal(not_aim,X,Y).
```

In case shooting, not just aiming, is imputed, a modified heuristic is invoked for predicate *explain(shoot)*.

Predicates of kind *goal*, whose first argument is the identifier of a particular excuse-generation heuristic, select and compose fragments of *canned text*, and display the resulting sentence in the context of the output alibi. One of the simples kinds of *goal* predicate is exemplified hereby:

```
goal(accidentally_aim,Kind_of_Gun,Victim):-
        write('I fell and accidentally the'),
        write(Kind_of_Gun), write('aimed at the'), nl.
```

where *nl* causes a new line to be started on the display.

More complex *goal* predicates invoke subservient predicates in order to select apt wording, and to take care, albeit rudimentally, of morphological or syntactic rules. One of the simplest examples is the selection of the word *myself* for insertion in an excuse where the *body* of the defendant was employed as an instrument (say, in case the defendant is claimed to have fallen or tho have been pushed, thus breaking the display-window):

```
select_object(A) :- A ==body, write('myself').
select_object(A) :- not(A==body), write('the'),write(A).
```

The following wording predicates employ *PROPEL* one of the primitive acts of Schank's *conceptual dependency theory*, as a general concept, and, according to the specific action imputed or claimed, select a suitable verb, and construct appropriate phrases - with *deep-case* arguments inserted - to accompany it:

```
select_PROPEL_act(hit,What) :- done(hit,_,What).
select_PROPEL_act(threw,Upon_what) :- done(throw,_,Upon_what).
```

that is

> *"Select verb* hit *or* threw *according to the accusation stressing one action or the other."*

```
select_PROPEL_act(threw,Target_of_PROPEL) :-
        done(throw,_,Upon_what),
        (is_part_of(Upon_what,Target_of_PROPEL);
        is_part_of(Target_of_PROPEL,Upon_waht) ).
```

that is

> *"A broader or a narrower concept may have been stated as being the target*

of the PROPEL act: a part of the whole may be hit implying that the whole was hit, or, vice versa, the whole may have been stated as being the target, while it is a particular component that was actually hit''.

Please observe that in *Prolog*, logical conjunction *(AND)* is indicated by a comma between predicates belonging to the antecedent of a rule, while, instead, logical disjunction *(OR)* is indicated by a semicolon.

So far, predicates considered handled *common-sense semantics;* in the following, *phrase generation* is performed:

```
select_PROPEL_arguments(threw,Instrument,Target_of_PROPEL) :-
        select_object(Instrument).
        write('on the '),
        write(Target_of_PROPEL).

select_PROPEL_arguments(hit,Instrument,Target_of_PROPEL) :-
        write('the'), write(Target_of_PROPEL),
        write('with'), select_object(Instrument).
```

These two predicates embody linguistic knowledge about the construction of phrases according to the template *threw...on the...* as opposed to the template *hit the...with...* Indeed, verb *throw* requires the prepostion *on*, while verb *hit* requires the direct object. Such predicates are invoked by *goal* predicates, in the antecedent part of the latter.

There are semantic predicates that state properties of objects: about *stone, gun* and *diamond* being *hard,* about *diamond* bein a *jewel* and about the latter being *small*, etc.

Predicates

```
can_be_tool(Object,hit) :- is_hard(Object).
can_be_tool(Object,hit) :- is_heavy(Object).
can_be_tool(Object,wound) :- is_sharp(Object).
```

tell that hard or heavy objects may be used as tools for hitting, while sharp objects can wound.

Predicates

```
is_part_of(glass,display_window).
        is_part_of(glass,windshield).
        is_part_of(glass,watch).
        is_part_of(windshield, car).
        is_part_of(diamond,diamonds_pack).
```

tell that glass is a component of display windows, windshields, and watches, that windshields are a component of cars, and that diamonds are found in a *diamonds_pack*: the latter, overspecific predicate about *diamonds_pack* is necessary because we have not adopted, for the time being, a more general representation that would analyze *a pack of diamonds* as being an instance of *pack*, with particular properties implied by the properties of the general concept *pack*.

5. Remarks on the Effects of Actions

There is a representation of the *effects* of the various actions (either compound, or atomic), and effects may have, in turn, further efffects. Effects are exploited to generate explanations, either

- for other actions that have taken place, or
- for actions claimed to have been planned.

For example, in one implemented exercise, a man is accused of having shot and wounded a shopkeeper, and then robbed him. Let us focus on *wounding*.

I) One effect is that *the wounded shopkeeper was unable to take care of his property.* The robber can claim that, having shot the shopkeeper accidentally,

a) he actually took away property,
b) in order to look after it and then return it.

II) Another effect is that *the wounded shopkeeper needed medical aid.* The robber can claim that

a) he actually ran away,
b) in order to get such help for the shopkeeper.

Please observe that in case the robber was caught while he was still in the shop, he did not manage to run away, but he can still claim he was about running away to get medical help. The potential action does not belong to the accusation, but it belongs to the alibi. In the prototype of *ALIBI*, we focused on actions that had taken place, and on possible causes and goals, while not handling, for the time being, actions that did not occur at all.

Effects are explanatory for events that occurred and were imputed, and in such cases, the choice among possible effects is determined by the event; instead, in devising an excuse for the defendant, it is possible to choose among possible alternative effects according to the interests of the defendant, while certain effects necessarily stem from situations ascertained or claimed, and cannot be denied even in a claim: indeed, like what we have already seen for the composition of actions, also the set of effects constitutes an *AND/OR tree*.

Breaking (an obstacle to access) has the physical effect that the obstacle is removed, and the associated social common-sense that, to the extent that the obstacle considered guarded property, there is no more such guard.

The effect of *threatening* is that the person threatened if *frightened*. Both an *injured* (and therefore *helpless*) person and a *frightened* person need help, and the offender can claim he meant to provide such help himself, or to look for somebody other able to provide it (e.g., professional help may be *medical treatment*), or for an instrumental object or abstract notion (for example, a phone is an instrument for conveying information, that in turn is a possible instrument for calling for help).

On the other hand, the efffects of *frightened* may be quite the opposite: a frightened person may be able and motivated to defend him- or herself beyond actual need, and thus he or she is dangerous; for example, some years ago, in Italy, a well-known football player, Re Cecconi, threatened, in jest, a jeweller who did not recognize him and therefore killed him.

6. Remarks on Compound Actions

Our approach to the decomposition of actions bears some resemblance to Burger's *Wordtree* approach. Burger's nomenclator (1984) – a harcopy book – is based on a binary branching process of definition of action verbs (including English neologisms that Burger defined to complete the hierarchy of concepts: the approach allows no lexical gaps). Sample definitions are:

MANAGE BY COMMANDING = To SMOKESTACK
MANAGE BY LISTENING = To AD-PRAC
FASTEN GROUNDED OBJECT = To STAKE
FASTEN HOOKED OBJECT = To CLASP
FASTEN SEWN OBJECT = To SUTURE

TO RIDE and MANEUVER = To JOCKEY
To CUT and SLIVER = To JULIENNE
To COOK and INFLATE = To SOUFFLÉ
To ESCORT and RESPECTABILIZE = To CHAPERON
To WORD and VOICE = To SAY

However, in *ALIBI* the decomposition of actions is not required to be a *binary* tree.

In a more accurate model, compound actions can be defined by a graph, including not only a list of more elementary actions (as we have seen in **Sec. 4**), but also their relationship in terms of time, enablement, causation, location and proximity, etc.: cf. (Schubert, Goebel and Cercone 1979: 5.1). Such a detailed, more realistic representation would require that the recomposition of *S* (the ascertained elements of an accusation) into alternative compound actions be performed according to a more complicated scheme than considered so far: sets of *S* should be matched against the graphs defining compound actions, used as templates.

Instead of analyzing actions by decomposing them into a small set of primitive acts, as done in the *conceptual dependency theory* of the Yale school (we have seen the primitive act *PROPEL*, in **Sec. 4**), we could adopt Alterman's approach in *NEXUS* (1985); according to that approach to representation, concepts are related according to various kinds of relationship:

- *taxonomic*: it is the *CLASS/SUBCLASS* relation, such as the one that links *taking* and *snatching*;
- *partonomic*: this kind of relation includes two categories:
  - *SEQUENCE/SUBSEQUENCE:* for example, *travel* has three subsequences: *depart, move* as being an intransitive verb, and *arrive*; and
  - *COORDINATE:* for example, a coordinate of *chopping* is *holding;*
- *temporal:* two categories belong to this kind of relationship;
  - there is the *Before* kind, that includes *ANTECEDENT/necessary* and *PREC-EDENT/plausible,*
  - and the *After* kind, that includes *CONSEQUENT/necessary* and *SEQUEL/ plausible.*

For example, an event that must necessarily occur before *dropping* is *holding*, that thus

is an antecedent of *dropping*. A consequent of *dropping* is *falling*, a necessarily following event. Instead, some other relations are just plausible: a precedent of *drinking* is *opening*, as sometimes a container has to be opened, while a sequel of *hitting* is *breaking*, a plausible but not necessary event following *hitting*.

In our present representation, the predicates of the input accusation include no argument for representing temporal relationships. Only some implicit knowledge is embodied in the sequence of constituent actions associated with compound actions. Future versions should include an explicit representation and exploitment of temporal information.

7. Intermediate Nodes, in the Hierarchy of Actions

Instead of always reaching for atomic constituent actions, it could be useful to allow intermediate nodes in the decomposition to be stopped at. Presently, intermediate nodes (i.e., non-atomic constituent actions) have to be recomposed, after having been totally decomposed: such nodes correspond to the root of the subtree shared by the decomposition tree and by the recomposition tree, and the difference does not seem so important. However, once extensions are introduced that handle argumentation to further depth, then it would be useful to introduce the capability of having coomonly occurring substructures that are not atomic, handled on their own as a whole.

As Mary Dee Harris points out (1985: p. 246), *flexibility* in the level of decomposition of semantic concepts is an advantage, and is subserved by representing events as a sequence of states that explicitly takes into account *time* relationships. For example, in the sentence *John stubbed his toe while walking down the sidewalk*, it is apropriate to consider *walking* as being a sequence of states that involves moving one foot forward while being supported by the other leg and so on, as this way, it is possible to relate, analytically, *walking* to *stubbing a toe*, the latter being a concurrent event that involves "*bringing the toe of the foot being moved forward into contact with an unmovable object with a force great enough to cause pain.*" *(ibid.)*. Instead, there is no point in breaking *walking* down to states that explicitly involve actions with feet, as far as the sentence *Terry and Susan watched the sunset as they walked along the beach* is concerned, while such decomposition would no longer be out of place if then *Susan stepped on a shell and cut her foot. (ibid.)*.

Representation formalisms we consider to be suitable candidates for adoption in the next phase of the *ALIBI* project, are:

- *extended semantic networks* as defined by Schubert, Goebel and Cercone (1979);
- *conceptual dependencies* and *memory organization packets (MOPs)* of the Yale school of natural-language processing (Schank 1982 a). A good introduction to conceptual dependency theory is Harris 1985: Chs. 8,9.

8. Deontic Liability Minimization

The representation and handling of the deontic dimension could be refined to various degrees.

- Rules could be reformulated as explicitly trying to *minimize* deontic liability, instead as just pretending that liability should be eliminated altogether – which is an oversimplification – just by eliminating from the set of recomposition trials, those recomposition trees that involve a deontically reprehensible node. Presently, *ALIBI* actually does not even try to eliminate liability altogether, as all alternatives are tried

out. A small step forwards would be having some reprehensible but not too incriminating possibilities allowed, just because they have not been explicitly labeled as being inculpating. The next step would be extending deontic labeling of action nodes, including nuances of offense according to a deontic meter.

- A qualitative jump in terms of artificial intelligence, would be the introduction of a *machine-learning* capability, according to the technological state of the art. In particular,
  - the system should be endowed with a component for the acquisition of new rules, embodying *patterns of deontic liability minimization*;
  - then, the system should be enabled to discover such patterns automatically. To do the latter, the system should be initially trained to recognize good patterns, by having the human expert approving of, or rejecting, patterns that the system has "discovered" by analyzing runs that have led to good "alibis" ("good", because approved by the expert),
    - ☐ first by adopting the overspecific pattern of given successful alibi-generation sessions, and
    - ☐ in the following, by trying to generalize patterns, which is tentatively done
      - ○ by eliminating elements from recomposition trees, or
      - ○ by replacing more general concepts – that is, nodes that are superordinate in the IS-A hierarchy – for nodes that occurred in the recomposition tree of the given successful session.

There is more, to discovering good liability-minimizations, than just

- avoiding high-liability nodes, or
- avoiding the cumulation of liability by mere addition of the degree of liability as associated with nodes.

Indeed, qualitative and quantitative evaluations are called for, that take into consideration such dimensions as:

- *credibility*
- *falsifiability*
- etc.

Falsifiability – that is, the possibility of ascertaining whether a given claim is false – is linked to both

- *documentability* – whose minimal value is *undocumentedness* – or the ability to produce witnesses or material evidence (by distinguishing between such an ability on the part of the defence as opposed to such an ability on the part of prosecution); and
- the possibility of having *complicity* involved, to the extent that it fabricates (or conceals) evidence, or provides, instead, a genuine alibi.

Quantification would enable the program

- to compare the legal punishability or moral reprehensibility of competing alibis;

and

- to compare alibis according to how easy it would be to prove or disprove them, along with qualitative considerations on enablement; an actual offender should compare advantages associated with risk, with making statements with the least commitment.

The prototype of *ALIBI* does not distinguish explicitly between

- argumentation *ad veritatem*, whose goal is to ascertain truth, whatever it is; and
- argumentation *ad personam*, that is, as convenient to persuade a given interlocutor (here, the judge or the jurors).

However, the implicit general aim is the latter. An explicit distinction between the two kinds of argumentation would subserve reasoning on interest and falsifiability. A rather sophisticated computational model of argumentation has been implemented by Flowers, McGuire and Birnbaum (1982) in the program *ABDUL/ILANA*, that simulates different roles in arguing about the conflict in the Middle East. This model handles *adversary arguments*, a class of argumentation in which *"neither participant expects to persuade or be persuaded. The participants intend to remain adversaries, and present their arguments for the judgement of an audience... In these arguments, an arguer's aim is to make his side look good while making the opponent's look bad." (ibid.)* This class of arguments is opposed to *persuasion arguments*, with participants that try to persuade the opponent, and are, in turn, willing to be persuaded.

Further topics that call for modelling are:

an explicit representation of what defence – embodied by *ALIBI* – is aware of or conjectures (to a certain degree of likelihood), including its knowledge or speculations about what the *prosecutor* is aware of or conjectures, and can or cannot prove, or is in the process of constructing a proof for;

a model of evaluation of *(un)supportiveness* of concomitant claims: this attribute is not necessarily binary, and intermediate degrees are possible;

a model of evaluation of possible *synergetic* effects of concomitant claims: in particular, the relation of concomitant claims to objective facts (possible or ascertained supportive or incriminating evidence), as opposed to considering those claims separately;

levels of abstraction concerning the *defendant's strategy*, the *prosecution strategy* as reconstructed by *ALIBI*, and *counterplanning* as incorporated in both strategies;

strategy adaptation, as estimated to be convenient because of current or global goals of the parts, and as constrained by previous commitment;

investment, other than modifications in terms of falsifiability as involved in manoeuvering. For example, it may be convenient to make believe the defendant is a fool, but a loss in terms of credibility is likely to be involved in the process;

an explicit formal structure for dealing with the variation in terms of mutual advantage; for example, a logic-based approach to *game theory* (cf. D'Amore

1976) is one possible formal approach. Classes of search procedures in artificial intelligence are related to operations research and game theory (Barr and Feigenbaum 1981: Vol. I, **II.B3**, **II.C5**);

we have not investigated cooperation or betrayal among acting characters yet: *complicity* of defendants, or the relationship between a defendant and accomplices that are not indicted yet, under certain respects involves a complication of the prisoner's dilemma, because of the lack of communication between accomplices, albeit we could admit even that communication occurs, to a certain degree, imperfect when earnest, and sometimes possibly misleading deliberately;

accounting for *ongoing action* on the part of police or accomplices;

coping with *miracles*, that is, external events that could occur in the world concomitantly with actions being carried out in the framework of the aftermath of the offense, but independently of the behavior of the parts in the offense, in investigation, and in the trial. A formal approach to *miracles* in theories of action is exposed by Lifschitz and Rabinov (1989), that correct a previous paper (Lifschitz 1987).

9. Implicit Knowledge on the Environment

Let us consider some way of enhancing the representation, and discuss it according to the following sample accusation:

> *"Standing on a hill near a road, the defendant injured the driver of a car by throwing a stone".*

For the time being, *ALIBI* is unable to analyze such an accusation, not only because it is formulated in natural language, but also because it does not involve just actions: the setting (the hill is near the road) is important, here.

From the viewpoint of alibi-generation, we would like to have *ALIBI* proposing an excuse like the following:

> *"I was aiming at a bird that was perched on the electricity wires, but I missed it, and the stone hit, instead, the glass windshield of a car that was passing by."*

In the explanation trial that generates this excuse, the defendant is assumed to be unable to deny it was him who threw the stone that hit the car driver. Anyway, the defendant denying his being the same as the person who threw the stone would lead to an excuse different from the one we have seen, while denying a stone was thrown by a person because, say, it was a meteorite, is very little credible even *prima facie*, and can be disproven by geological analysis.

Because of several reasons, the present version of *ALIBI* is unable to generate an excuse such as the one we have singled out above. Indeed:

- the accusation we have been considering allows no innocent explanation; therefore, the defendant claims the intended target of the action was another one, such that targetting it is less unlawful (albeit not lawful altogether);
- subserviently, the defendant is able to perform a relatively complex evaluation of

deontic effects, and then to replace a criminal action with several negative pretended actions, whose addition would yield a less severe penal judgement:

* premeditated injury is a fellony that would be punished severely;
* the same injury, procured without malice aforethought, is still a fellony, but would be punished less severely;
* tampering with public electricity wires is punishable, but is a misdemeanor instead of a fellony, and possibly punishement would be just a fine;
* hitting a bird is morally reprehensible, and, if performed, could have involved at most a fine, if there are laws forbidding causing suffering to animals under certain conditions; this introduces also to the necessity of a representation of the diversity of legal systems and possibly of cultural contexts, too;
* just having had the intention of hitting a bird is not punishable;
* admitting to having missed the bird may at most hurt the defendant's pride, as it perhaps witnesses of his ineptitude at aiming.

- Implicit common-sense knowledge is involved about, among other things, the structure of physical objects and of the environment:
  * about the structure, layout, and use of common physical objects such as a car: a driver has to be assumed to be found sitting behind the fragile glass windshield of every moving car;
  * about commonly found environments: the setting of the story is near a road, and roads are likely to be flanked by electricity poles, that are high and substain wires where birds are often perched;
  * in mechanics, even though at an elementary and naive level: the trajectory of thrown objets, with or without hitting an obstacle;
  * about the presence of wild (and, by default, healthy) birds in a given location being just ephemeral: the bird must have flown away even before the car driver was hit, if a bird there was, which is immaterial, because birds are used to be temporarily perched on wires; the point is that the defendant claims he was aiming at the bird. The very presence of the bird is unlikely to be proven.

Augmenting *ALIBI* with such knowledge would require the addition of some more architectural components, but *is* feasible, at the present state of the art of artificial intelligence.

10. When Several Characters are Involved, and: Grotesque, or Futility?

Let us consider the following story,that will allow us to point out some extensions that are called for.

> *"A car driven by* Man1 *stopped near a jeweller's shop. Two policemen were standing by.* Woman1, *sitten near the driver, stated kissing* Man1; *this compelled the policemen to turn their back, as they are gentlemen. Then* Man2, *who was concealed on the rear seat, got down and broke the display window of the shop, stole some jewels, and jumped back into the car, that then* Man1 *drove away."*

In the present version of *ALIBI*, we have considered only the goals and plans of the defendant, while not goals of other characters: victims, accomplices, witnesses, or other people. This way, the prototype does not account for *complicity*. The story we have seen cannot be handled by the present program. What is interesting to observe about *complicity*, in the story considered, is that *Man1* and *Woman1* committed no felony, if we consider their actions on their own, ignoring the actions of *Man2*, that is, ignoring complicity.

This story includes a grotesque assumption about the kindness of gentlemen: the thieves relied *a priori* upon the discretion of gentlemen being unquestionably active in policemen, even when contrasting with the latter's professional goals. However, it does not necessarily take unlikelihood to make a story *funny*; cf. Dyer, Flowers and Reeves 1989, that exposes a particular computational approach to *irony recognition* in narratives, based on the failure of plans of characters. Beside this aspect of the story, the policemen are considered as though they were a single character (same motivations, same plan, same actions), which is also unlikely, and contributes to the grotesqueness of the story. On the other hand, even with a likely story, a bad alibi can be funny, or – instead – just *incongruous* or *futile*. We have not tried yet to characterize these qualities coherently, but on *futility*, see **Sec. 3**.

11. Alternative Explanations, Beyond Application to Legality or Morality

In real life, excuses are not necessarily meant to deny legally or morally reprehensible behavior. Indeed, negative traits such as *incompetence* in a given domain – or even features that are just considered as being negative by the prejudices hold by specific environments: for example, belonging to or originating from certain social groups – could also be denied by an individual by inventing an alternative explanation of symptoms of the feature he or she wants to conceal. A possible extension of *ALIBI* could not only have deontic values as labels associated with actions in the representation of semantic concepts, but include also a similar indication of other features of human behavior or of people. Possibly, an order of gravity among features would also be established: for example, legal reprehensibility is likely to be considered, by the defendant, as being worse than just moral reprehensibility, that in turn is worse than incompetence (unless an individual prefers to claim his or her shrewdness caused a certain effect, while actually having behaved incompetently).

However, in general, just a decomposition of actions is not enough to produce alternative explanations. Common-sense facts and rules are required, that can be structured variously. Let us consider the following joke, translated from *La Settimana Enigmistica* 1973:

> *"In a provincial cinema hall, a placard announces the new program: 'A great color documentary film,* The Last Cannibals. *There is no trick. When the image is shaky, it is because the cameraman was afraid.'"*

Let us analyze this story. We have not implemented this analysis in a program; programming the specific case in *Prolog* would be rather easy, but developing a remarkably more general program would be very difficult.

The setting of the story, *a provincial cinema hall*, is qualified as *provincial*, and some likely – while not necessary – properties of the latter are: (technical) *incompetence, low standards* (of a certain class of services, and of shows in particular), *gullibility* (of the public), *charlatanism* (in shows). Against such a background, the following reasoning – (**1a**) to (**1f**) – is plausible:

1a) • image is shaky, in a film, if the camera was unsteady while filming; • the camera is unsteady if the hand of the cameraman is unsteady; • if the hand of the cameraman is unsteady, then the cameraman is professionally incompetent; • an incompetent cameraman is unlikely to get a job requiring much competence as a cameraman (*because* it is easy to notice incompetence, by testing: *see* **1b**). *Therefore*, • the image is shaky, in film *F*, *because* • the hand of the cameraman, *C*, of *F*, was unsteady, *because* • cameraman *C* is an incompetent in his profession; *therefore* • it is very unlikely that *C* would be hired as a cameraman for shooting a film requiring much competence.

1b) • It is easy to test the incompetence (that is, to obtain a negative evaluation of competence) of a cameraman as due to his hand being unsteady, *because* • testing can be done on a very short and inexpensive trial film, and • one easily notices a shaky image even in a very short and inexpensive film.

On the other hand,

1c) • *F* is claimed to be a film involving true cannibals; • cannibals belong in an exotic setting; • a film with true cannibals has to be filmed in their exotic setting; • shooting an exotic film in place is expensive; • producing an expensive film is justified only if the final product (the film) is of the highest quality; • a product of the highest quality has to be manufactured only by comptent personnel; *therefore* • the competence of prospective personnel has to be checked; *for which see* (**1b**).

1d) • Filming *X* requires being near *X*; • in *F*, *X* is claimed to be true cannibals; • being located near cannibals is dangerous; *therefore* • filming cannibals is dangerous. *Therefore* • the above claim about *F* implies a claim to the effect that filming *F* was dangerous. • It implies also a claim to the effect that *C* is professionally competent; *indeed*, • shooting a film in dangerous conditions requires competent personnel, *and, as we know from* (**1a**), • *C* is – by definition – the cameraman that filmed *F*.

1e) However, *because* of (**1a**) and (**1c**), as *C* was found to be professionally incompetent, *we conclude that* • *F* is not of the highest quality; *therefore*, • *F* is not an exotic film taken in place; *therefore*, • actors in *F* are not true cannibals; *which is contrary to claims on F* (see **1d**). *Therefore*, • *F*, the way it is claimed to be, is a cheat.

1f) Besides, *because* of (**1a**) and (**1d**), as *C* was found to be professionally incompetent, *we conclude that* • *F* was not filmed in dangerous conditions. *Therefore*, • *F*, the way it is claimed to be, is a cheat.

The placard in the story proposes an alternative explanation:

2a) • image is shaky, in film *F*, *because* • the camera that filmed *F* was shaking, *because* • the hand of the cameraman, *C*, was shaking, *because* • *C* was afraid, *because* • *C* was filming cannibals, *and – as already seen in* (**1d**) – • *C* had to be located near those cannibals, *and* • being located near cannibals is dangerous. *On the other hand, as we know from* (**1d**), • shooting a film in dangerous conditions requires competent personnel.

2b) Beside reasons (**1d**), also reasons (**1c**) are resorted to, here.

2c) *Therefore*, • *F* was filmed by professionally competent people (including *C*), *and* • *F* is of the highest quality.

There is an important point to be observe about the use of *afraid*, in the story

considered. In a way, it testifies of the courage of the cameraman, that has – knowingly, it can be presumed – put himself in danger, for the sake of his profession, or as a test of courage for its own sake. A suitable model is called for, to account for such seemingly paradoxical phenomena.

12. Pragmatic Tactics on what Information to Expose or Conceal

Assumptions on *mutual belief* are of different use in

- question–answering systems, such as interfaces of databases or knowledge–bases, where a governing criterion is that the tool should provide answers that satisfy the interaction goals of the user without misleading him/her even by correct answers (cf. Webber 1986 b: Sec. 4.3); versus
- systems that simulate a role for which subtle deception is a legitimate means.

For the first, Joshi's modification (1982) to Grice's Maxim of Quality (1975):

> *"If you, the speaker, plan to say anything which may imply for the hearer something that you believe to be false, then provide further information to block it."*

provides the criterion for the level of informativeness required to satisfy Grice's Maxims of Quantity (1975):

- *"Make your contribution as informative as is required for the current purposes of the exchange."*
- *Do not make your contribution more informative than is required for the current purposes of the exchange."*

Now, let us consider a system that may try to mislead on purpose: *IMP* is a dialogue–system impersonating an *informant*, and modelling the pragmatics of negotiation about some goods that a prospective client (an *evaluator*) is trying to assess: the program interacts with a user and tries to present the goods in a favourable light, that is, to convey a good impression about the given object (having a *positive bias* about that object); if necessary, *IMP* tries to mislead, without lying by an act of commission, while attempting to maintain a neutral image of itself and an impression of completeness for its own answers (even simulating insulted surprise if an intervening question by the customer seems to imply that the informant is concealing information if not prompted by a direct and detailed question). In particular, *IMP* has been exemplified in the literature (Jameson 1983a) as applied to the rental of furnished rooms. *IMP* is well–informed about the various properties of moderately priced rooms offered for rent in Hamburg, having been supplied in advance with an evaluation form (a data–structure organized hierarchically); besides, it behaves as if it believed that the customer, too, is well–informed and possesses the same information, against which he or she assesses the qualities of the room considered. Out of the various heuristics of question–answering included in the pragmatic knowledge embodied in *IMP*, here is a representative example: if *IMP* is asked a specific question abut one of the properties of the room – that the program is able to compare with the average (stated in the evaluation form it has) for the various classes of rooms on offer for rent in the same city – *IMP* answers this question directly, but if the program expects that some element in that answer would produce a (possibly slight) downward impression shift, then the program itself adds, if possible, a (true) claim that is meant to shift impression back

upward.

For a future upgraded version of *ALIBI*, the task of tuning impression and selecting information would be more difficult, because the database of facts imputed is not as static as an evaluation form with data on cooking facilities or noise in a room. *Planning* concerns no longer just shaping the final utterance according to pragmatic knowledge, but trying to convince about a plan of action, according to those objective "symptoms" upon which the input accusation, that interprets them according to a reconstructed criminally motivated plan, is based.

13. Plan Recognition in Natural-Language Processing

Plan recognition plays an important role in natural-language processing, for example, in order to allow a natural-language interface to answer questions appropriately, according to:

- the pragmatics of discourse,
- a general model of the user, and
- a specific model of the user constructed according to episodic memory.

A short introduction to plan recognition in the framework of the pragmatics of interaction is provided by Webber (1986 a: Sec. 2). On modeling users for the purposes of subserving the process of generating explanation for the output an expert system provides, see, for example, Kass and Finin 1988; McKeown 1988. Another aspect of the involvement of plan recognition in natural-language processing is the role of plans in analyzing multi-paragraph action-stories (cf. Dyer 1983).

Upgrading *ALIBI* to involve *planning* and *counterplanning* on the part of both defence and prosecution requires an explicit representation of plans intended to achieve certain (sub)goals, and of what may thwart plans. In *POLITICS*, a program of the Yale school, Carbonell (1981) included *scripts* (sequences of plans or events) augmented with *goals*. Planning involves also *counterplanning*, and Carbonell (*ibid.*: pp. 280-283) defines, and describes in graphs, two kinds thereof:

- *obstructive counterplanning:* actor *X* attempts to block actor *Y* from achieving Y's goal *G(Y)* with plan *P/G(Y)*;
- *constructive counterplanning:* actor *X* tries to achive his goal *G(X)* in spite of attempts (potential or actual) by actor *Y* to thwart *X*'s plans for pursuing *G(X)*, or to directly nullify the goal state *G(X)*.

Also in the framework of the Yale approach to natural-language processing, Dyer (1983) has defined *thematic abstraction units (TAUs)*, that are generalized patterns of action in terms of goals and plans. A TAU is a memory construct that serves as a basis for story characterization by means of a morale, sometimes corresponding to an adage; this is useful, as it allows to store cross-contextual episodes, and, *"once an episode has been indexed under a TAU, its recovery/avoidance heuristics become available for use in completely different situations."* (Dyer 1983: p. 32). Two TAUs, drawn from Dyer 1983: pp. 32, 41, follow:

TAU_HYPOCRISY

x is counter–planning against y
x is trying to get a higher authority z to
either block y's use of a plan P–1
(or to punish y for having used P–1)
by claiming that P–1 is an unethical plan
y claims that x has used an unethical plan P–2
similar to P–1
therefore, x's strategy fails
(adage: "The pot calling the kettle black.")

TAU_RED_HANDED

x is committing a violation V
of a contract (between x and y,
where y may be society)
x requires secrecy to achieve V
y directly ATTENDs (eyes or ears) to V
x suffers a goal failure
with respect to V
x and y now have an explicit
goal conflict (where y may
retaliate)
(adage: "Caught red–handed.")

where ATTEND is one of the primitive acts of the conceptual dependency theory, and indicates the act directing a sense organ; direct ATTENDing corresponds to witnessing. Sec. 3 in Dyer 1983 presents a general scheme for recognizing TAUs, whose processes are activated by goal failures, plan failures, or expectation failures (including false beliefs). The scheme is based on categorizing TAUs according to eleven planning metrics: *enablement* (e.g., plans may fail because of a false expectation), *cost* (sometimes it is prohibitive, or not commensurate to goals), *efficacy* (too weak measures can cause plans to fail), *risk* (too high risks may make a plan unfeasible), *coordination* (e.g., in *Romeo and Juliet*, Juliet tries to fool her opponents, but deceives Romeo, too), *availability* (for plans that consume a limited resource, or that impair future credibility, or where partial giving in does not appease an opponent), *legitimacy* (as in cases involving hypocrisy or of unjust retaliation), *affect* (as in the case of plan–failure due to exceeding cautiousness), *skill* (as in the case of plan–failure due to inexperience), *vulnerability* (as in the case of causing a greater harm), and *liability* (as in the case of an implicit contract stemming from the adoption of a given plan). TAUs, as well as some other approaches of the Yale school, including TOPs (Schank 1982 b,a) are based on Schank's assumption (1981) that episodic memory is organized around *failures*, not *successes* of plans.

Wilensky (1983: Sec. 4) has proposed a scheme for *meta–planning*: a meta–level planner should embody knowledge on how to select or coordinate plans. The meta–planner reasons in terms of meta–goals and meta–plans, and possesses particular meta–goals under situations termed *meta–themes*. For example, because of the meta–theme ACHIEVE AS MANY GOALS AS POSSIBLE, the meta–planner would have the meta–goal of resolving a goal conflict whenever it comes into being (*ibid.*: p. 30). Other meta–themes Wilensky lists are: DON'T WASTE RESOURCES, MAXIMIZE THE VALUE OF THE GOALS ACHIEVED, and AVOID IMPOSSIBLE GOALS. The latter causes a *Goal Detector* to initiate a *RESOLVE–CIRCULAR–GOALS* meta–goal for goal G1, and the *REPLAN* meta–plan would try to choose another plan that would achieve G1 without spawning circular subgoals detected (*ibid.*: p. 47). Wilensky describes a collection of programs called *PAM* (for *Plan Applier Mechanism*); *PAM* analyzes story texts, tries to find explanations for the events described there, answers questions about the text, and paraphrases the text from the viewpoint of the different characters involved in the text. Let us focus on the latter task; an example (*ibid.*: Sec. 5.3.) follows:

Input text: WILLA WAS HUNGRY.
SHE PICKED UP THE MICHELIN GUIDE
AND GOT INTO HER CAR.

(Various user's questions, and the respective answers of the program, are omitted)

Output: THE STORY FROM WILLA'S POINT OF VIEW:
I WANTED TO GET SOMETHING TO EAT, BUT I DIDN'T KNOW WHERE A RESTAURANT WAS. SO I PICKED UP THE MICHELIN GUIDE, AND I GOT INTO MY CAR.

PAM can handle also multi-paragraph stories (cf. p. 150).

Story-telling is another area of natural-language processing, and is also strongly related to research in planning and counterplanning. Yazdani (1983) proposes a model with five interacting processes: a *plot-maker*, a *world-maker*, a *simulator*, a *narrator*, and a *text-generator*. Yazdani's system *ROALD* corresponds to the simulator. *ROALD* has a representation of the world with the involved characters. It is given the general guideline of the plot. A planner is associated with each character, that plans the actions that the character would like to carry out in the world. The point is that those plans are thwarted (else, there would be no story). The simulator has a problem-solver, that receives the plans of all characters, and introduces *props*, *distractions*, and *accidents*, so that the actions of other characters are thwarted. *ROALD*'s overall goal is to make sure that the simulation follows the path required by the input plot (*ibid.*; cf. Yazdani 1982 a,b,c). An earlier system, *TALE-SPIN*, described by Meehan (1977, 1976), was a simulator allowing for only one active agent in the world of stories. Another story-teller is described by Lebowitz (1983, 1985).

Available technology, as shortly surveyed in this section, would allow to construct enhanced versions of *ALIBI*, accounting for planning aspects as occurring in action stories. These technical resources were not employed in the construction of the prototype of *ALIBI*, that, instead, focuses on the decomposition and recomposition of actions.

In particular, a particular structure devised by the Yale school, *Memory Organization Packets (MOPs)*, seems suitable for application and adaptation, for the purposes of allowing the program to evaluate how *credible* an alibi is: a MOP describes a standard compound action and can be represented by the union of

- a directed graph
  - whose vertices are a set of goals (of the various roles involved), plans, and possibly events; and
  - whose arcs, termed *intentional links*, bear labels that identify their kind: subsets of specific kinds of arc can occur between ordered pairs of vertices of given kinds; for example, event vertices may *motivate*, or *achieve*, or *thwart* a goal vertex; some arcs may point to elements in other MOPs, to avoid overspecific repetitions;
- a set of items, such as objects or information involved.

MOPs were defined by Schank (1982 a). A good exposition of MOPs can be found also in Dyer 1983. A standard sequence of plans is assumed by default to occur, but

exceptions considered possible and not unlikely to occur can be also represented. For the purposes of the task of *ALIBI*, the representation should be augmented with a representation of the likelihood of exceptions, omitting those considered unlikely beyond a certain threshold. A scoring formalism would compose scores of credibility for an alibi while its construction is resorting to a set of MOPs and other, more elementary, disaggregated actions or events.

14. Planning as Involved in Local Projects other than 'Alibi'

Another program, co-developed by one of the authors, is also concerned with planning human actions: the *BASKETBALL* expert system (Nissan, Simhon and Zigdon 1989), an adviser for a basketball coach, tries to predict the strategy of the rival team (according to data on the human resources of the latter), and then proposes a strategy for the coach's own team and recommends those players that belong to the team and that it would be better to have playing as the opening five in the considered match. *BASKETBALL* is a stand-alone system, but it fits in the framework of *The Sport Kit*, that includes expert assistants for training or practice in various sports, implemented by Nissan's students under his direction. The prototype of the *TENNIS* expert adviser - simpler than *BASKETBALL* - is intended to advise a player, by considering resources of that player and of his adversary. No adversary player is involved, instead, in the task of the *SKINDIVING* expert system (Aviel and Nissan 1989), an adviser for going skindiving that takes into account the sea and weather conditions, the season, hour, moon phase, as well as the user's skills. Anyway, the architectural and programming approach adopted in *The Sport Kit* (coded in the *C* programming language) is very different from our approach in *ALIBI* , where a semantic and pragmatic description of actions is exploited by means of the *Prolog* built-in inference engine.

Nissan (in Schmidt) deals - among other things - with the representation of etymological conjectures (*not* with automated hypothesis-generation). Strictly speaking, there is no predefined "plan", but patterns of morphological, phonological, phonetic, and semantic change could be considered as events composed in a hypothesized sequence; sequences of events can be, sometimes, handled like sequences of plans (cf. Dyer 1983). A loose relation to the concerns of *ALIBI* can be perceived, because a shared goal is the reconstructions of past events or plans according to a set of ascertained facts underlying an accusation, while only one interpretation corresponds to what really happened, also only one possible etymological conjecture is true, while often other etymologies or folk-etymologies can be hypothesized, too.

15. Artificial Intelligence Systems and Deontic Applications

We have knowledge of no program with the same task as *ALIBI*; however, computing is not new to deontic applications. The following refers to projects and approaches described in the literature.

- We are aware of two other projects that apply computing to *moral problems,* each project in its own way:
  - Thomas Harold Murray (1985) has developed a computational model of unethical behavior, involving biographic simulation.

* Here in Israel, Yehuda Eisenberg (1987) of Bar-Ilan University has developed, in 1987/8, a menu-based interactive educational program that proposes a story with a moral dilemma, and the considerations and viewpoints involved, to children according to their level of awareness.

- On the other hand, *the law*, that used to be an area of application for information retrieval, is now a thriving sector of application of expert systems. Cf. Moulin and Rousseau (1989); Gardner 91985). Specialized conferences are held (Franson and Smith 1989). In particular, on logic programming for the law, see Sharpe 1985. The paper Sergot *et al.* 1986 exposes a formalized representation in *Prolog* of the British Nationality Act 1981, which defines British citizenship. Stamper (1980) deals with modelling legal rules in *LEGOL*, a logic-based programming language specialized for legal applications. Gordon (1985) exposes a logic-based object-oriented approach.

16. Conclusions

We have exposed *ALIBI*, an artificial intelligence program whose task is to find excuses for a defendant. Events as stated in the input accusation are separated from deontic interpretations, and are integrated in the framework of broader plans, motivated by goals different from the illegal ones assumed in the accusation. Sometimes the same plans are motivated by different goals, while sometimes fortuitous events can be invoked as an excuse. We have discussed also several extensions that are called for. The model embodied in *ALIBI* can be refined, either by adding components, or by replacing components with more sophisticated versions. Assumptions the model is based upon fall into two categories:

- assumptions on the universe modelled, and
- assumptions adopted to reduce specifications of the domain modelled to a currently manageable degree of difficulty. This second kind of assumptions has to be revised as the system is upgraded.

ALIBI differs from legal applications of computing because it focuses not so much on *regulations*, as on theories of *action*, and on the common-sense pragmatics of *social behavior*. While *ALIBI* is of interest for models of *argumentation*, more generally the topic is interesting for the synthesis of *explanations*, as well as for *abduction* in artificial intelligence or human cognition: *abduction* is like interpreting partial symptoms to reach higher abstraction levels, which is the opposite of *deduction* under certain respects.

Acknowledgements

ALIBI was developed under *UNIX*™ in *Edinburgh Prolog*, on a *VAX*™ computer at the Computing Laboratory of the Department of Mathematics and Computer Science of Ben-Gurion University.

The first version of the program was implemented during the summer of 1987, by Kuflik and Puni under the direction of Nissan; since then, the latter developed the program furtherly. The work originally carried out in the framework of *Expert Systems* and *Prolog* courses taught by Nissan. Extensions were then specified and partly designed by him and by Kuflik in 1988.

References

Alterman, R. 1985 *A Dictionary Based on Concept Coherence. Artificial Intelligence* 25,2: p. 153–186.

Aviel, A. and E. Nissan 1989 *Predictive Advice in the SKINDIVING Expert System*. Technical Report, Ben-Gurion University (Beer-Sheva).

Barr, A. and E. A. Feigenbaum 1981 *The Handbook of Artificial Intelligence*, Vol. 1. HeurisTech Press (Stanford, California) and W. Kaufmann (Los Altos, California); presently: Morgan Kaufmann (Los Altos, California).

Brodie, L. and J. Mylopoulos, Eds. 1986 *On Knowledge Base Management Systems: Integrating Artificial Intelligence and Database Technologies*. Springer-Verlag (Berlin).

Burger, H. J.1984 *The Wordtree: A Transitive Cladistic for Solving Physical and Social Problems*. The Wordtree (10876 Bradshaw, Overland Park, Kansas).

Carbonell, J. 1981 *POLITICS* (Sec. 11: pp. 259–307) and *Micro POLITICS* (Sec. 12: pp. 308–317). In: Schank, R. C. and C. K. Riesbeck, Eds., *Inside Computer Understanding: Five programs Plus Miniatures*. Lawrence Erblaum Associates (Hillsdale, New Jersey).

D'Amore B. 1976 *Applicazione della logica proposizionale alla teoria dei giochi*. Appendix in: D'Amore, B., *Elementi di teoria dei giochi*. Collana di Matematica Moderna, 12. Zanichelli (Bologna): p. 142–157.

Dyer, M.G. 1983 *In-Depth Understading: A Computer Model of Integrated Processing of Narrative Comprehension*. The MIT Press (Cambridge, Massachusetts).

Dyer, M.G., Flowers, M. and J. F. Reeves 1989 *A Computer Model of Irony Recognition in Narrative Understanding*. In: Nissan, E., Ed., *Advances in Computing and the Humanities, Vol. 1: Language and Speech*. JAI Press (Greenwich, Connecticut).

Flowers, M., McGuire R. and L. Birnbaum 1982 *Adversary Arguments and the Logic of Personal Attacks*. In Lehnert and Ringle 1982: pp. 275–294.

Franson, R.T. and J.C. Smith, Eds. 1989 *Proceedings of the Second International Conference on Artificial Intelligence and Law*, University of British Columbia, Vancouver (Canada), June 13–16, 1989.ACM SIGART.

Eisenberg, Y. 1987 *The Computer as Managing a Dialog on Moral Dilemas*. Technical Report, Integrated Division of Jewish Studies, Bar-Ilan University (Ramat-Gan, Israel). In Hebrew.

Gardner, A.V.D.L. 1985 *Overview of an Artificial Intelligence Approach to Legal Reasoning*. In Walter 1985: pp. 247–274.

Gordon, T.F. 1985 *Object-Oriented Predicate Logic and Its Role in Representing Legal Knowledge*. In Walter 1985: pp. 163–203.

Grice, H.P. 1975 *Logic and Conversation*. In: Cole, P. and J. L. Morgan, Eds., *Syntax and Semantics*. Academic Press (Orlando, Florida).

Hamburger, H. 1988 *Natural Language and Expert Systems*. A special issue of the *International Journal of Expert Systems: Research and Applications* 1,4. JAI Press (Greenwich, Connecticut).

Harris, M. D. 1985 *Introduction to Natural Language Processing*. Reston Publ. Co. (Reston, Virginia).

Jameson, A. 1983 *Impression Monitoring in Evaluation-Oriented Dialog. The Role of the Listener's Assumed Expectations and Values in the Generation of Informative Statements. Proceedings of the Eighth International Joint Conference of Artificial Intelligence (IJCAI'83)*, Karlsruhe, West Germany. Morgan Kaufmann Publ. (Palo Alto & Los Altos, California): Vol. 2, pp. 616–620.

Joshi, A.K. 1982 *Mutual Beliefs in Question Answering Systems*. In: Smith, N. Ed., *Mutual Belief*. Academic Press (Orlando, Florida).

Kass, R. and T. Finin 1988 *The Need for user Models in Generating Expert System Explanation*. In Hamburger 1988: pp. 345–375.

La Settimana Enigmistica 1973 Anno 42, No. 2128 (Milano, 6 gennaio 1973): pag. 4 *(Leggendo qua e là...,* 19553).

Lebowitz, M. 1983 *Creating a Story-Telling Universe. Proceedings of the 8th International Joint Conference on Artificial Intelligence (IJCAI'83),* Karlsruhe, West Germany. Morgan Kaufmann Publ. (Palo Alto, California): pp. 63–65.

Lebowitz, M. 1985 *Story Telling as Planning and Learning.* Technical report CUCS-202-85, Department of Computer Science, Columbia University (New York).

Lehnert, W. and M. Ringle, 1982 *Strategies for Natural Language Processing.* Lawrence Erlbaum Associates (Hillsdale, New Jersey).

Lifschitz, V. 1987 *Formal Theories of Action.* In: *The Frame Problem, Proceedings of the 1987 Workshop:* pp. 35–57.

Lifschitz, V. and A. Rabinov 1989 *Miracles in Formal Theories of Action. Artificial Intelligence* 38,2: pp. 225–237.

McKeown, K.R. 1988 *Generating Goal-Oriented Explanations.* In Hamburger 1988: pp. 377–395.

Meehan, J. R. 1976 The *METANOVEL: Writing Stories by Computer.* Ph. D. Thesis, Technical Report 74, Department of Computer Science, Yale University (New Haven, Connecticut).

Meehan, J. R. 1977 *TALE-SPIN: An Interactive Program that Writes Stories. Proceedings of the 5th International Joint Conference on Artificial Intelligence (IJCAI'77).* Morgan Kaufmann Publ. (Palo Alto, California).

Moulin, B. and D. Rousseau 1989 *Design of Deontic Knowledge-Bases of Querying Texts Dealing With Regulations. Proceeding of the Ninth International Workshop on Expert Systems and Their Applications (General Conference),* Avignon (France), May 29 – June 2, 1989. EC2 (Nanterre, France).

Murray, T.H. 1985 *Unethical Behavior: A Model Based on Personality Theory and Validated by Biographic Simulation.* Ph. D. Thesis, The American University (University Microfilms International, NDG86-12443).

Nilsson, N.J. 1982 *Principles of Artificial Intelligence.* Springer-Verlag (Heidelber & Berlin) and Tioga, presently Morgan Kaufmann (Palo Alto, California).

Nissan, E. (to appear) *Structure and Representation for a Multilingual Semitic Machine-Dictionary: Issues in Engineering a Lexicon Meant for Symbolic Manipulations.* In: Schmidt, K.M., Ed. *Collected Articles and Papers on Conceptual and Content Analysis by Computer.* Two thematic volumes in *Advances in Computing and the Humanities.* JAI Press (Greenwich, Massachusetts).

Nissan, E., Simhon, D. and N. Zigdon 1989 *Resource Evaluation and Counterplanning With Multiple-Layer Rulesets,* in the *BASKETBALL* Expert System. Technical Report, Ben-Gurion University (Beer-Sheva).

Schank, R. C. 1981 *Failure-Driven Memory. Cognition and Brain Theory* 1.

Schank, R. C. 1982a *Reminding and Memory Organization: An Introduction to MOPs.* In Lehnert and Ringle 1982.

Schank, R. C. 1982 *Dynamic Memory: A Theory of Reminding and Learning in Computer and People.* Cambridge University Press (Cambridge, U.K. & New York).

Schubert, L. K., Goebel, and J. J. Cercone 1979 *The Structure and Organization of a Semantic Net for Comprehension and Inference.* In: Findler, N.V., Ed., *Associative Networks: Representation and Use of Knowledge by Computers.* Academic Press (Orlando, Florida): pp. 121–175.

Sergot, M.J., Sadri, F., Kowalski, R. A., Kriwaczek, F., Hammond, P. and H.T. Cory 1986 *The British Nationality Act as a Logic Program. Communications of the ACM* 29,5: pp. 370–386.

Sharpe, W. P. 1985 *Logic Programming for the Law.* In: Braines, M.A., Ed., *Research and Development in Expert Systems. Proceedings of the Fourth Technical Conference of the British Computer Society Group on Expert Systems,* Warwick. Cambridge University Press (Cambridge, U.K.).

Stamper, R. 1980 *LEGOL: Modelling Legal Rules by Computer*. In: Niblett, B., Ed., *Computer Science and the Law*.Cambridge University Press (Cambridge, U.K.): pp. 45–71.

Walter, C., Ed., 1985 *Computing Power and Legal Reasoning*. West (St. Paul, Minnesota).

Webber, B. W. 1986a *Natural Language Processing: A Survey*. In (Brodie and Mylopoulos 1986): pp. 353–363.

Webber, B. W. 1986b *Questions, Answers and Responses: Interacting with Knowledge-Base Systems*. In Brodie and Mylopoulos 1986: pp. 365–402.

Wilensky, R. 1983 *Planning and Understanding: A Computational Approach to Human Reasoning*. Addison-Wesley (Reading, Massachusetts).

Yazdani, M. 1982a *How to Write a Story. Proceedings of the ECAI'82 European Conference on Artificial Intelligence*, Orsay, France. Also: Research Report R. 104, Department of Computer Science, University of Exeter, U.K.

Yazdani, M. 1982b *Story-Writing by Computer*. Research Report R. 106, Department of Computer Science, University of Exeter, U.K.

Yazdani, M. 1982c *Generating Events in the Fictional World of Stories*. Research Report R. 113, Department of Computer Science, University of Exeter, U.K.

Yazdani, M. 1983 *Language Generation for Story Telling*. In: Rösner, D., Ed., Abstract kit of the *International Workshop on Language Generation,* Burg Stettenfels, West Germany. Report, Projekt *SEMSYN*, Institut für Informatik, Universität Stuttgart, West Germany.

António J. R. P. Ferreira
Lisboa (Portugal)

Atitudes Musicais e Metáforas Computacionais: uma visão pessoal

Resumo

Partindo da idéia da metáfora como geradora de motivação e entusiasmo, observa-se a predisposição histórica da música à aplicação de conceitos inspirados nos formalismos computacionais. As relações entre a ideia de gramática e a música como linguagem. Os processos de decisão na composição musical e a pesquisa em IA. Conclui-se com uma apresentação das atitudes de dois compositores perante a máquina e com uma breve apreciação final sobre o conceito de criatividade e a sua possível cristalização formal.

Abstract

Starting from the idea of metaphor as origin of motivation and enthusiasm, I look, in a brief way at the historical predisposition of music in being influenced by computer formalisms. Also to the relationship between grammars and music, decision process-making in musical composition and related, ending with a brief judgement about the concept of creativity and possible formalizations.

Introdução

Muitos de nós ouvimos música todos os dias. Além do omnipresente equipamento reprodutor/produtor de sons, existente num grande número de habitações, somos banhados permanentemente por uma paisagem sonora resultante da actividade humana, quer esta seja de 'inspiração' musical ou não.

Na interpretação deste ambiente sonoro servimo-nos muitas vezes de certas analogias retiradas ao vocabulário normalmente usado noutros campos. Para as ideias que aqui vou apresentar é pertinente realçar que as apreciações tecidas ao longo do texto são feitas do ponto de vista da composição musical. Reflectem portanto a minha prática e as minhas influências. Nesta base, definamos as metáforas que, na minha opinião, provocaram determinadas atitudes por parte de compositores contemporâneos.

Aristóteles, citado por Paul Ricoeur, afirmava que uma metáfora é a aplicação a uma coisa de um nome que pertence a outro. A metáfora seria assim uma figura de retórica que, na sua forma 'ampliada', leva à comparação. Paul Ricoeur expande estes conceitos, falando de metáfora de tensão e de substituição, concluindo que não é um ornamento do discurso e que para lá do valor emotivo, ela diz-nos algo de novo acerca da realidade (Ricoeur 1976, 1987).

Esta realidade é representada, neste meu texto, pelos conceitos científicos e técnicos que acabam, por percolação, a influênciar as intenções dos compositores.

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 511/518
1992, APDC & SPF, Lisboa

Assim, por metáfora, eu quero designar uma forma significante, onde poderei acomodar noções ainda vagas.

Ora a música, um campo de intensa actividade humana, com uma larga mas informal base teórica e com critérios para apreciações, é fértil na criação de resultados audíveis, resultantes do 'transplante' de várias metáforas usadas noutros campos. Na área à qual eu estou a restringir estas considerações, a da música ocidental, o aparecimento dos computadores, máquinas resultantes do pensamento matemático e o consequente conceito de programa, chamaram inevitavelmente a atenção para a estrutura dos processos de decisão musical, quer em composição, quer em interpretação.

Ideias e metáforas usadas na apreciação das capacidades cognitivas do ser humano vão, pela via das tentativas de sua modelização em máquinas programáveis, fascinar os compositores. Para muitos o interesse reside na ampliação dos métodos tradicionais de composição (maior controle na produção sonora, audição imediata de resultados). Outros vão partir de bases matemáticas e de sistemas formais, interessados na capacidade do computador em 'criar' estruturas que valem por si e isto fora de qualquer utilidade específica, manifestações poéticas de máquinas toscas (François 1982).

É sobre as condições iniciais desta sedução discreta que tratarei a seguir.

O convite à viagem

Num artigo recente, Luiz Moniz Pereira fala de como a descontinuidade entre o Homem e a Máquina, ponto de diferenciação e de referência, se encontra em vias de atenuação.

De facto, o mundo tecnológico que nos rodeia está limitado, por questões práticas, pela natureza (limite de recursos, ligação imediata da actividade técnica às suas consequências). A recusa da reconciliação, no próprio acto técnico, da negação e afirmação da natureza (Roqueplo 1982) será o radicalizar de posições e a impossibilidade da aceitação harmoniosa de um mundo industrializado (Pereira 1988). Historicamente falando, a metáfora da máquina na explicação do ser humano teve, a partir do séc. XVI, um grande impacto nas imaginações. Inspirada nos progressos em matemática e mecânica, a arte de realizar autómatos, utilizando mecanismos complexos de relojoaria, atinge o auge no séc. XVIII com Vaucanson. Curiosamente um dos seus autómatos mais célebres era um intérprete de flauta <travessa> que executava as melodias de um modo análogo ao de um intérprete real. Talvez não seja de grande espanto, pois a música, desde o tempo de Pitágoras, sempre se baseou, consciente ou inconscientemente, nas noções matemáticas de simetria, série e grupo.

Também o facto de dispor de uma linguagem 'discreta' que codificava mais ou menos a continuidade sonora (o caso da notação) a predispunha a possibilidades de manipulação mais abstractas.

E eis que, no final do mesmo século, J. J. Hummel publica o seu livro intitulado "Instruções em Quatro Línguas sobre como Compor Valsas e Contradanças com Um Par de Dados" (Hummel 1973). A sua técnica consistia em fornecer um certo número de motivos musicais, que eram escolhidos por lançamento de dados e ordenados segundo as regras fornecidas. O primórdio da composição algorítmica.

É assim relativamente fácil de compreender que as primeiras tentativas de 'composição' utilizando o computador surjam na mesma altura do seu aparecimento neste século (suite ILLIAC, Hiller 1956).

Se nos séculos anteriores os 'feitos' dos autómatos levaram a pensar que não haveria

limites às possibilidades da mecânica, depressa se verificou que tal não era verdade. No entanto, um programa actual de computador poderia ser idealizado por um 'engenho de computação' mecânico construido com rodas dentadas. É, claro está, muito mais fácil construir um programa num terminal usando uma linguagem de programação. O computador, isto é, a máquina física onde o programa correrá, é então uma espécie de fonte de energia para a execução do mesmo.

Se atendermos ao facto de que sempre existiram alguns compositores que, num sentimento comum de abstracção, tentaram basear os seus métodos composicionais em conceitos matemáticos - ver o caso de Joseph Schillinger (1895-1943) que baseou o seu sistema de composição na lógica matemática e no sistema de coordenadas cartesianas ou então o exemplo mais conhecido de Iannis Xenakis - não será de espantar que tenha havido uma imensa curiosidade em tentar representar o conceito que cada compositor tem do seu trabalho na máquina virtual que é o computador. O convite está feito. Avancemos pois no tempo, ultrapassando as primeiras experiências (Barbaud, Phillipot, G.M. Koenig) e olhando para uma metáfora importante para a música e para a própria ciência da computação: a linguagem natural.

Linguagem como metáfora para a música

A ideia de pensar a música em termos de uma gramática musical não é propriamente nova. Será no entanto nos anos 50, a idade de 'ouro' da sintaxe nas pesquisas em linguística e especialmente com o trabalho de N. Chomsky "Syntatic Structures" (1957) que, paralelamente ao aparecimento de modelos computacionais para a compreensão da linguagem natural, a sua aplicação ao campo musical começa a despertar interesse.

Aqui, não estarei muito preocupado em apreciar o facto de a música ser ou não uma 'linguagem'. O próprio Chomsky numa conferência na Universidade de Stanford em 1979 respondeu à mesma questão dizendo que tudo dependeria das definições de cada um (citado em C. Roads 1985). Mais concretamente, apontarei as semelhanças dos problemas e soluções que esta aproximação à música tem no processamento da linguagem natural por meios computacionais.

Assim, vejamos que aplicações da teoria linguística à música resultaram desta analogia (o que se segue é referido em Roads 1985):

- Análise sonora empregando instrumentos teóricos da fonologia.
- Teorias cognitivas da música, em que se podem mencionar várias atitudes: considerar a música como uma prática mais ou menos directamente relacionada com a linguagem falada, ponto defendido por L. Bernstein que relacionou as letras do alfabeto a notas, substantivo a melodia, verbo a ritmo e adjectivo a harmonia (Bernstein 1976), usar linguagens formais para construir uma teoria cognitiva da audição musical (Laske 1975) e usar uma metodologia linguística (modelos generativos) aplicada à análise musical (Lerdahl e Jackendoff 1977 e 1983).
- Por fim, o campo de interesse para aqui, a composição musical em si.

Neste último caso, considera-se a estrutura musical como possível de ser caracterizada em termos de expressões abstractas usando linguagens formais (Jones 1981) ou, mais concretamente, assenta no considerar a música como um sistema com as suas regras sintácticas.

As gramáticas são assim criadas com o objectivo de gerar estruturas musicais e podem ser expressas por regras de reescrita mais ou menos complexas. Geradores de melodias ou motivos podem ser construídos de uma maneira análoga a geradores de frases de

linguagem natural. O processo pode ser expandido: examinando a estrutura de uma gramática formal - um grupo de símbolos terminais e não terminais, um conjunto de regras de reescrita, que especificam as combinações dos símbolos e um símbolo que inicie o processo generativo - podemos associar uma probabilidade a cada escolha das possibilidades de geração. Esta aproximação foi usada em composição por K. Jones (1981), e por M. Ligabue na criação de um sistema de regras que modelizassem a improvisação no Jazz (Ligabue 1986).

Mas, e em paralelo com o estudo do processamento da linguagem natural, torna-se evidente que uma aproximação puramente sintáctica é extremamente limitada. Apesar da música se nos apresentar, por via da sua notação, de uma maneira segmentada e 'discreta' a verdade é que o som é um sinal muito mais 'denso' e multidimensional. O problema da semântica na música foi posto por M. Minsky no seu artigo "Music, Mind and Meaning", que, à pergunta: "porque é que nós gostamos de certas melodias?", contrasta duas respostas, normalmente apresentadas como justificativas:

- elas contêm certas características estruturais, isto apesar de não existir uma sintaxe clara que nos permita interpretá-las da mesma maneira como fazemos quando da apreciação da correção sintáctica de uma frase falada em linguagem natural.

- elas assemelham-se a outras melodias de que nós gostamos, e aqui entra a problemática da representação de conceitos ou 'melodias' no cérebro de cada um de nós.

Conclui ele que métodos generativos são só significativos para usos analíticos simples: "... o objectivo de teorias músicais orientadas para a sintaxe segue uma direcção errada, pois elas aspiram a descrever as frases produzidas pelos nossos cérebros, sem tentarem descrever como essas frases são produzidas. Significado ('meaning') é muito mais do que a estrutura das frases" (Minsky 1981).

Para quem compõe, utilizando a capacidade do computador de fazer 'viver' estruturas abstractas, este é um ponto a ter em consideração, apesar de em composição eu considerar que tudo é útil. O aspecto multidimensional da linguagem natural, que tal como nas primeiras experiências em música, não recebeu grande atenção, levou a que se passasse de uma fase estatística/numérica nos anos 50, para outra baseada na sintaxe, acabando por ser reconhecida a necessidade de recorrer a formulações das representações do conhecimento implícito em cada um de nós. Isto gerou um fértil campo de pesquisa em ciência de computação, no qual se empregaram sistematicamente técnicas e filosofias do que se convencionou chamar Inteligência Artificial. As insuficiências acima citadas fazem com que alguns olhem a música como zona de eleição para a aplicação das mesmas técnicas.

Assim, Mira Balaban considera o meio musical como uma área típica para o estudo de IA - a actividade musical como actividade humana à qual falta uma teoria completa e devidamente organizada (Balaban 1985). Uma nova metáfora, referente ao modo como o ser humano emprega os seus conhecimentos, resultou em maneiras mais ou menos sofisticadas de resolver o problema de uma inteligência musical que, de acordo com Otto Laske, é distinta de outras inteligências e não é derivada de competências linguísticas ou lógico-matemáticas (Otto Laske 1988). É o que me proponho observar em seguida.

"O Homem é a medida de todas as coisas"

Do esforço empregue na tentativa de modelização das fontes de 'energia' usadas pelos seres humanos na resolução de situações e problemas, tornou-se evidente que a flexibilidade humana no uso de diferentes formas de representação, com o fim de observar um problema segundo perspectivas diversas, é difícilmente codificável nas linguagens

formais usadas correntemente em programação.

A actuação mais comum é, e de acordo com as características da situação a resolver (complexidade, predictabilidade, densidade de soluções), aplicar varias estratégias que façam sentido em relação a essa mesma situação (dividindo o problema em subproblemas, usando e criando regras de produção, representando analogias, impondo restrições, etc.). Estes métodos, usados regularmente em IA, implicam que se considere a composição musical como um processo em que se tentam resolver problemas o que, no caso da música tonal, significa a procura, no vasto espaço de 'soluções', das estruturas que satisfaçam as restrições definidas pela pessoa que compõe (Steels 1986).

Não há, no entanto, um único 'processo criativo'. Tal como já indiquei acima, é bem possível que a estratégia criativa de alguém envolvido na música seja diferente de, por exemplo, um homem de ciência; objectos, horizontes e fins são diversos (Hargreaves 1986). Dentro da própria música é sabida a dificuldade de pianistas com uma educação clássica em 'aprender' o tipo de improvisação usado no Jazz.

Assim,a ideia de representações que permitam pontos de vista múltiplos, proporcionando uma melhor consideração do contexto (musical ou outro), ganhou terreno, criando em algumas pessoas a vontade de explorar os processos mentais, na maior parte inconscientes, envolvidos na composição musical.

No seu artigo "Learning the craft of musical composition" (1986) Luc Steels considera duas formas empregues na resolução dos problemas que se deparam ao longo do processo de composição:

- uma é considerar que o compositor parte de princípios fundamentais e explora o espaço de estados usando técnicas 'fracas' (tentativa-e-erro, exames sistemáticos até à solução, método 'subida da colina', etc.).
- a outra forma consiste em usar tácticas menos 'brutais', recorrendo a heurísticas (estas designam o conjunto de métodos díspares e mal definidos, postos em jogo pelos seres humanos ao 'activarem' os processos de descoberta e aprendizagem; podem, eventualmente, ser verbalizadas e, portanto, formalizadas).

As restrições impostas à progressão dos acordes pela teoria da harmonia clássica (não existência de oitavas ou quintas paralelas, movimentação de tons respeitando distâncias mínimas) são exemplos de regras que podem ser exploradas pelos métodos 'fracos'.

A outra atitude toma, normalmente, a forma de regras de produção "se...então...", usadas para explicitar uma dada base de conhecimentos. Usando estes métodos, criaram-se assim sistemas periciais que tentam ilustrar o processo de composição de música tonal, caso de L. Steels, ou que se prendem à resolução de problemas aparentemente 'acessíveis', como a harmonização de corais de Bach (Ebcioglu 1988, Marilyn Thomas 1985, Lischka 1986).

Este último parte do conceito de que se existe um sistema com restrições, então existe uma aplicação musical. O formalismo das restrições pode ser visualizado como uma rede de processos ligados entre si por 'fios', onde circulam as informações. Um processo actua somente na informação ao seu dispor e 'emite' novos resultados pelos 'fios' a ele ligados (Roads 1985). Processos qualitativos podem também ser representados desta maneira 'distribuida'.

A resolução do 'problema' (a harmonização correcta) é assim associada à satisfação de determinadas restrições (de cuja escolha, diga-se, dependerá o sucesso do programa). Esta atitude deve muito à metáfora de um programa como uma sociedade em que comunicam peritos e especialistas, tal como foi formulada por Marvin Minsky.

Para uma atitude composicional propriamente dita, estas realizações despertam-me a

pergunta da sua relevância para o resultado sonoro final. Haverá algo de intrinsecamente fascinante em ouvir os resultados sonoros de modelos computacionais (gramáticas, álgebras, cadeias de Markov, sistemas periciais)? E que critério me levará a preferir um modelo formal a outro? Afirmei acima que quem compõe pode encontrar utilidade em qualquer coisa. Irei, seguidamente, mencionar duas atitudes de compositores, para depois falar um pouco mais da questão da relevância.

1. Paul Berg (Conservatório Real de Haia, Holanda) tem, nas suas próprias palavras "uma afinidade por restos e lixo (junk)". Na verdade, as suas atitudes reflectem um percurso pessoal e original, do qual poderemos ter uma ténue pista ao ouvirmos que "...este 'lixo' pode-se encontrar na minha casa e na minha música".

A sua posição é por ele resumida, ao afirmar que "não estou particularmente interessado em trabalhar com notas, escalas ou ritmos, mais interessante é a possibilidade de criar uma qualquer estrutura e em seguida ouvir o som dessa mesma estrutura."

É um pouco a estética do 'objecto encontrado', aplicada à composição musical, embora as estruturas que podem produzir esses 'objectos sonoros' (Scheaffer 1966) sejam perfeitamente determinadas. Isto ressalta nas afirmações de Herbert Brun, com o qual P. Berg estudou. Diz ele que "Uma coisa é a procura de acontecimentos (events) que produzam o som que nós queremos, outra, completamente diferente, é descobrir o som dos acontecimentos que nós desejamos" (Brun 1985).

A tal atitude o computador presta-se bastante bem e tanto um como o outro criaram as suas próprias linguagens formais de composição: no caso de Brun SAWDUST e PILE 4 no caso de Berg.

Esta última (criada em 1979 e 'defunta' em 1986) negava qualquer partitura, sendo desenhada como uma linguagem de controle de sistemas em tempo real. Era uma linguagem de composição muito próxima do computador (Berg 1979).

Nas notas de programa de uma das suas composições, "String Quartet", de 1985, Paul Berg reafirma a sua atitude: "... esta composição não é a expressão de emoções pessoais." A atitude musical é considerada assim como criadora de estruturas, sendo a música "...restos, poeira deixada pelo trabalho de composição" (Brun 1982).

2. George Lewis é um trombonista e compositor, nascido em Chicago (1952). Depois de ter estudado filosofia (Yale) e de ter sido director da programação musical do Kitchen Center em New York (80-83), colaborou e colabora em inúmeros projectos de vários músicos (S. Lacy, A. Braxton, J. Zorn). Desde o seu primeiro contacto com computadores (1978) esteve sempre interessado em programas interactivos. Aqui entra a sua história de improvisador (a sua relação com a AACM de Chicago, grupo de músicos dedicados à improvisação) que o levou a apreciar, do ponto de vista do 'performer', a interacção entre os músicos aquando de improvisações colectivas. A sua escolha de pequenos sistemas portatéis (Apple II e ou Atari) prende-se com a sua opinião de que assim "... uma pessoa não está dependente de uma instituição, nem precisa de quantidades desmesuradas de dinheiro... portabilidade é outra das vantagens, especialmente para quem executa as suas próprias composições. Uma grande quantidade de pessoas acaba assim por ouvir a minha música" (Lewis 1983).

As suas composições, envolvendo computadores, tem como característica a utilização de programas que modelem o processo de audição que ocorre entre músicos numa improvisação: "... eu não gosto de dar concertos de trombone solo. Prefiro tocar com mais alguém. Eu queria fazer um programa para ver que espécie de coisas um improvisador tem

de saber em ordem a tocar eficazmente com outro improvisador'' (G. Lewis entrevista com C. Roads 1983).

Para atingir sucesso nesta atitude, a estratégia chave deve residir, ainda segundo G. Lewis, na escuta selectiva de certos factores referentes ao contexto da própria música (''... bons improvisadores não podem ouvir tudo...''). Apesar da sugestão de que o estudo da improvisação musical poderá ser um campo de interesse para os investigadores de inteligência artificial (''... muitos aspectos do processamento humano de informação estão 'ao trabalho', numa base de decisões tomadas momento a momento, aquando da performance de música improvisada'' - Lewis 1983), os seus programas, implementados em FORTH e 'alimentados' por um sistema de conversão analógica-digital, com algum pré-processamento, recusam técnicas mais complexas como as que são usadas correntemente em IA. A sua aproximação ao problema da interacção músico/computador é, segundo as suas próprias palavras, ''cheap and dirty'' (G. Lewis citado em Roads 1986).

Considerações finais

Devo sublinhar, uma vez mais, que não quis aqui questionar a utilidade dos sistemas e teorias apresentadas, mas sim apreciar como elas são usadas. Por via das metáforas utilizadas por teóricos do conhecimento (M. Minsky e Howard Gardner por ex.) foi realçada a importância da música enquanto actividade humana, ''... precisamente porque ela comunica o desígnio ou propósito do ser humano, sendo, ao mesmo tempo, mais ambígua e potencialmente mais eloquente do que a linguagem ou as representações visuais'' (Machover 1986).

Essa actividade humana de criação musical encontra agora uma atracção poderosa na capacidade abstracta da máquina computacional. Tal como J. Baudrillard afirma no seu livro ''Amérique'' essa máquina torna-se, para o intelectual, o equivalente do walkman para o adolescente, uma ''videografia dos seus conceitos''. O conceito de criação, que, de uma maneira geral, pode ser considerado do ponto de vista do produto e do ponto de vista das características da pessoa que produz (ver Hargreaves 1986), encontra agora tentativas de formalização em sistemas periciais que procuram 'descobrir' novas informações. É o caso de EURISKO, escrito em 1980 por D.B. Lenat, que é guiado nas suas buscas por regras heurísticas, uma das quais, de uma maneira geral, diz ''para se olhar a extremos''. Estas regras são 'criadas' pelo próprio programa ao longo da sua evolução (Lenat 1984, 1988).

Mas, sistemas periciais que realizam a noção cartesiana de inteligência (concretizada nesses programas, pela separação entre base de conhecimentos e o formalismo que a manipula - o 'motor' de inferência), não são muito adequados a aplicações musicais. Na minha opinião, o processo criativo é contínuo e permanentemente 'sustido'', isto é, não creio que um músico tenha uma 'visão criativa' e depois a concretize unicamente aplicando regras, técnicas ou formulas. Estas 'emergem', de um modo contínuo, no decorrer do processo. A questão reside pois, em equiparar determinada concepção da actividade criativa à mecanização dessa mesma prática. Estas considerações levam compositores, que tendo realizado e utilizado tais sistemas, a considera-los mais como antagonistas amigáveis no processo de composição (Cope 1987). Se, nesta discussão, distinguirmos originalidade (como 'novidade', infrequência estatística) de criatividade (como resposta original que tem utilidade, no sentido mais lato da palavra - Hargreaves 1986), poderemos dizer que a relevância do uso de tais sistemas depende de um certo acto de fé por parte dos compositores: ''A idealidade 'computacional', como processo universal do pensamento, é

montada face ao indivíduo, como um guia, poder-se-ia dizer, moral" (Sfez 1988).

Argumentou-se que a característica do processo criativo se baseia não tanto na resolução de problemas, mas na descoberta de novos problemas. Isto reflecte-se na intencionalidade do acto, intencionalidade essa que é exterior ao programa (R. Born 1987), embora este seja pensado e executado em vista de determinados fins e objectivos.

Taft, citado em Hargreaves (1986), tentou aplicar a ideia de cognição 'quente' e 'fria' ao processo criativo. Resultavam assim dois estilos, um, mais pausado, baseado na ideia de resolução de problemas na sua aproximação ao desenvolvimento de novos conhecimentos, sendo o outro uma expressão mais emocional e comparativamente mais incontrolada. As duas atitudes podem ser vistas em acção em muitas das realizações musicais empregando computadores. Estes, como sistemas físicos que são, e pela sua capacidade manipulatória de símbolos, levam ao fascínio da simulação de comportamento 'inteligente' (mais concretamente aquelas características que nós associamos à ideia de inteligência no comportamento humano - compreensão da linguagem, aprendizagem, resolução de problemas...), que, no caso da composição musical, eu considero mais como um uso inteligente das maquinas do que a criação de 'inteligência' em si. Como, por ex., em certos programas que tomam decisões em situações económicas complexas, a apreciação da correção das conclusões por eles produzidas necessita do conhecimento intuitivo acerca da situação em causa, o que leva a considerações éticas no uso dos mesmos (ver Rainer Born 1987).

Assim, nada como a audição para detectar a RIGIDEZ de um estrutura, nada como a tentativa de criar uma certa competência musical, recorrendo aos vários métodos de redução computacional da complexidade para se SENTIR a multidimensionalidade das acções humanas.

Referências e Bibliografia

Paul Ricoeur - *Teoria da interpretação,* trad. port. Edições 70, 1987.

Jean-Charles François - *"La note musical et la synthese digitalisée de l'onde sonore"* em *Rethoriques de la technologie,* Traverses/26, 1982.

Luiz Moniz Pereira - *"Inteligência Artificial: Mito e Ciência"* Colóquio/Ciências n. 3, Outubro 1988.

Phillipe Roqueplo - *"La Nature demeure-t-elle l'Horizon de la Technique?"* em *Rethoriques de la technologie,* Traverses/26, 1982.

La Recherche en Intelligence Artificielle, número especial, Outubro 1985.

Curtis Roads - *"Grammars as Representations for Music"* em C. Roads e J. Straw, eds. 1985, *Foundations of Computer Music,* MIT Press.

Kevin Jones - *"Compositional Applications of Stochastic Processes"*, Computer Music Jour. 5(2), 1981.

Marco Ligabue - *"A System of Rules For Computer Improvisation"* em P. Berg, ed. 1986, Pro. of the 1986 International Computer Music Con.

Marvin Minsky - *"Music, Mind and Meaning"*, Computer Music Journal 5(3), 1981.

Lucien Sfez - *Critique de la Communication,* Seuil 1988.

Mira Balaban - *"Foundations for Artificial Intelligence Research of Western Tonal Music"*, B. Truax, ed. 1985, Proc. of the 1985 ICMC.

Otto E. Laske - *"Introduction to Cognitive Musicology"*, Computer Music Journal 12(1), 1988.

Luc Steels - *"Learning the Craft of Musical Composition"* em P. Berg, ed. 1986, Proc. of the 1986 ICMC.

David J. Hargreaves - *The Developmental Psychology of Music,* Cambridge University Press, 1986.

Kemal Ebcioglu – *"An Expert System for Harmonizing Four-Part Chorales"*, Comp. Music Journal 12(3), 1988.

Christoph Lischka, Hans-Werner Güsgen – *"MvS C: A Constraint-Based Approach to Musical Knowledge Representation"* em Paul Berg, ed. 1986, Proc. of the 1986 ICMC.

Marilyn Taft Thomas – *"VIVACE: A Rule Based Al System For Compositon"* em B. Truax, ed. 1985, Proc. of the 1985 ICMC.

Curtis Roads – *"Improvisation with George Lewis"* em C. Roads, *Composers and the Computer,* 1986.

Herbert Brun – *"The Premise is that there be Music"* em B. Truax, ed. 1985 Proc. of the 1985 ICMC.

Tod Machover – *"Être compositeur aujourd'hui* em *InHarmoniques 1,* IRCAM, 1988.

Douglas B. Lenat – *"Computer Software for Intelligent Systems"*, Scientific American, (Trends in Computing), pp. 68–75, 1988.

David Cope – *"An Expert System for Computer-Assisted Composition"*, Computer Music Journal 11(4), 1987.

Reiner Born (ed.) – *Artificial Intelligence: the Case Against*, C.R., 1987.

Sherry Turkle – *The Second Self,* Simon & Schuster, 1984.

Zenon W. Pylyshyn – *Computation and Cognition*, The MIT Press, 1984.

Charles Dodge, Thomas A. Jerse – *Computer Music: synthesis, composition and performance*, Schirmer Books, 1985.

III

IMPACTOS GLOBAIS

GLOBAL IMPACTS

Jesús Mosterín

Dep. of Logic, History and Philosophy of Science
Univ. of Barcelona, Barcelona (Spain)

The role of communication in the dynamics of cultural change

Abstract

The culture of an individual (i.e. the set of cultural traits or memes [Dawkins, 1976, 1982] available to him in his brain) is subject to continuous change. The sum or union of the cultures of all the members of the group being the cultural pool of the group, the culture of a group is also changing all the time.

The cultural evolution (of the individual or of the group) is a fact, which can be described. If we wish to go from description to explanation, we have to appeal to forces or factors which explain the changes in memes and constitute the dynamics of cultural change.

Every human group is a subgroup of that maximal human group, the humankind. If we look at the evolution of human culture (the culture of humankind), we can see periods of drift and of convergence. Drift and convergence are not random, they depend very strongly on the facility of communication. The easier the communication, the greater the convergence. The more difficult the communication, the greater the drift.

The most probable outcome of the actual technological conditions (barring generalized wars or comparable social catastrophes) seems to be the planetary growing up of a unique world culture, able to supply individuals everywhere with an extremely rich and varied array of cultural alternatives, geographically modulated by statistical variations in the distribution of memes as a function of ecological variables.

Resumo

A cultura de um indivíduo (i.e. o elenco de traços culturais ou memes [Dawkins 1976, 1982] que lhe são disponíveis no seu cérebro) está sujeita a mudança contínua. Sendo a soma ou união das culturas de todos os membros de um dado grupo o lote cultural desse grupo, a cultura do grupo está sujeita também a uma permanente mudança.

A evolução cultural (dos indivíduos ou do grupo) é um facto que pode ser descrito. Se quizermos passar da descrição à explanação, devemos apelar às forças ou factores que explicam as mudanças nos memes e que constituem a dinâmica da mudança cultural.

Cada grupo humano é um subgrupo do grupo humano máximo: a humanidade. Quando consideramos a evolução da cultura humana (a cultura da humanidade), podemos detectar períodos de deriva e períodos de convergência. Nem deriva nem convergência são ocasionais, dependem muito fortemente das facilidades da comunicação. Quanto mais fácil a comunicação, maior a convergência; quanto mais difícil, maior a deriva.

A resultante mais provável das condições tecnológicas actuais (se excluirmos guerras generalizadas ou catástrofes sociais comparáveis) parece ser o crescimento à escala planetária de uma cultura mundial única, capaz de proporcionar aos indivíduos, onde quer que se encontrem, uma panóplia rica e variada de alternativas culturais, modulada geograficamente por variações estatísticas na distribuição dos memes como uma função de variáveis ecológicas.

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 523/530
1992, APDC & SPF, Lisboa

The culture of an individual (i.e., the set of cultural traits or memes [Dawkins 1976, 1982] available to him in his brain) is subject to continuous change. He is learning and forgetting, adopting and rejecting memes all the time. So his culture is in evolution, as long as he is alive, in contrast to his genome, whose genes remain the same from conception till death.

The culture of a group is also changing all the time. The sum or union of the cultures of all the members of the group is the cultural pool of the group. The cultural pool of the group changes each time someone leaves the group (because of death or emigration or resignation or dismissal) or enters the group (because of birth or immigration or admission). It changes also each time the culture of any of the members changes.

The cultural evolution (of the individual or of the group) is a fact, which can be described. If we wish to go from description to explanation, we have to appeal to forces or factors which explain the changes in memes and constitute the dynamics of cultural change.

Every human group is a subgroup of that maximal human group, humankind. If we look at the evolution of human culture (the culture of humankind), we can see periods of drift and of convergence. Drift and convergence are not random, they depend very strongly on the facility of communication. The easier the communication, the greater the convergence. The more difficult the communication, the greater the drift.

Actual and virtual culture

An individual's culture at a certain moment is the totality of transferred information of non-genetic origin – the totality of memes – which is available to that individual at that moment. But the word "available" is here ambiguous. In a strong sense, a certain bit of information is available to an individual when it is actually stored in his brain and can be immediately retrieved at his will. An individual's *actual culture* is the whole of the information which is available to him in this strong sense. In a weaker sense, a bit of information is available to someone when he knows how to get it from an external artificial carrier at his disposal. The information available to him in this weaker sense is the individual's *virtual culture*.

My actual culture only covers the four telephone numbers I know by heart, but, thanks to the telephone directory on my table, my virtual culture covers all the telephone numbers in Barcelona (or even in the whole of Spain or the world, taking into account the telephone information services). I know the earth is round, I know I have to drive on the right (in my country) and to stop when the traffic lights turn red, I know how to ride a bicycle, how to use a screwdriver and how to sing the Marseillaise. All this belongs to my actual culture and the pertinent information is stored in my brain. I do not know exactly where Plymouth lies. But I have an atlas at hand, on the shelf, and I know how to find the situation of Plymouth: I have to look for "Plymouth" in the index, which will refer me to the page with the right map and to the relevant coordinates, so allowing me to locate Plymouth on the map and to properly interpret that location. I do not know how the sophisticated alarm-clock I have just received works, but with it comes a detailed brochure with instructions and explanations, written in a language I understand. The information about Plymouth's situation and the use of the new alarm-clock belongs to my virtual culture. Once retrieved and retained, that information becomes part of my actual culture.

All virtual culture is founded on certain actual culture, namely, on the actual (brain) information on how to retrieve the information stored in the external carrier. I can only

retrieve and use the virtual information contained in the telephone directory, the atlas or the instructions brochure, if I previously dispose in my brain of enough actual information on how to find and consult the book or brochure, on how to read and interpret it. i.e., if I can read and understand the language in which it is written and the cartographic code of the map, and if I am familiar enough with the handling of this type of sources.

Most of the information available to the lawyer, the accountant, the historian, the astronomer, etc. is virtual information. And an important part of the student's education consists in learning to locate the sources and to use the methods which will allow him to exploit the huge accumulations of information (libraries, data bases, etc.) virtually at his disposal, i.e., in enlarging his virtual culture.

Group's culture

The virtual culture depends on the actual one, and actual culture is information stored in a brain. But only individuals – no groups – have a brain. So strictly speaking culture resides in individuals, not in groups. But it is usual (and expedient) to speak sometimes of the culture of a social group (or set of individuals). For this *façon de parler* to make sense, it is necessary to define the social group's culture as a function of the cultures of the individuals which make up the group.

The culture of a social group is the culture of its members. This can mean different things. In its broadest sense, "the culture of its members" can mean the totality of the cultural information present in any members of the group, even in a single one. The group's culture, understood in this broad sense, will be called the *cultural pool* of the group as the cultural information which is shared by all and everyone of the members of the group. Very often when we speak of a group's culture, we mean neither its cultural pool not its unanimous culture, but something in between, nearer to the second, but less stringent. As a matter of fact, the unanimous culture of any territorially or politically defined social group is usually empty, as almost always there are infants and minority individuals which have not acquired the majority memes. A group's *shared culture* to the degree n or with probability n (where n is a number between 1/2 and 1) is the majority culture of the group. A certain meme or culture trait belongs to this shared culture if the ratio between the individuals of the group who have that meme in their individual cultures and all the members of the group is greater or equal to n. Shared culture can be more or less shared. In the extreme case, a group's shared culture to the degree 1 is identical to its unanimous culture.

The culture of an individual or a group changes (or at least can change) with time. And so, any precise definition has to include a temporal reference. Let us call $M(x,t)$ the set of all the memes which make up the culture of the individual x at the moment t. It is not our intention to give here formal definitions, but it would not be difficult to arrive at them. For example, we can define the cultural pool of the social group G in the moment t as the union of the cultures of its members, i.e.,

$$\text{Cultural pool of } G \text{ in } t = \bigcup_{x \in G} M(x,t)$$

The unanimous culture of the social group G in the moment t can be defined as the intersection of its member's cultures, i.e.

$$\text{Unanimous culture of } G \text{ in } t = \bigcap_{X \in G} M(x,t)$$

The probability or relative frequency $p(m, t)$ of a meme m in a social group G at a time t is the ratio between the number of individuals of G which have the meme m in their cultures and the total number of individuals of the group, i.e.

$$p(m,t) = \frac{|\{ X \in G \,/\, m \in M(x,t) \}|}{|G|}$$

Strictly speaking we should provide the letter p with a subindex G and so write pg, instead of just p, to make explicit the dependence of p on the reference group, but let us assume it without writing for the sake of simplicity. We can then define the other concepts as a function of p:

Cultural pool of G in $t = \{ m \mid p(m, t) > 0 \}$

Shared culture (to the degree n) of G in $t = \{ m \mid p(m, t) > n \}$

Unanimous culture of G in $t = \{ m \mid p(m, t) = 1 \}$

We have already remarked that the unanimous culture of a territorially or politically defined group – the European population, the Brazilian nation, the Dutch people – is probably empty, the empty set of memes. On the other hand, the unanimous culture of a culturally defined group – the French speaking community, the Islam, the tailor's guild or the college of doctors – is never empty, for it always contains at least the meme complex – the French language, the Islamic religion, the tailor's trade, the medical lore – which enters into its own definition. If someone lacks those memes in his culture, he does not belong to the group, by definition.

If a social group G is part of another group H (as for example, the group of the Dutch is included in the group of the Europeans), then the cultural pool of G is also included in the cultural pool of H (the Dutch culture pool is part of the European culture pool). On the other hand, if G is a part of H, then it is H's shared or unanimous culture which is included in G's shared or unanimous culture, and not the other way around. We can conclude that the larger a group is (along the same inclusion line), the larger is also its cultural pool, but the smaller its shared culture.

Diffusion and drift

The basic needs of the human being are given in his nature and are practically universal. The satisfaction of those needs presents a permanent challenge, to which humans have always and everywhere tried to react. The different answers to that challenge are the different cultures, which are like the different solutions to the multivariate equation of our needs. The same natural needs can be satisfied through different cultural means. And the different cultures are different solutions to the same problems.

Cultures not only supply solutions to the natural problems. They also create their own problems (and pseudoproblems) and bring up their own solutions (or lack of solutions).

A group's cultural tradition is always conservative. However revolutionary the change of a certain meme can be, the majority of memes remain stable. The invention, the discovery, the novelty in cultural transmission have normally been (at least till fairly recently) rare. A sociocultural group's cultural progress has more often taken place through the assimilation of foreign inventions than through its own inventions. The transmission of cultural information from one sociocultural group to another is called diffusion. Through the assimilation of a foreign meme, the group often achieves a better solution of a

preexistent problem, for which it already had its own – albeit less successful – way of dealing. But the diffusion not always has a progressive character. Also pathological memes expand from group to group through diffusion, as in the case of tobacco smoking.

Diffusion has been metaphorically characterized as cultural contagion. And the closer the contact, the easier contagion is. When several sociocultural systems establish contact and their members get easy opportunity of communication among themselves, the diffusion of cultural traits becomes unavoidable. The diffusion between two sociocultural systems can be mutual, but if there is a considerable difference in overall development between them, the diffusion will generally follow the direction from the more to the less developed group, not only in the cases of intrinsic advantage of the first group's cultural solutions, but also in many other irrelevant cases, due to the superior general prestige associated with the first system.

An extreme case of diffusion is the forced diffusion or acculturation, which often goes together with colonialism, war and oppression. The individuals are forced against their will to assimilate and adopt cultural traits imposed by the colonial lord, the conqueror or their own government. The Spanish conquest of America, for example, represented a flagrant case of acculturation or forced diffusion of not asked–for foreign memes (like the cultural complex of catholicism) among the indigenous populations. But acculturation does not always come from the outside. Nationalist movements and governments try to culturally homogenize the population living in a given territory, imposing on them (often against their will) the cultural traits (language, religion, etc.) which they identify with their presumed nation.

If contact induces diffusion, isolation produces drift. The continuous reproductive isolation of several subpopulations of a given population of organisms induces an increasing genetic differentiation among the different subpopulations, together with an increasing genetic uniformity among the individuals inside each subpopulation. Something similar happens with culture. The separation of a culturally more or less homogenous population into isolated subpopulations determines the cultural drift, i.e., the differential evolution and the fragmentation of the initial culture. A specially clear example is the geographical fragmentation of the Latin language into different Romance tongues, as a consequence of the isolation into which the different south European populations fell after the collapse of the Roman Empire, and of the subsequent abandonment of roads and bridges and infestation of land ways and sea routes by bandits and pirates.

Comparable and incomparable memes.

When two cultural traits are alternative solutions to the same problem, sometimes they can be weighed and compared in an objective way, and sometimes they cannot. The Chinese and Japanese eat rice with sticks, the European and Americans with a fork. Both systems fulfil their function with equal ease (once one gets habituated). The English drive to the left, the French to the right. Obviously it is necessary for traffic security that all car drivers going in the same direction drive to the same side of the road. But whether this side is the left or the right is a matter of indifference. Europeans express their mourning wearing black cloths; the Chinese, wearing white garments. But none of these colors is intrinsically more connected with mourning than the other. All these are instances of unweighable, incommensurable, incomparable memes.

Other cultural traits or memes are weighable, commensurable, comparable. They have a clearly defined task to perform or problem to solve, and their efficiency at performing

that task can be objectively measured and compared. The function of a knife is to cut. The better it cuts, the better knife it is. The same can be said of an ax. And it is clear than a steel knife or a steel ax cuts much better than a stone knife or a stone ax. Every time a cultural group with stone cutting instruments has come into contact with another one with steel instruments, it has rushed to adopt them. Even the primitive Yanomami of the upper Orinoco, reluctant to adopt any foreign memes, have adopted the steel knives and axes. Stone knives and steel knives are comparable memes, and the second are objectively better than the first. Traditional Chinese medicine prescribes to swallow a living tadpole in order to avoid getting pregnant. But, of course, condoms, the pill, sterilization and other modern methods of contraception are much more successful in achieving their goal. It is because of that superior efficiency that even the Chinese have adopted them.

The evolution of group's culture takes place by means of (1) the invention or introduction of new cultural dimensions or tasks, (2) the invention or introduction of new memes for the already available tasks, and (3) the variation in the relative frequency of alternative memes in the course of time. If we compare cultural with biological evolution, we could compare the first type of change with the macromutations producing polyploidy, the second with normal mutations in an already available locus of the karyotype, the third with the evolution of genic frequencies studied by population genetics.

Almost all formal quantitative studies of cultural evolution [Cavalli–Sforza & Feldman 1981, Boyd & Richerson 1985] are centered in the analysis of the third type of changes, i.e., in the analysis of the changes of the frequency distribution or a fixed stock of memes in time, in analogy to population genetics. Drift and natural selection are the main forces in population genetics. Drift also plays an important role in cultural groups subject to isolation. What is here the analog to natural selection? For the comparable memes, it is something like rational choice. For the incomparable memes there is not anything like natural selection. The acting forces are something like fashion or tradition, which perhaps are more akin to drift than to natural selection.

Isolation increases drift, communication leads to diffusion and convergence, leading to an increased supply of cultural varieties and to change in memes frequency as a function of rational choice in view of local ecological specificities [Mosterín 1986, 1988].

Convergence

Isolation and difficulty of transport and communication determine the cultural drift, the fragmentation of the world in different cultural areas. Contact and facility of transport and communication determine the diffusion of memes and the convergence of cultures. We say that two cultures converge (or that two social groups converge culturally) if the number of memes they share is increasing all the time.

During most of history the different human groups have been tremendously isolated from each other by impassable geographical barriers. China was surrounded by frozen taigas, dry deserts, inaccessible mountains and the pacific Ocean. So it is no wonder that the Chinese identified the civilized world with China, beyond which only lay inhospitable deserts, sparsely populated by brutish savages. That explains also that the Chinese secret of the silk production could be kept for millennia, even after the stuff itself began to arrive in the West along the arduous silk route. America was an island, geographically and culturally isolated, ignorant of such foreign developments as the foundry of metals or the use of wheels. More surprising still is the fact that the obviously most useful invention of the wheel took two thousand years to be diffused from Mesopotamia (where it was already

known about – 3500) to the nearby Egypt, where it only arrived around – 1570 (already in the New Empire).

Progress in communication has been exceedingly slow in the past. Napoleon could not send his news and directives from Rome to Paris any quicker than Caesar. News could not travel at a higher speed that its carriers (a man riding a horse in Europe and Mongolia, a man running in Africa, America and much of Asia), and that means 100 km. per day in the best of cases. The revolution in transport and communication which has taken place in the last century and a half has changed all that. Now we can go anywhere on the planet in the same day. And news can travel at the speed of light, i.e., for all practical purposes, instantaneously.

Cultural convergence and cultural drift are not random processes. Cultural convergence is determined by contact and communication. Cultural drift is determined by isolation. It is true that there are still pockets of isolation on our planet, some imposed by abject poverty and others by the political will of totalitarian political systems. But this notwithstanding it is beyond doubt that in our time we have reached an unprecedented facility of travel, contact and transport, specially transport of information, i.e., of communication. Because of that it does not come as a surprise that we also observe an unprecedented cultural convergence. This convergence began in the 16th century and has increasingly accelerated ever since, reaching already now a spectacular level.

At this very moment we are witnessing the constitution of a unique world culture, in which melt and on which converge the different traditional ethnical or national cultures. The process can be halted for a while here or there, but in its overall effect it is irreversible (unless a world war destroys the very technological base which sustains it and makes it unavoidable).

The genetic difference between human populations is normally much smaller than the genetic variability inside a single human population [Ayala, 1980]. As it is the human genetic pool which defines the human nature and determines our basic needs and motivations, whose satisfaction poses the problem which the different cultures try to solve, and given the relative similarity of the genetic pools of the different human populations, we could expect to find a certain cultural uniformity among human populations, only varied by ecological adaptation to local conditions such as wether and soil. But it is not so. The remarkable cultural dispersion and fragmentation of a humankind genetically relatively homogenous is a paradox, only explainable by the tremendous isolation in which most human populations lived till recently, which made very difficult and improbable the transfer of memes between different groups and the cultural contagion. In those situations of isolation the random forces play a very important role, and determine a cultural divergence (randomness acts each time in a different direction) which goes much beyond simple adaptation to local conditions.

As the isolation between groups breaks down and cultural convergence increases, it is to be expected that the role played by random forces in the evolution of culture will decrease. And this evolution seems to point in a direction in which the world as a whole will become more and more homogeneous, as its different areas become more and more similar. By the same process each area will become internally more heterogeneous, as the cultural supply inside it diversifies and becomes richer with the arrival of lots of new foreign memes which come to join the autochthonous ones. The positive aspect of the process is the increased opportunity for each individual of finding available in his area the peculiar combination of memes which would best suit his personal taste, idiosyncrasy and circumstances, and so contribute to his happiness. The danger implicit in the process is the

risk that the cultural traits of small groups disappear, so decreasing the total cultural pool and cultural variability of mankind.

Anyway, the most probable outcome of the actual technological conditions (barring generalized wars or comparable social catastrophes) seems to be the planetary growing up of a unique world culture, able to supply individuals everywhere with an extremely rich and varied array of cultural alternatives, geographically modulated by statistical variations in the distribution of memes as a function of ecological variables.

References

Ayala, F. J.: 1980. *Origen y evolución del hombre.* Madrid: Alianza Editorial.

Boyd, R. & Richerson, P.: 1985. *Culture and the Evolutionary Process.* The University of Chicago Press.

Cavalli–Sforza, L. & Feldman, M.: 1981. *Cultural Transmission and Evolution: A Quantitative Approach.* Princeton University Press.

Dawkins, Richard: 1976. *The Selfish Gene.* Oxford University Press.

Dawkins, Richard: 1982. *The Extended Phenotype.* Oxford University Press.

Mosterín, Jesús: 1986. *"Cultura como información".* In J. Sanmartín (ed.), *La sociedad naturalizada. Genética y conducta.* Tirant lo Blanch. Valencia.

Mosterín, Jesús: 1988. *"Ontological Queries and Evolutionary Processes". Biology & Philosophy, v. 3, n. 2.*

Valdemar W. Setzer

Dep. of Computer Science,
Univ. of São Paulo, São Paulo (Brasil)

The Computerized Society: New Life or Destruction?

Abstract

In this paper we analyse certain aspects of the problems that may arise with the indiscriminate introduction of computers in society. We show that some applications are not intrinsically harmful, contrary to others, mainly due to the impoverishment of the information. In particular, we make an analysis of the consequences of treating information through computer programs. Some attention is given to the replacement of human beings by computers, with special attention to the use of these machines in education. Some positive aspects of computers are mentioned, specially the possibility of every person with some education getting to know their logical functioning, and the fact that they permit to be totally controlled, contrary to other machines. We conclude with the need to develop a new kind of thinking to deal with the present problems that afflict humanity, mainly in respect to technology.

Resumo

Neste trabalho analisamos certos aspectos dos problemas que podem surgir com a introdução indiscriminada de computadores na sociedade. Mostramos que algumas aplicações não são intrinsicamente perigosas, contrariamente a outras, principalmente devido ao empobrecimento das informações. Em particular, fazemos uma análise das consequências de processar-se informações através de programas. Alguma atenção é dada à substituição de seres humanos por computadores, com especial atenção ao uso dessas máquinas em educação. Alguns aspectos positivos dos computadores são mencionados, especialmente a possibilidade de cada pessoa com alguma formação poder conhecer o seu funcionamento lógico, e o fato de que eles permitem ser totalmente dominados, ao contrário de outras máquinas. Concluímos com a necessidade de desenvolver-se um novo tipo de pensamento para tratar os problemas correntes que afligem a humanidade, principalmente em termos da técnica.

The computer and the computerized society

Every society has employed information. Before writing was developed, this information was completely informal; writing was the first step to the increasing formalization and abstraction of information. In more recent times, up to the 1950's, societies became increasingly based upon information, a good part of it, semi-formal, because it was written in some natural language. Another part of the information, essentially scientific, employed some formalism in its representation, sometimes of mathematical (including geometrical) form. More recently, information began to be stored in and processed by computers. Let us characterize a computerized society as a society

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposium, 1989, 531/541
1992, APDC & SPF, Lisboa

which is strongly based on information stored in computers, and its treatment by these machines.

The information introduced in computers may be divided into two classes: 1) texts in natural language, coded sound and coded printed figures; 2) programs. There is an enormous distinction between these two classes. In the first, the computer is merely used as a storage device. In the second, one stores the procedures which the information is going to be processed with, either just transformed or used as the basis for the generation of new data.

In the first class, there may exist non-formal elements. For instance, a poem may be stored in the machine. Evidently, its significance transcends any formal interpretation one may give to the stored symbols. Storing natural language phrases just uses the poor quantified nature of written letters. On the other hand, the necessary codification of sound or figures is in general far from their original representation and produces the lost of some of its richness.

The second class of information stored in computers has a totally different aspect: it is completely formal and abstract, having and exact, mathematical meaning. In terms of the formal interpretation which one may give to any programming language, that meaning is self-contained. It is important to recognize that computers work only by executing formal, mathematical instructions – or, better, logical instructions, which are grouped sequentially into programs. A computer program is a mathematical model of something, and represents a total alienation with respect to reality. In this sense, one may say that a such a program is pure syntax and has no semantics. Moreover, the computer is a machine that produces nothing, that is, no usable, consumable goods, unless it is connected to another machine, controlling it, and the latter produces something. In this sense, the computer does not do anything which cannot be done by humans, either mentally or with paper and pencil. In other words, it is always possible for a person to simulate the functioning of computers – considering its data processing function. The only difference, and eventual limitation in this comparison, is speed, much slower in the human being when he is executing exactly the same instructions. It is also important to recognize that the computer is a discrete machine, a quantified system, whose data, programs and results must by necessity be reducible to "numbers".

We consider the invention of motor-driven machines the *first revolution in the human work*. It coincides with the so-called industrial revolution, and may be characterized by the replacement of man's physical effort by machines. This was added up to the increase of sensorial capacities which occurred previously with the telescope's invention. The *second revolution in the human work* happened with the introduction of the digital computer. It did not replace man's physical effort, but his mental effort.

Due to the first revolution, using a power shovel a very weak person, asthmatic, rachitic, is able to dig a big hole in the ground in just a couple of minutes. This task would have required a long work executed manually by many strong, healthy men. Due to the second revolution, a mental retarded person, with a minimum of intellectual and memorizing abilities is nowadays able to execute tasks for which a reasonable amount of knowledge and intellectual abilities, were previously required. The recent "ironic" languages make it even possible for illiterate people to use computers in quite complex tasks, choosing functions through selection of figures displayed on a screen. Thus one sees that computers may replace some mental activities. But, by no means one should say that they may replace *every* such activity. According to the two classes of information described before, which may be stored in computers, we may characterize the first one, comprising

texts, figures and sound, as simple extensions of previous existing forms of storing them: written texts, printed figures, records and magnetic tapes, and a combination of these as movie films. This "simple extension" means here just the increase in recovery speed and easiness of use.

It is with the second class of informations that something new appears with the introduction of computers: the possibility of introducing into the machine a kind of thought, under the form of a program. It is evident that anything produced by man, be it a house, a machine or a piece of art work, contains some thoughts in its form. However, in the case of computers the thinking that is introduced into the machine under the form of programs is of an extremely particular nature: it is a strictly formal, mathematical-logical, discrete and finite kind of thinking. A lathe also contains thoughts - its design, its construction, form and functionality. But its action is not totally formal, because it produces articles under a certain precision. That is, it is not possible to antecipate with mathematical exactitude the result of a lathe's work. As with any other machine except for the computer, the lathe's functioning depends on physical nature, in this case, through the materials being worked by the machine and its own materials. It is a technical marvel that a machine like the computer has been built with physical materials but its functioning, being mathematical, transcends physical constraints.

As we have pointed out, the mathematical functioning of a computer does not cover the whole of mathematics: we can only introduce into the machine the algorithmic mathematics which is discrete and finite in its description and "execution", or in any other mathematical model reducible to these restrictions. Thus, a program must be by necessity the outcome of a quite limited kind of thinking. Considering that we are able to think other subjects than mathematics, and even in the latter, parts of it are not algorithmic or finite, one may conclude that the computer cannot replace every type of mental activity.

It is interesting to note that the methods adopted by physicists since Descartes and Newton have established limits to our knowledge of nature (wouldn't that be enough of a sign to begin searching for or adopting new scientific methods?), as for instance Heisenberg's Uncertainty Principle (theorem) or the equations of Quantum Mechanics. In the latter, some terms have no classical limit and thus cannot be understood based on our physical perceptions (the "spin" of an electron is no spinning whatsoever). Moreover, Mathematics and Computer Science have also established limits to their possibilities of solving problems. For instance, it is possible to prove that it is impossible to write a computer program that reads another program and proves that it executes a function also given as input. It is also impossible to develop a program to sort a certain number of words (or numbers) in linear time (which is the minimum time complexity when there is a variable number of input data). There also exist combinatorial limits, that is, for a certain problem the number of combinations a program is required to make is so great that the time required to solve the problem makes it infeasible, even with the fastest imaginable machines.

With these considerations, and taking into account the characterization of a computerized society given at the beginning, one may conclude the following. If the computer replaces human activities of the first class - storing texts, sound and figures without transforming or analizing them - and also those activities which are intrinsically of mathematical-algorithmic nature, in principle it will not be influencing society in a negative way. An example of the latter is accounting: it deals with an "a priori" quantified subject, which is money. It is not by chance that there has been a happy marriage between computers and the banking system (as there has been between TV and advertising). It has permited banks to inflate themselves to a degree which was not foreseeable some years ago.

We are aware of the objections that may be stated by some people: sounds emitted by things or beings (that is, excluding sound-producing electronic equipment) reveal non-formal inner characteristics of its emitters, and its codification in a computer may eliminate this subtle richeness. The problem here is that records and tapes have already been having this bad effect, so the computer does not introduce any novel change in the process. In this case we would have to deal with a more general problem - that of the society's "machinezation".

Now, if the computer is programmed to process, to transform text, sound or image, or even if it is programmed to analize them reaching some result, we are not anymore in the simple usage of the machine as a means of just storing and transmitting information. In this case we are acting in the region of the other class of information introduced into a computer: the human thought converted into programs which produce that transformation. The problem is that this transformation is purely formal and mathematical. There lies a deep destruction of the society.

The destruction

If the computer is programmed to transform or analize information it may be contributing to the destruction of the society. This happens because it will be necessarily impoverishing that information, treating it as mere mathematical-algorithmic, quantified objects. Obviously, we are using an axiom: there exists information which is non-mathematical, and non-algorithmic. For instance, the concept of a "tree". Fortunately, in early education it is kept from being trivialized by a formal definition, as is the case with the concept of an "island". In our country it is given as "a piece of land surrounded by water from all sides", which is incorrect, because there is no water above and below it, and if it is not polygonal, there are only two planar sides: inside and outside. Besides being wrong, this definition "kills" the child's imagination because, as any definition, it is dry and dead. (Looking at a 1960 College Edition of Webster's Dictionary, being used while we are writing these lines, we encounter the same error.)

Let us examine one example of an improper use of computers in a highly computerized society, along the lines presented above. Let us suppose that, as in the film "War Games", the decision of lunching a nuclear rocket against another country be taken by computers, without human interference. This is an example of the so-called "computer decision making". It happens that it is not possible to properly handle social and managerial informations in a mathematical fashion, because they are not mathematical objects. D. Warnier, in his last, extraordinary book has stated that computers do not take decision, only humans do it; computers only make "logical choices" (1). Social and managerial decisions should be essentially based upon non-formalized knowledge and judgement. Mathematics may help, but the ultimate decisions have to be intuitive, based upon global knowledge. Any formalization in this area tends to impoverish the information. The mentioned film shows, in deep metaphors, what could happen if decision making is relegated to computers. In Brazil, one may see how formal reasoning in Economics always lead to disasters: society simply refuses to follow the formal model economists tend to force upon it, and some imponderable element is always left out of the equations, leading to negative, unexpected results.

So we consider as the highest danger in the computerized society its cultural impoverishment and the assignement of essentially human activities to computers. There are various examples already under development. The most common, and certainly one of

the dangerest is the use of computers in education. We have already had the opportunity of writing our ideas about this subject (2,3). We will only mention some of the arguments here.

Education is an essentially human activity. Any replacement of human beings in this activity envolves a dangerous impoverishment and the treatement of a child as a machine. A bicyle also "treats" a child like a machine, but only partially, because when driving it he acts in an infinite number of informal, unconscious ways that may eventually be simulated by a machine, but certainly using other inner processes. On the other hand, using or programming a computer requires always a formal activity; the thinking process has to be reduced to those thoughts which can be introduced into the machine.

Being a professor, we have had many experiences of getting an incorrect answer from a student, but we intuitively knew that, even though the formulation was wrong, the knowledge was correct. A machine will never accept a wrong answer as a correct one, unless it is programmed to do so for very specific, forseeable cases. It happens that a lesson should never be foreseeable, because in this case the instructor and students will be reduced to machines, the computers. A lesson "given" by a computer has to be totally and formally planned: all the possible interactions with the student have to be foreseen, because if the program which is controlling the whole process receives an unexpected data input, it will give unpredictable results, probably undesired and certainly wrong. Thus, this lesson has to be planned having in mind an abstract student, and cannot get out of the previously defined tracks. That is, the student will be treated as an abstraction. The way education has of escaping this pattern, which is in fact the old Programmed Instruction introduced by Skinner in 1954, is to make the student learn how to program the computer. This is exactly the activity the student is subjected to (or forced to follow) when he is using Papert's LOGO system(4). Following their master Papert, its proponents consider this activity quite superior to computerized programmed instruction, because they think that programming a computer is an open activity. In fact it is, but in an exclusively mental space, and a very restricted one. In the mentioned books, we had the opportunity of analysing the fact that the computer forces upon the student a formal language and a restricted mathematical thinking. Based on this fact, and using as a background Rudolf Steiner's Waldorf Education (being practiced since 1919 in over 400 schools in the western world) and his concepts of the evolution of each human being, we tried to answer the question: is there an appropriate age to subject children to computers? Our answer was yes, and this age should not be less than 17. After that purely intellectual conclusion, we had the opportunity of testing its adequacy. We organized a short introduction to computers and computing, called "Computing Day", inviting classes of senior high school students for an 8-hour short course (5 hours of class lectures and 3 hours of practice with microcomputers) given in one day in our department at the University of São Paulo. Sometimes younger students sneaked into the program, and once by mistake we had a full class of 15 years old high school students. The result confirmed our conclusions: the younger students were not interested in the concepts we were trying to transmit; they only wanted to play with the machine as a toy. They had no objectivity in regard to the computer, and were not interested in its structural aspects, which were one of our main educational goals. This shows that they did not have the necessary maturity to face the machine.

There is a point where we agree with Papert's followers: the computer has an impact on the education of thinking. But we consider it a bad influence if used with young children or early teenagers, because at these ages they should not be forced to practice a formal, mathematical thinking whith deterministic aspects of cause-and-effect. Even mathematics

can be introduced in an artistic way and maintaining its relation with the real world.

We consider the use of computers in elementary education an important step in the destruction of society, because it negatively influences that activity which is the highest and most universal of our inner processes: thinking. We fear the consequences which may occur if the mentality of adults will be formed by the type of thinking that one is necessary to use in order to program a computer. Will these addults have the (intuitive) feeling that humans are essentially different from machines? Will these people delegate social decisions to computers, admitting that these machines are faster and practically do not make mistakes? Will they find it natural that these machines should control humans, and we will end up having a much worse "1984" than Orwell's, if a real "War Games" produced solely by these machines does not finish with this world beforehand?

There are other forms of destruction produced by computers. For instance, one could suppose that if a machine replaces a man in a sub-human kind of work that machine would be doing him some good. This is only true if we give that person a more dignified work to do. In general, this is not happening, because it does not belong to the capitalistic ethics (is there one?) or, in the "other side", human dignity cannot be covered by a (coherent) materialistic Weltanschauung.

It would take too much space to cover other types of destruction that could be produced by computers in the individual and in society. Let us just finish this chapter with the statement that our biggest concern along these lines is the impact of computers in people's mentality(5). It seems to us that they decisively contribute to diminish the image man should make of himself, of his grandeur, and the recognition that he transcends the mineral, plant and animal nature. It may induce the unjustified opinion that man is a machine. What does a layman feel when he faces a machine which apparently thinks, speaks without clearly repeating a voice recording, etc? The problem here is not just with laymen: for instance, observe the mentality of the great majority of technicians and researchers who work in the area of Artificial Intelligence, AI (we prefer to interpret these initials as "Automated Imbecility", because the machine always act in totally imbecile way - intelligent were the people who devised and programmed the algorithms of that field). This unfortunate nomenclature was one step further in the degradation of the image man makes of himself, and in the raise of the machine to a level where it does not belong. Let us add that we consider AI to have a great merit: any AI system shows that we humans certainly *do not* function in that way. Knowing what we are not maybe helps forming a better image of what we are...

The new life

We are not against machines or against technology. We are against their misuse, the most frequent nowadays. Let us demonstrate our position covering some positive role the computer could exert in society.

It seems to us that the computer permits that we develop a new consciousness towards machines and permits a new degree of dominance over them. The new consciousness comes from the fact that one may understand with relative easiness the computer's logical functioning. We are preocupied with the fact that there is an increasing ignorance in relation to the principles of machine functioning. How many people know the working principles of internal-combustion motors? We do not advocate that everyone should know how to repair his car, but consider it to be very dangerous to have an almost total unconsciousness in relation to machines. This may degrade the human being, because it

causes a kind of mental paralysis. In fact, the normal atitude of a person that looks at a car is to ask himself "how does this thing work, how does its motor work, and how does it transmit impulse to the wheels?" and proceed to take actions to clarify his doubts. Unfortunately, in general that "normal" does not happen: if the questions are made, the complexity of machines prevent the search for the answers; worse than this, it seems that these questions occur less and less. With it, man ceases to be moved by a curiosity of knowing, thus violating a fundamental characteristic of human beings: not becoming satisfied with a mere sensorial perception, and providing, through thinking, a bridge to a concept. About this, see Steiner's magnificent work "Philosophie der Freiheit"(6).

If we look at a brown rectangular form, somewhat larger than ourselves, with a small metal protuberance with an "L" shape, our thinking builds a bridge between the corresponding mental image ("Vorstellung") and the concept of a door. We become satisfied and act according to the functionality associated with the concept. The incapacity of building this bridge should lead us to look for knowledge. Maybe the concept is missing, maybe the thinking process is not adequate, maybe the sensorial perception is not adequate. If the human being is lead to eliminate this search, he will stop being a man and will enter increasingly in a state of animal automatism, without mental consciousness.

The computer, contrary to other machines, permits a total conceptualization of its logical functioning because, as referred above, it is not sujected to nature's physical "laws" (if they exist), and works externally in a complete mathematical way. Thus, the computer introduces the possibility of a new life of people in relation to machines, because it permits its use with a complete comprehension of its working principles. Moreover, as we have already pointed out, as long as it does not control other machines, the computer produces nothing material. Adding up the fact that it is a strictly deterministic machine, we have something which does not exist in other machines: we may have total control over its functioning. For the first time man has become the total master of a machine. He is not the master of cars, because these have undesirable side-effects (pollution, isolation of passengers in relation to the environment, psychological effects due to the enormous flow of visual impressions, etc.). Automobiles do not do solely what we ideally would like them to do, that is, just transportation without any other side-effect, and the analogy applies to any other machine. On the other hand, the computer permits that one forces it to do exactly what one wishes from him, and nothing else.

These two factors, the possibility to easily getting to know how the machine works, to a level inexistent in other machines, and the possibility to totally dominate it could mean a new position of regarding machines and technology. It was this new positioning that led us to propose that every high school student should not finish the school without a basic knowledge of what computers and programs are and how one can use this machine well or badly. This was the origin of the "Computing Day" mentioned before.

Conclusions

A new consciousness in relation to machines, now possible with the computer, could mean a new life for society. However, we are sure that this consciousness would be against an exagerated computerization of society. This consciousness would show the brutal impoverishment which such an exageration would bring to our life, as was pointed out by the philosopher H. Skolimowsky, from the U. of Chicago(7). He states that the total automation and programming by the computer is not an expression of freedom, but of slavery, and he states that it is time to think the deep consequences of a general introduction

of these machines in our everyday life. In our opinion, the computer permits that every person becomes a king–philosopher, as long as he has had the proper education along the lines exposed here.

What is the use of technology? We have posed this question to the audiences of almost every lecture we have been giving lately involving social aspects of computers. Invariably, the answer we get is "to facilitate man's life, to give him more comfort", referring to the satisfaction of our animal instincts. We believe that the deep meaning of technology is not that one, but to give man more freedom in relation to nature. Would it not be for printing technology, few people would be able to enter in contact with the ideas exposed here. Notice that the written word gives the reader total freedom of examining and analysing the contents of the text, in a calm environment, without any impositions. We would be satisfied if just one person could be touched by this paper, but much more than if it would have been a consequence of attending one of our lectures. In this case, not only our ideas, but our enthusiasm, and the unconscious relation one establishes with another present personality, could have influenced the hearer, diminishing his freedom.

It is urgent to create a new consciousness in relation to machines and technology. For instance, one should eliminate the falacy that "machines are neutral; the benefit or harm they may cause depend on their use". An automobile may be excelently driven: it will continue to pollute and cause psychological harm. (Let us take advantage of the occasion to mention the subject of our other battle against the misuse of technology: TV.) Neurophysilogical research has demonstrated that TV produces on the viewer a state of drowsiness, a semi–hypnotical state, independent of the program being transmitted. Our studies about this apparatus have led us to the conclusion that there is no positive effect of its use as a mass communication medium or, in some special cases, that could not be obtained with more advantages and less negative effects by other means. In particular, as we have shown, the computer forces a certain type of thinking and a formal language, extremely dangerous if improperly applied. It is also the most mistifying of all machines, because it simulates certain types of thinking, and may cause undue admiration to machines and induce an unjustifiable materialistic mentality. Moreover, if used professionally by programmers and systems analysts in an intensive way, it may cause serious psychological effects, as for instance fixed thoughts, unsocial behaviour, etc. (the so–called "Data Processing Syndrome").

It also urges to recognize that, under a holistic point of view, technology is not improving conditions of life and is impairing humanity's evolution. As E.F. Schumacher, the famous author of "*Small is Beautiful*" said, for each problem it solves, science creates a host of new problems, to the point where "at the height of success, with technological competence unequalled in all human history, we read nothing but questions of survival ... so we have to ask ourselves if something has gone wrong in the development of technology, that finest thing we've got for our *material life*"(8) (our italics). In our opinion, what happened was, from one side, the unconscious adoption of that technology without criticism. From another side, technology began to be used outside the material life, as for instance in arts (canned music, movie films), in education (TV, computers) and even in religion (for instance, the cheap transmition of religious cerimonies, sermons, etc.). J. Weinzenbaum got dismayed when he realized that some psychiatrists were seriously proposing using his famous program *ELIZA* in the treatment of the mental disturbed, which prompted him to write his extraordinay book(9).

N. Postman has shown in an excellent book how there has been a brutal impoverishment of our culture due, in which almost everything has been transformed into a "show",

mainly due to TV(10). He showed that it was not Orwell's prophecy (the state controlling the citizen) but Huxley's (pleasure controlling the individual) that came into reality. The indiscriminate computerization of our society will produce not only cultural, but also social impoverishment. For example, MacLuhan's "integrated village" linked through a computer net does not enrich social communication, on the contrary. Communication through computer terminals is not integrated, it does not produce social cohesion, it separates people. The government that tries to look at citizens through the computer prism will have the tendency to treat them as machines and not as human beings. Let us not forget that the computer permits a tremendous, instant control over the citizens, following quantitatively their life and movements.

We also believe that it is fundamental to recognize that the problems which press humanity are only part technological (in general, that part caused by technology itself...). Most of them are human problems. The serious problems aflicting humanity will not be solved by technology, and much less by computational technology. It is really fantastic that so many people still believe in what we call "scientific romanticism", dreamming about the day in which science and technology will solve every human problem. On the other side, it is also dangerous to face history as composed of repetition cycles, and that man is the same as he was thousands of years ago. The critical situation of our times has never existed before, and it is not by crossing our arms, waiting for the end of the present cycle, that we will solve our problems. Even worse is to try to go back to ancient traditions; old costums and practices transposed to later times always tend to create degeneration, as one may see in some religious sects. The time requires changing something, because if we continue in the same way chaos seems to be inevitable. It is not technology which will produce the necessary improvement. Looking back at the recent past, one may extrapolate with some certainty that technology will only worsen the situation.

The necessary change has to deal with mentality and social consciousness, that is, we need a new type of thinking, more impregnated with life and less with death, as it is the case with the formalistic scientific thinking of our times. The society's computarization will not bring these changes, because the mentality imposed by computers is evidently materialistic and anti-human, and they always treat the user as a machine. In particular, as we have shown, the computer affects directly the thinking process, and thus it is, of all machines, the most dangerous, preventing the necessary change in the present way of thinking. Even worse, it directs thinking toward an area completely alienated from reality. It represents a true pollution of thinking, and because it affects this exclusively human process, the "highest" of our inner activities, it requires an even acuter consciousness than with other machines, in order to place it in its correct position.

A recommendation, which is simple to formulate but difficult to be realized, is that the computer should not replace human activities which elevate, upgrade man (as for instance social work, medicine, health, education, arts and religion), but only those that degrade him. As G. Boyd has put it, "the development that made standardized appliances, vehicles and ready-to-wear clothing cheaply available to everyone is perceived as "good". Unfortunately there is a profound difference between articles whose main functions are of sustenance (i.e. food, clothing, furniture, etc.) and those whose main function is to inspire hope, such as music, art and education. Automating motor car production mainly eliminates sub-human uses of human beings, but to try to automate artistic, musical and teaching performance is to eliminate the quintessentially human uses of human beings"(11). The difficulty lies in the fact that our simple rule requires something that does not exist anymore for almost the whole of humanity: a Weltanschauung adequate to our modern times, giving

a conceptual and practical understanding of what are the human entity and the universe. This view was provided in ancient times by the direct semiconscious vision, by the communications at the mystery places, miths and religions. Those forms do not satisfy contemporary man anymore, because he has enormously developed his capacity of objectively observing his environment and his thinking, has separetad himself from nature and abandoned all the traditions which gave him sustenance up to the beginning of this century. Besides that, he requires a conceptual approach (but not formal, and much less mathematical, because these are too restrictive) to be able to understand his own role in the universe, to take actions and to act in a conscious and free way.

Science tried to replace religion, but the result was the elimination of man's positioning in relation to himself and to the universe, without providing satisfactory and encompassing concepts. The reason for this is that science has only given incomplete, materialistic explanations (that is, based on our physical senses and mathematical abstractions). However, everyone may verify in himself the existence of inner processes with great evidences of not being physical, as willing, feeling, thinking, apparently infinite, layered memory (machines have only one layer – they either have some data stored in them, or do not have that data) and self-consciousness (which even animals do not have).

We should not throw away machines and technology; there is no sense in going back to a tranquil, passed away world. They have to be redeemed and put essentially to the service of giving man more and more freedom, so that he becomes more and more human, less and less animal. Goethe touches this mystery in his Faust I, scene of the "Studierzimmer", in which, being asked who was he, Mephistopheles answers to the scientist Faust: "Ein Teil von jener Kraft, die stets das Böse will und stets das Gute schafft" (a part of that force, which always wants the evil but always creates the good). As Faust well notes in the sequel, Mephisto calls himself "a part" but seems to appear in his entirety. Technology is one of his parts. It could lead man to destruction if he abandons himself to it in an unconscious way. However, if he starts using it adequately, he could elevate himself, and redeem the evil that hides, disguise behind it. We are not just accountable for ourselves, but also for the universe.

References

(1) Warnier, D. – *Computers and Human Intelligence*. Prentice Hall, Englewood Cliffs 1986.

(2) Setzer, V.W. – O Computador no Ensino: Nova Vida ou Destruição? [Computers in education: new life or destruction?] in E.O.C. Chaves & V.W. Setzer, *O uso de Computadores em Escolas: Fundamentos e Críticas* [The use of computers in education: fundamentals and criticisms], Editora Scipione, São Paulo 1987, pp. 69–123.

(3) Setzer, V.W. – *Computers and Education*. Floris Books, Edinburgh 1989.

(4) Papert, S. – *Mindstorms: Children, Computers and Powerful Ideas*. Basic Books, New York 1980.

(5) Setzer, V.W. – O Computador como Instrumento do Cientificismo [The computer as an instrument of scientism]. *Anais do I Simpósio Anual,* Academia de Ciências do Estado de São Paulo, São Paulo 1976, pp. 69–88.

(6) Steiner, R. – *The Philosophy of Spiritual Activity*. Rudolf Steiner Publications, West Nyack, New York 1963.

(7) Skolimowski, H. – *Society Supplied with Informatics. The Era of Philosopher-kings is coming. Vital Speeches* Vol. 1, Nº 16, June 1984.

(8) Schumacher, E.F. – *Good Work.* Harper, New York 1979.

(9) Weizenbaum, J. – *Computer Power and Human Reason: from Judgement to Calculation.* W.H. Freeman, San Francisco 1976.

(10) Postman, N. – *Amusing Ourselves to Death: Public Discourse in the Age of Show Business.* Penguin, New York 1986.

(11) Boyd, G. – *Education and Miseducation by Computer*, in *World Yearbook of Education 1982/83, Computers and Education.* Nichols, New York 1983 and Kogan Page, London 1984, pp. 50–54.

John N. Warfield

Inst. for Advanced Study in the Integrative Sciences
George Mason Univ., Fairfax, Virginia (USA)

Artificial philosophy and the decline of technology

Abstract

Historically, there has been a time lag between the conception of a technology and development of the underlying science that thoroughly explains the technology.

Until the advent in the 20th century of large-scale systems that involve significant hazards to large segments of the public, and which produce disasters that involve significant loss of life or major economic losses or both, the time lag did not warrant major attention. Now that these ***invasive systems*** *are demonstrably the cause of a growing number of* ***disasters****, this time lag is appropriately designated as the* ***hazard gap****.*

It is shown that the hazard gap stems from an ***artificial philosophy*** *that energizes advocates and designers of invasive systems. Unless genuine philosophy replaces artificial philosophy as key part of the basis for mitigating the impact of the hazard gap, technology will decline as a force for the improvement of life.*

Resumo

Historicamente, sempre existiu um atraso temporal entre a concepção tecnológica e o desenvolvimento da ciência de base que explica em profundidade a tecnologia.

Este atraso temporal não suscitou atenção de maior até ao advento, no século XX, de sistemas de grande escala que envolvem riscos significativos para largos segmentos da população e que produzem desastres susceptíveis de implicar perdas importantes de vidas, de recursos económicos ou de ambas as espécies. Agora que estes ***sistemas invasores*** *são, comprovadamente, a causa de um crescente número de* ***desastres****, aquele atraso temporal é designado apropriadamente por* ***intervalo de risco****.*

É mostrado que o intervalo de risco se alimenta de uma ***filosofia artificial*** *que estimula defensores e congeminadores de sistemas invasores. A menos que uma filosofia genuína substitúa a filosofia artificial como componente chave de bases para se mitigar o impacto do intervalo de risco, a tecnologia não deixará de declinar como força de melhoria da vida.*

1.0 Preliminaries

It is known to all that modern natural language is ambiguous, not only because ambiguity is inherent in terms, but also because of the existence of multiple (and sometimes inconsistent) definitions of terms. This paper uses a certain core set of defined ideas to underlie its content. Some of them have been given the principal definition from a dictionary. Others have been taken from published articles or books. Still others have been constructed specifically to clarify the arguments given in this paper. The following

TI & Sociedade / IT & Society, Simpósio Internacional de Lisboa / Lisbon International Symposyum, 1989 542/550
1992, APDC & SPF, Lisboa

definitions apply in this paper (except in quotations of other authors, where the terms reflect those author's meanings):

Abduction. The form of inference introduced by C. S. Peirce (1) as the only form that creates new ideas, especially hypotheses.

Belief. That state of mind which, according to C. S. Peirce (1) is the origin of rules of action, which themselves underlie habits.

Casuistry. The determination of right and wrong in questions of conduct or conscience by the application of general principles of ethics.

Disaster. A situation involving substantial loss of life or of life-sustaining resources (wealth, natural resources, environment) or both.

Generic Design Science (2). A science which presents the foundations, theory, and methodology of non-area-specific design; as a supplement to Specific Design Sciences that are allied with academic disciplines.

Habit (1). Behavior governed by ingrown rules of action stemming from belief whose origins no longer are readily available to the conscious mind.

Hazard Gap. The (unknown) period of time (which may be months, years, decades, or even centuries) between (a) the introduction of a potentially hazardous new system that involves new technologies and (b) the adequate development of the science that underlies and explains those technologies.

Hubris (3). Excessive pride or arrogance (and, in the Greek version, violence).

Intuition. An unarticulated and, usually, unarticulatable, basis for a decision.

Invasive System. A system having these properties: (a) it may affect many people who rendered no approval of its influence on their lives, (b) it is large in scale, (c) it is expensive, (d) the costs are often borne by people who did not accede to its introduction, and (e) it has the potential for causing **disaster** (as defined above).

Philopraxia. The love of action (as contrasted with philosophy).

Philosophy. The love of wisdom (as contrasted with philopraxia).

Possible Decision Bases. The set of concepts honored by archival discussions which have been put forth as possible foundational explanations of human decisions (as opposed to ad hoc or more superficial explanations).

Praxiology. The archival body of knowledge that results from the efforts of those who exhibit philopraxia.

Recursive Order. A count of the number of times a concept appears in nested form within a statement.

Representation (4). An **articulated** description of a situation or an aspect of a situation (as opposed to an implicit sense of that situation or aspect).

Rule of Action (1). A belief embedded in the mind which is automatically invoked in a situation to govern individual behavior.

Science (2). An organized body of knowledge consisting of foundations, theory, and methodology, in which the foundations steer the theory and the theory steers the methodology; which has applications to a class of situations; which is continually being evaluated in those situations; and which is modified according to those evaluations.

Sophology. The body of archival knowledge that results from the efforts of those who exhibit philosophy.

Subsume. To integrate within a larger framework.

Subsumptive Transformation (5). A process that subsumes several concepts enroute to the production of a new framework of thought.

Supersume. To assign to a concept a significance, meaning, or context that

substantially exceeds its proper scope; thereby denying an appropriate subsumption of that concept, and propagating underconceptualized contexts.

Supersumptive Transformation. A process that transforms a concept to a status well beyond its true status.

System [Willard Gibb's definition] "Any portion of the material universe which we choose to separate in thought from the rest of the universe for the purpose of considering and discussing the various changes which may occur within it under various conditions"(6).

Technomyopia (5). A form of mental dysfunction characterized by inability to think about things that lie even slightly beyond the native imagination of the affected individual.

Transcendentalism. The belief that knowledge of reality is derived from intuitive sources rather than from objective (sensory) experience.

2.0 The Hazard Gap

2.1. *Invasive Systems.* It is our central concern to limit the sphere of our discussion to a particular class of decisions: decisions about the advocacy and design of what we will call *invasive systems.* Examples of invasive systems that have already had negative impact upon many people are: the Chernobyl nuclear plant; the Bhopal chemical plant; the Challenger space vehicle; the Stealth bomber; the savings and loan associations in the United States that went bankrupt and are in the process of being "bailed out" by the taxpayers and the healthy savings and loans; the tanker that went aground in Valdes, Alaska; governments that insist on continuing regional wars that make life miserable for many; the Third Reich and the naval technological system that encouraged U. S. sailors to shoot down an Iranian passenger aircraft.

Typically invasive systems are created by people , without any clear disciplining standards (7, 8), who are much fewer in number than those potential victims of the failure of such systems.

Technologies in the invasive systems are constantly misrepresented by those who advocate them. Among the reasons are: self-interest, ignorance, fear of what will happen if they are not used, genuine belief in the merits of the technologies, and combinations in varying proportions of the foregoing. Many of the advocates of invasive technologies and systems no doubt are driven by combinations of these factors. Whatever mix is involved for any one advocate, we can almost always be sure that two factors will be included: *self-interest and ignorance* of the unknown fundamental aspects of the technology.

2.2. *Decisions Concerning Invasive Systems* (7). It is our purpose here to inquire into the *decision situation* in regard to the design of such systems. Whatever conclusions may be drawn are not necessarily relevant to non-invasive systems or other targets of design activity.

2.3. *Technology Leads Sciences.* Ignorance about invasive systems and their technologies differs only in scope and potential impact from the historical ignorance that has accompanied new technologies. As Vickers (9) has stated, "...throughout almost the whole of human history, technology has progressed with an uncanny ignorance of the scientific principles which were guiding it." The hazard gap is not a new concept (10) -- it is only the scope of the potential consequences that is new, and creates the necessity to explore the decision basis for advocating and designing such systems.

Continuation of the growth and extent of invasive systems likely will cause the decline of technology worldwide, as the public begins to sense that the human and financial costs

are too great compared to the benefits. Overreaction will prevent or slow the introduction of truly worthwhile technology, due to public inability to sort it out from the rest. It is essential now to reexamine the system of beliefs underlying invasive systems, and take corrective measures. We have to engage in "framebreaking and remodeling" in order to correct at least some percentage of the abuse potential of invasive systems. *Even small percentage gains will repay the effort.*

3.0 Framebreaking Orientation

Our orientation toward framebreaking is comprised of several parts. Let us address them singly.

3.1. *The Doctrine of Fundamentals.* We shall assume the validity of what we call the Doctrine of Fundamentals: namely, that *every human decision and action is grounded in fundamentals which, however, are not necessarily within the scope of awareness of those individuals.* This Doctrine is a matter of sophology. As Rorty (11) has indicated, "uncovering the presuppositions of those who think they have none is one of the principal means by which philosophers find new issues to debate".

3.2. *The Doctrine of Necessity* (12). We shall further assume the validity of what we call the Doctrine of Necessity: namely, *that there is an ordering to concepts, such that if a concept A cannot be understood without understanding another concept B, then necessarily B is prior to A; and through the sequential consideration of this Doctrine, sequences can be discovered that reveal progressively more fundamental concepts, continuing in a sequence of priors leading to a "deep logic"* (2, 5).

We have already published considerable experimental evidence to show that whenever complex issues are considered by people, they are able to generate this deep logic under carefully designed conditions, and the form that it takes indicates clearly that people generally are unable to do this without such conditions (5). Moreover we have shown that, at the outset of investigation, there is almost always substantial diversity of belief concerning relative importance of factors in these issues (5). We attribute the diversity to issue complexity, reflected in the absence of clear understanding of the sequences of priors.

3.3. *The Doctrine of the Implicit.* The Doctrine of the Implicit holds that (a) in essence, implicit bases for decisions are equivalent to secret bases for decisions, (b) implicit or secret bases have in common the capacity to form a shield against discovery and correction of misperceptions and underconceptualization, thereby promoting preventable disasters and (c) implicit or secret bases guarantee suspicion and distrust (12). Advocates and designers of systems with implicit or secret bases convey their indifference to the stated consequences, thereby making them all the more culpable when the systems or designs lead to disaster.

3.4. *The Doctrines of Origins of Belief.* Robert Coles in describing the work of Charles S. Peirce said that it "straddles delicately the world of faith and the world of precise scientific observation". In reflecting a balanced sensitivity to the human condition and the world of science, Peirce's ideas attained a unique credibility. In his view, the fixation of belief (1) took place through precisely four methods: (a) the Method of Authority, (b) the Method of Tenacity, (c) the Method of Metaphysics, and (d) the Method of Science. Of these, he felt that only the last could stand the test of time.

Peirce felt that all belief accrued through inference, but he distinguished four distinct types (1): immediate perception, abduction, deduction, and induction. Of these, only abduction creates anything basically novel, being the basis for the generation of hypotheses (which can later be tested).

Other concepts set forth in sophology as bases for belief and, especially, for decision,

include (e) transcendentalism (the singular rule of intuition), (f) casuistry (the singular rule of ethics), and (g) esthetics (the singular rule of sensory appeal).

In today's disciplines we see, as competing bases for belief, decision, and guidance; these highlighted areas: (h) sophology (the meaning of wisdom), (i) praxiology (the meaning of action) and (j) technology (the meaning of technique).

3.5. *Relevance of the Doctrines to Decision Bases for Invasive System Advocacy and Design*. Supersumption logically is a totally inadequate mode of sustaining a conceptual basis. Nevertheless it accounts for sustaining authority, tenacity, casuistry, esthetics, praxiology, and transcendentalism. Metaphysics and invasive systems technology are both sustained by transcendentalism which, itself, is sustained by supersumption. In other words, from among the ten decision bases listed above, **eight are at root sustained by supersumption**. Only two of them, namely science and sophology, do not share this fatal flaw.

The proof of the assertion that eight of the members of the set are sustained by supersumption is readily articulated: while one can easily find particular cases or instances where any member of the subset of eight has dominated human belief, apparently to good ends; it is also true that one can easily find instances where any particular member is irrelevant or even destructive in a particular case. The central conclusion is that without abusing any of these eight members by questioning the appropriateness of their membership in the set of decision bases, every one of them must be subsumed in the set of decision bases, every one of them must be subsumed in a larger framework, in order to lead to a general validity. But this is precisely what drives the evolutionary development of sophology and science.

Sophology is sustained by a never-ending quest for deep logic, and because of this, it marches along with and sustains science (2). The latter is constantly involved in subsuming prior knowledge within larger frameworks. The frequently-heard assertion to the contrary typically reflects thinking about short time spans, rather than the long-term advance of science. Moreover the often-stated argument that science is value-free is gradually becoming recognized as an intolerable absurdity that merely obscures the presence of ethics in the Universal Priors (12) to science, in the person of the scientists themselves. It is one thing to say that ethics is there, and another to see it take its rightful role in what goes on in science and in technology. Churchman (13) has stated the case in his illustration of another instance of supersumption, wherein praxiology is allowed to override ethical discussion.

3.6. *Artificial Philosophy*. As defined, philosophy is a state of mind involving the love of wisdom. Artificial philosophy is a substitute state of mind, in which the appearance of philosophy is given, but neither the love of wisdom nor the wisdom is likely to be there. Instead they are replaced with the supersumptive bases described earlier. Neitzsche (3), with considerable prescience, caught the right tone in saying that "our whole attitude toward nature, our violation of nature with the help of machines and the heedless ingenuity of technicians and engineers, is hubris ...". Soudek (14) summarized Solzhenitsyn's description of the personality aspects involved in engineers in the Soviet Union before and after the revolution. A reading of the before and after characteristics will lead to a congruence between the idea of supersumptive behavior that lacks a sound philosophical basis and subsumptive behavior that strives for such a basis. In today's world where each supersumptive actor uses the supersumptive behavior of another to justify such behavior, only framebreaking will alleviate the situation.

4.0 Remodeling Orientation

If framebreaking can be achieved by discrediting supersumptive assumptions and arrogant behavior that grow out of artificial philosophy, then remodeling must provide the subsumptive basis for a new model that will guide decisions about invasive systems. The basis for remodeling includes several components.

4.1. *Recursive Order in Scholarship and Practice.* Let Y represent "to make a decision" and let X be the subject of the decision. Then Y(X) can literally represent "to make a decision about X".

If X does not refer to the decision-making act itself, we can say that Y(X) represents a "first-order decision situation".

Suppose, however, that the act of decision-making is, itself, the subject X. Then we have the case Y(Y), "to make a decision about making a decision". We call this a "second-order decision situation".

A first-order decision situation is the kind that habitual behavior typically manages transcendentally (i.e., intuitively) through artificial philosophy.

A second-order decision situation not only questions how a decision will be made, but invokes the recursive need to know an answer in order to arrive at an answer. The questions: "How should one make a decision?" itself requires a decision. The terms "nesting" and "recursion" apply to this and higher-order situations.

Similar ideas have been discussed in "second-order cybernetics", where principles of cybernetics are invoked to explore cybernetics itself. Clearly iteration is a necessity in these recursive situations, but this is precisely what is involved in science itself, as Peirce has definitively indicated in discussing the fallibility of scientists and the need for a continuing community of evaluators (15).

Willingness to explore thoroughly the bases for decision-making is necessary but not sufficient to overcome the artificial philosophy that is destroying the social fabric. It is also necessary to explore thoroughly the processes that are used, and to open them to the full array of relevant information needed to construct sound decisions and system designs. Second-order considerations are also relevant to process selection. The study of an appropriate process to select an appropriate process, carried out subsumptively in the light of what is already in the literature, is critical as well.

Third-order recursion may be needed, or even higher order, until a model of action supported by sophology and science is sufficiently high in quality that it is beyond improvement in the near term.

4.2. *A Clear Perception of What it Means to Know.* "To know is to represent accurately what is outside the mind; so to understand the possibility and nature of knowledge is to understand the way in which the mind is able to construct such representations. Philosophy's central concern is to be a general theory of representation, a theory which will divide culture up into the areas which represent reality well, those which represent it less well, and those which do not represent it at all (despite their pretense of doing so)", according to Rorty (11). To this we add that Peirce's concept of semiotics, demands that representations shall not be evaluated only by those who create them. Rather the full spectrum of what is required to communicate must be taken into account in judging the quality of representations. It is not enough simple to construct a representation of something. It is necessary to assess the capacity of the human mind to comprehend that representation--a quality control condition that demonstrably has escaped the majority of advocates and designers of invasive systems.

4.3. *Adopting New Habits.* According to Peirce (1) "Belief... involves the establishment in our nature of a rule of action, or, say short, a *habit*" Boulding (16) said "... there are some simplifications which may be advantageous. Habit is a good example. Habits are economizers of decisions." We should not expect decisions perpetually to be carried out in higher-than-first-order decision situations. The alternative is to adopt new habits. The conscious adoption of new habits can evolve from outcomes of recursive studies and the choice of sophology instead of artificial philosophy as a decision base. These studies can justify a selection of the new processes and learning situations needed to upgrade substantially the quality of decisions and designs. And finally through the practice of the outcomes of such studies, habits can accrue, freeing the mind for whatever non-habitual behaviors it may seek to pursue.

4.4. *Distinguishing Science from Technology.* To attain the clarity of thought needed for the remodeling, it is necessary to constantly distinguish sharply between science and technology (17, 18). Northrup has facilitated this distinction by noting that:

"Science proceeds in two opposite directions from its many technical discoveries. It moves forward with the aid of exact mathematical formulation to new applications, and backward with the aid of careful logical analysis to first principles. The fruit of the first movement is applied science; that of the second, theoretical science. When this movement toward theoretical science is carried through for all branches of science we come to first principles and have philosophy."

4.5. *Distinguishing Da Vinci from Comte.* It is always possible that some will confuse the use of logical analysis with the positivist views of Comte. Such confusion stems from a built-in belief that creativity and logic are incompatible. Any study of Da Vinci's work will show the subsumption of these two valuable concepts. It is especially important to make major revisions in the education of professionals, in order to replace supersumptive thought with subsumptive thought, and to incorporate creativity and logic within a new model, rather than pretending that they are incompatible (19).

4.6. *The Sciences of the Artificial.* While Simon's famous work, *The Sciences of the Artificial*, may have induced some to believe that sciences of the artificial exist, no one should think that Simon himself made claims that adequate sciences for dealing with system design were available. On the last page of the body of his book (20), he made clear that in dealing with systems it would be necessary to provide "substance...to go with the name" in regard to a science of systems.

No one should think that artificial philosophy can be a proper base for a science of the artificial.

5.0 Conclusions

Artificial philosophy currently provides a widely-ingrained decision basis for advocacy and design of invasive systems. The performance of these systems shows its inadequacy, and promotes the decline of technology. Framebreaking and remodeling must lead to a new basis for advocacy and design of invasive systems, if they are going to continue to be imposed in society.

6.0 Acknowledgements

Appreciation is extended to my long-time colleague, Dr. Alexander Christakis, for his help in formulating some of the ideas given in this paper.

7.0 References

(1) T. A. Goudge, *The Thought of C. S. Peirce,* New York: Dover, 1969.
(2) J. N. Warfield, *A Science of Generic Design,* forthcoming, 1989.
(3) F. Nietzsche, *Genealogy of Morals,* III, p. 9.
(4) Richard Rorty, *Philosophy and the Mirror of Nature,* Princeton, NJ: Princeton University Press, 1979.
(5) J. N. Warfield, "Underconceptualization", *Proceedings of the Conference on Support, Society, and Culture: Mutual Uses of Cybernetics and Science,* (G. De Zeeuw and R. Glanville, Eds.), Univ. of Amsterdam, March, 1989, 15–39.
(6) Muriel Rukeyser, *Willard Gibbs,* Woodbridge, Conn.: Ox Bow, 1988 (reprint of the 1942 book published by Doubleday, Doran).
(7) J. N. Warfield, "What Disciplines Large–Scale System design?", *Proceedings of the 1987 Conf. on Planning and Design in Management of Business and Organizations* (P. C. Nutt, Editor), New York: American Society of Mechanical Engineers, 1987, 1–8.
(8) G. Ostberg, "Evaluation of a Design for Inconceivable Event Occurrence", *Materials & Design* 5, April/May, 1984, 88–93.
(9) G. Vickers, *Responsibility--Its Sources and Limits,* Seaside: Intersystems, 1980.
(10) J. B. Conant, *Two Modes of Thought*, New York: Trident, 1964.
(11) Richard Rorty (Ed.), *The Linguistic Turn*, Chicago: The Univ. of Chicago Press, 1967.
(12) J. N. Warfield, "Implicit Aspects of Much Systems Thinking", *Systems Research* 5(4), 1988, 333–342.
(13) C. West Churchman, *Thought and Wisdom,* Seaside, CA: Intersystems, 1982.
(14) I. Soudek, "The Humanist Engineer of Aleksandr Solzhenitsyn", *IEEE Technology and Science Magazine*, 3(5), 1986, 19–22.
(15) Karl–Otto Apel, *Charles S. Peirce: From Pragmatism to Pragmaticism,* Amherst: Univ. of Massachusetts Press, 1981.
(16) S. Aida et al, *The Science and Praxis of Complexity,* Tokyo: The United Nations University, 1985.
(17) J. N. Warfield, "The Domain of Science Model: Evolution and Design", *Proc. Society for General Systems Research,* Salinas: Intersystems, 1986, H46–H59.
(18) F. S. C. Northrup, *The Logic of the Sciences and the Humanities,* Westport: Greenwood Press (reprint of the 1947 Macmillan book), 1979.
(19) J. L. LeMoigne, "The Paradoxes of the Contemporary Engineer", *European Journal of Engineering Education* 6, 1981, 105–115.
(20) H. A. Simon, *The Sciences of the Artificial*, Cambridge, MA: MIT Press, 1967.

CORRIGENDA

Pág. / Linha		Onde se lê	Leia-se
34	11	*Responsability*	*Responsibility*
76	17	converations	conversations
145	41	An ⊢	An ⊢ C
145	42	A1,...,	A1,..., Ai
145	45	Cn, B1	Cn, B
155	19	intencionality	intentionality
161	44	emphasis	emphasize
219	3	fixes	fixed
229	9	com também	como também
271	26	*"percursores"*	*"precursores"*
306	14	caractistica	caracteristica
326	23	h is given	ℵ is given
508	19	*Understading*	*Understanding*
512	39	1973	1793

Na reprodução do "slide" da Fig. 1, p. 318, resultaram ilegíveis os dizeres seguintes, incluídos nas duas molduras rectangulares superior e inferior, respectivamente:

INSIDE WORLD (non-observable, immaterial)
MIND
PRINCIPLE OF ONE (Synthesis, levelling)

OUTSIDE WORLD (observable)
MATTER, ENERGY
PRINCIPLE OF TWO (analysis, separation)

Por estes lapsos de composição e outros que não tenham sido detectados apresentam-se desculpas aos autores e aos leitores.