Maria Eunice Quilici Gonzalez

Informação e Conhecimento Comum

Uma análise sistêmica dos processos criativos auto-organizados.

UNESP-Universidade Estadual Paulista
Faculdade de Filosofia e Ciências
Campus de Marília

Informação e Conhecimento Comum

Uma análise sistêmica dos processos criativos auto-organizados.

Maria Eunice Quilici Gonzalez

Coletânea de artigos apresentada como parte dos requisitos exigidos para o concurso de título de Livre-Docente na disciplina de Teoria do Conhecimento

1410001627

Fac. de Filosofia e Ciências - BIBLIOTECA - Campus de Marília

Agradecimentos

Aos meus pais, irmãos e irmãs pelo apoio incondicional
nos meus caminhos e descaminhos;
Para Mariana Broens, Betty Milidoni, Myrna Desidério e Edna Bonini,
sem as quais este concurso não teria, jamais, se materializado;
Para Pim, que me ensina a preciosa arte do ajuste na vida;
Para Zeljko Loparic, Balthazar Barbosa Filho e Noel Sharkey,
meus mestres queridos que me iniciaram no universo filosófico interdisciplinar.
Para o querido Michel Debrun e os meus colegas do CLE e do GAEC,
com quem aprendi a perceber a beleza da auto-organização nos (meus) caminhos caóticos.

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

Nada é mais importante do que
observar as origens da invenção,
as quais são, na minha opinião,
mais interessantes que as próprias invenções.

Leibniz

Apresentação

A presente coletânea foi estruturada a partir da seleção de textos sobre Informação, Criatividade e Conhecimento Comum que elaborei, nos últimos quinze anos, como resultado de pesquisas realizadas junto ao *Grupo de Estudos Cognitivos (GAEC)* da Unesp, campus de Marilia, do *Grupo CLE de Auto-Organização,* do Centro de Lógica, Epistemologia e História da Ciência, da Unicamp e, mais recentemente, junto ao Nijmegen Institute for Cognition and Information (NICI), da Universidade Católica de Nijmegen, Holanda, em colaboração com o professor Pim Haselager. Em tais textos investigamos a natureza dos processos de aquisição do conhecimento comum a partir de uma perspectiva interdisciplinar, no domínio da Ciência Cognitiva e Filosofia da mente.

Dada a sua natureza interdisciplinar, a presente coletânea reúne elementos de várias áreas do saber, principalmente da Filosofia, Epistemologia, Psicologia, Ciência Cognitiva, Física e da Ciência da Informação. Além disso, os dois últimos artigos desta coletânea foram elaborados em parceria com pesquisadores de áreas distintas do saber que, tendo um objetivo comum, compartilham informação sobre um mesmo tema de estudos.

A pesquisa *interdisciplinar*, tal como a compreendo, requer a satisfação de pelo menos duas condições: (a) o abandono de qualquer pretensão á busca de verdades absolutas e a adoção de um *método perspectivista* de análise que possibilite o estudo de um mesmo tema a partir de diferentes visões de mundo, sem recair no relativismo; (b) a especificação de uma abordagem integradora, sistêmica, dos fenômenos que unifique a multidimensionalidade da experiência cognitiva. Foi justamente na tentativa de satisfazer estas duas condições que desenvolvemos grande parte de nossa pesquisa sobre a natureza dos processos de aquisição do conhecimento, cujos resultados parciais apresentamos aqui em três etapas.

Na Parte I introduzimos alguns dos problemas centrais da Teoria do Conhecimento relacionados à natureza do conhecimento e sua possibilidade de fundamentação racional. Como tais problemas têm sido objeto de análise filosófica por muitos séculos, remontando pelo menos aos gregos antigos (sem mencionar as preciosas reflexões filosóficas da

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

tradição oriental), iniciamos nossa investigação com uma análise do diálogo platônico *Teeteto* sobre a distinção entre conhecimento, sensação e opinião verdadeira. Conforme ressaltamos nesse capítulo, o texto platônico nos deixa sem uma solução para o problema da definição e justificação racional do conhecimento, uma vez que tal justificação parece, inevitavelmente, necessitar o auxílio do conhecimento que se procura definir ou justificar. Tal dificuldade, que envolve uma circularidade indesejável, ficou conhecida na tradição como *O problema do Teeteto*. Esse problema é discutido no capítulo 1, que apresenta também uma tentativa de sua resolução proposta por Fred Dretske, na obra *Knowledge and the flow of Information.* O mesmo problema foi sucintamente tratado no último capítulo desta coletânea (no artigo *Conhecimento comum e auto-organização*), sendo desenvolvido no *Capítulo I – Conhecimento e Informação.* Neste capítulo, após estabelecer um paralelo entre aspectos do pensamento platônico e a abordagem informacional dretskeana, na tentativa de resolver o problema do Teeteto, argumentamos que Dretske deixa sem solução, assim como Platão, o problema em questão.

Concluímos o Capítulo 1 com uma crítica às tentativas de definição e justificação do conhecimento em termos de crenças verdadeiras racionalmente justificadas e propomos uma análise do conhecimento em termos da dinâmica de geração, estabelecimento e alteração de hábitos constitutivos da ação individual e coletiva, na difícil tarefa de sustentação da vida. Essa dinâmica, como argumentamos no capítulo 2, está na raiz do pensamento sistêmico criativo.

Nesse contexto, introduzimos os pressupostos fundamentais da Teoria Geral dos Sistemas, tal como originalmente formulados por Ludwig von Bertalanffy (1969), da Teoria da Auto-orgnização proposta por W.R. Ashby (1962), Michel Debrun (1996), Herman Haken (1999), entre outros, e da Metodologia Perspectivista elaborada por Donald Peterson (1996). Tais pressupostos estão presentes nos textos: *Complexidade e Criatividade: uma abordagem sistêmica dos processos auto-organizados*, bem como em *Auto-orgnização e Perspectivismo: algum acréscimo à Ciência Cognitiva?,* publicados em 2004 e 1996, respectivamente.

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

Em *Complexidade e Criatividade: uma abordagem sistêmica dos processos auto-organizados,* apresentamos inicialmente a estrutura conceitual básica da sistêmica aplicada ao domínio da cognição, que inclui: a busca de leis gerais unificadoras dos fenômenos da natureza, a caracterização do princípio de mutualidade entre sujeito e meio ambiente, bem como os processos de emergência espontânea de novas formas de organização no fluxo dos eventos que constituem os sistemas dinâmicos. Tendo como pressuposto tal estrutura, argumentamos em defesa da hipótese que o processo criativo de obtenção do conhecimento pode ser definido a partir da dinâmica auto-organizadora que se estabelece no domínio da ação de sujeitos, incorporados e situados (em contraposição à concepção transcendental de sujeito), em um meio ambiente que possui informação disponível para tornar coerente a existência dos mesmos.

Em *Auto-organização e Perspectivismo: algum acréscimo à Ciência Cognitiva?*, discutimos inicialmente alguns dos pressupostos metodológicos subjacentes à abordagem vigente na Ciência Cognitivista dos anos noventa, nos estudos dos processos cognitivos, e endereçamos algumas criticas formuladas por filósofos da mente, como John Searle, ao funcionalismo computacional que dominava (e ainda domina) boa parte do projeto explicativo dessa ciência. Em seguida, argumentamos que a pesquisa cognitivista poderia ser enriquecida através da adoção das hipóteses da Teoria da Auto-organização, em especial aquelas propostas por Michel Debrun na sua caracterização do sujeito cognitivo. Finalmente, os preceitos da metodologia perspectivista desenvolvida por Donald Peterson foram propostos como uma alternativa para a ampliação do limitado universo metodológico da então vigente Ciência Cognitiva.

Em resumo, na Parte I introduzimos problemas centrais da epistemologia e indicamos possíveis contribuições da teoria dos sistemas e da auto-organização para o estudo desses problemas, as quais, argumentamos, se aplicam nas mais diversas formas de análise dos processos de auto-organização da ação e do pensamento criativos. O objetivo dos textos foi introduzir o universo da sistêmica e do perspectivismo ao leitor interessado nos problemas investigados na ciência cognitiva, sem a pretensão de oferecer uma análise detalhada da natureza da criatividade ou dos problemas centrais da ciência cognitiva.

unesp

Uma análise mais elaborada da natureza da criatividade foi posteriormente desenvolvida, em colaboração como professor Pim Haselager, no texto Creativity: Surprise and Abductive Reasoning , publicado no Journal of Semiotics, 2004. Nesse artigo explicitamos a dinâmica do raciocínio abdutivo, tal como formulado por Peirce e Hanson nos seus estudos da lógica da descoberta científica, dando especial ênfase à relevância epistemológica da experiência da surpresa no processo de criação. Nele discutimos também a possibilidade de máquinas complexas, como robôs auto -organizados, realizarem ações criativas.

Na **Parte II**, analisamos, no contexto da Ciência Cognitiva e da Filosofia da Mente, questões epistemológicas relativas à natureza do conceito de informação e seu emprego nos processos auto-organizados de aquisição do conhecimento. Em especial, focalizamos: (1) A natureza das representações mentais; (2) A metodologia de investigação dos modelos conexionistas das representações mentais e (3) O problema da relação mente-corpo. Este percurso é apresentado através dos textos: *O nascimento da Ciência Cognitiva e suas raízes na física do século XIX*; *Um estudo cognitivo informacional das representações mentais; Ação, causalidade e ruído nas redes neurais auto-organizadas*; *Informação e Cognição: uma proposta de (dis)solução do problema mente-corpo.*

Em *O nascimento da Ciência Cognitiva e suas raízes na Física do século XIX*, propomos uma análise histórica do método de análise adotado na Ciência Cognitiva e argumentamos que este possui suas raízes na Física do século XIX, em especial no que diz respeito ao emprego de métodos experimentais e recursos matemáticos para explicar e prever o comportamento do sistema sob investigação. Dada a divulgação restrita dos estudos da Ciência Cognitiva no Brasil, e considerando nosso propósito interdisciplinar, apresentamos neste artigo, como de praxe, uma breve introdução à Ciência Cognitiva, enfatizando as duas áreas centrais que a constituíam na década de 1990, a saber, Inteligência Artificial e Conexionismo. A novidade que julgamos relevante no presente artigo diz respeito à aplicação de hipóteses da Teoria Matemática da Informação nos modelos de redes neurais artificiais. Nesse texto esboçamos uma teoria, cujo desenvolvimento se daria posteriormente em nossas investigações acerca da natureza da informação na estruturação dos sistemas cognitivos.

unesp

Em *Um estudo cognitivo informacional das representações mentais,* investigamos o conceito de informação inspirado na noção de informação ecológica desenvolvida por Gibson (1996) com o objetivo de explicitar o processo de formação de padrões de atividade em redes neurais auto-organizadas. Dividida entre a tradição cartesiana representacionista da mente e o anti-representacionismo radical gibsoniano, argumentei em defesa da possível anuência de Gibson aos modelos conexionistas de representação que incorporam padrões informacionais disponíveis no meio ambiente. Atualmente, porém, ao ler este trabalho, vejo com ceticismo tal possibilidade, uma vez que, como toda análise da cognição centrada em redes neurais, ela perde o horizonte dinâmico da relação sujeito-corpo-meio ambiente, considerado essencial por Gibson para o entendimento da relação informacional experienciada por organismos situados no meio.

Em *Ação, causalidade e ruído nas redes neurais auto-organizadas* introduzimos os pressupostos fundamentais do neo-conexionismo que ressaltam a capacidade de aprendizagem e auto-organização dos modelos de redes neurais explicativos da atividade inteligente. O objetivo central deste artigo é verificar em que medida os processos de auto-organização ocorrem de fato em tais modelos. Após apresentar os pressupostos básicos da teoria da auto-organização tal como proposta por Ashby (1962), Atlan (1979) e Debrun (1996), dentre outros, argumentamos que apenas algumas dessas propriedades podem ser a eles atribuídas. Isto porque tais modelos não possuem princípio motor autônomo para a ação, o que os torna bastante frágeis enquanto modelos da atividade inteligente.

Finalmente, em *Informação e Cognição: uma proposta de (dis)solução do problema mente-corpo*, retomamos o conceito de informação, agora no contexto da epistemologia proposta por Dretske (1981), que enfatiza a noção de dependência ou correlação nômica na caracterização do conceito de *conteúdo informacional*. Procuramos indicar os passos desenvolvidos por Dretske na sua proposta de formulação das condições necessárias e suficientes para a formulação do conceito de conhecimento perceptual, oferecido como alternativa como a concepção tradicional do conhecimento em termos de crenças verdadeiras justificadas. A partir das idéias de Dretske, sugerimos uma aplicação do

unesp

conceito de informação no estudo do problema mente/corpo, sugerindo a dissolução deste problema a partir da perspectiva informacional.

Na **Parte III,** que reúne os textos**:** *Auto-organização, Autonomia e Identidade Pessoal,* e *Conhecimento Comum e Auto-organização*, questionamos a tradição platônica na caracterização do sujeito cognitivo. Tal concepção caracteriza o sujeito cognitivo em termos de estruturas simbólicas abstratas que são manipuladas por mecanismos de percepção passiva, tal como no mito da caverna – segundo o qual os indivíduos são seres acorrentados, receptores passivos de sombras que lhes chegam através de imagens luminosas projetadas na parede da caverna. Uma visão alternativa é proposta em termos da Teoria da auto-orgnização e das hipóteses da Ciência Cognitiva Dinâmica.

.Em *Auto-organização, Autonomia e Identidade Pessoal,* elaborado em parceria com Mariana C. Broens e Juliana Serzedello, analisamos o problema da Identidade Pessoal, na sua formulação atual na Filosofia da Mente. Com base nos princípios de causalidade circular e de condicionalidade dos estados disposicionais, emergentes da dinâmica presente no acoplamento entre percepção e ação, propomos uma estratégia de análise do processo de formação de identidade pessoal em sistemas cognitivos. Especial ênfase é dada ao caráter auto-organizador do processo de geração de hábitos que constitui, em nosa concepção, a marca principal da identidade pessoal.

Em Conhecimento Comum e Auto-organização, desenvolvido em parceria com Pim Haselager, retomamos o problema platônico, apresentado no dialogo *Teeteto*, sobre a definição do conhecimento em termos de crenças verdadeiras justificadas. Argumentamos que o conhecimento do senso comum não pode ser propriamente concebido de tal modo, mas, sim, em termos de disposições. Tais disposições resultam de acoplamentos dinâmicos e auto-organizados entre o corpo de um ser em ação e o mundo.

Concluímos esta coletânea situando o debate em torno da natureza do conhecimento comum, no âmbito da Ciência Cognitiva e da Filosofia da Mente, indicando as raízes platônicas da concepção de conhecimento que, entendemos, estão presentes na Ciência Cognitiva tradicional (em especial nas bases do funcionalismo computacional).

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

Nosso propósito de pesquisa futura é aprofundar o estudo do papel cognitivo do corpo, entendido como uma unidade sistêmica, em termos de relações que se estabelecem entre agentes/sujeitos ativos e o mundo com o qual interagem informacionalmente através de suas ações histórica e ambientalmente estruturadas.

PARTE I

Conhecimento, Criatividade e Perspectivismo: Uma análise sistêmica

unesp

CONHECIMENTO E INFORMAÇÃO

Maria Eunice Quilici Gonzalez
gonzalez@marilia.unesp.br

Apresentação

A reflexão filosófica sobre a natureza do nosso conhecimento revela uma série de problemas que são objeto de investigação da Teoria do Conhecimento. Entre os mais conhecidos estão os seguintes: (1) O que é conhecimento? (2) Qual a diferença entre conhecimento e crença verdadeira? (3) Como é possível o conhecimento? (4) Quais são os fundamentos do conhecimento? (5) Qual é a diferença entre conhecimento científico e conhecimento comum? (6) Qual é a diferença entre conhecimento prático e conhecimento puro? (7) Quais são os critérios de verificação do conhecimento? (8) Em que diferem conhecimento e sensação?

Os problemas acima têm sido objeto de discussão na filosofia por séculos, remontando pelo menos aos gregos antigos, sem uma resolução consensual. A maioria desses problemas, relacionados à natureza do conhecimento e sua possibilidade de fundamentação, enfrenta a dificuldade central de se obter, racionalmente, os princípios a partir dos quais o conhecimento pode ser demonstrado, sem cair em um processo de regressão infinita. Disputas infindáveis entre dogmáticos e céticos ilustram o cenário filosófico em que se desenrolaram os debates acerca de tais problemas.

Neste capítulo, discorreremos sobre o problema da natureza do conhecimento em três momentos. No primeiro, ilustramos parte deste cenário no contexto da Grécia antiga, através da análise do diálogo platônico *Teeteto*, e apresentamos, em seguida, uma análise contemporânea do problema do *Teeteto*, tal como formulado por Fred Dretske, em "*Knowledge, and The Flow of Information*". Argumentamos que o dualismo platônico é traduzido na proposta informacional desenvolvida por Dretske em sua tentativa de solucionar o problema do Teeteto. Finalmente, na terceira parte, concluímos o capítulo com uma análise especulativa, sobre possíveis conseqüências da postura dualista – que caracteriza o trabalho filosófico através da busca de verdades – no estudo do conhecimento comum e da ação ambientalmente situada em agentes incorporados e dotados de uma história própria.



I - O Problema do *Teeteto*[1].

No *Teeteto*, Platão apresenta um diálogo sobre a natureza do conhecimento atribuindo grande ênfase à habilidade filosófica de diferenciar o verdadeiro do falso. Nesse contexto, Sócrates, um dos personagens que dialoga com Teeteto, se compara às parteiras (embora se julgue, naturalmente, superior a elas), cujo trabalho ele acredita ser propriamente realizado pelas mulheres que, ao atingirem certa idade, já não podem procriar mas conhecem, melhor do que as outras, quando uma mulher está grávida e o que fazer para ajudá-la no parto. Assim, no *Teeteto* encontramos a seguinte passagem:

> "Eis aí a função das parteiras; muito inferior à minha. Em verdade, não acontece às mulheres parirem algumas vezes falsos filhos e outras vezes verdadeiros, de difícil distinção. Se fosse o caso, o mais importante e belo trabalho das parteiras consistiria em decidir entre o verdadeiro e o falso..." (*Teeteto*, VI, p.10, 149b)

A suposta superioridade do parto das idéias em relação ao parto natural residiria, justamente, na sua capacidade de auxiliar a reflexão filosófica na difícil tarefa de determinar critérios de distinção entre o verdadeiro e o falso. Na seguinte passagem, Sócrates explica em que consiste sua alegada superioridade:

> "Sócrates -... A minha arte obstétrica tem atribuições iguais às das parteiras, com a diferença de eu não partejar mulheres, porém homens, e de acompanhar as almas, não os corpos, em seu trabalho de parto. Porém a grande superioridade da minha arte consiste na faculdade de conhecer de pronto se o que a alma dos jovens está na iminência de conceber é alguma quimera e falsidade ou fruto legítimo e verdadeiro. Neste particular, sou igualzinho às parteiras: estéril em matéria de sabedoria, tendo grande fundo de verdade a censura que muitos me as assacam, de só interrogar os outros, sem nunca apresentar opinião pessoal sobre nenhum assunto, por carecer, justamente, de sabedoria..." (*Teeteto*, VII, p.10, 150 c-d).

O pressuposto platônico, segundo o qual a tarefa do filósofo consiste em auxiliar na busca de verdades, deixou raízes profundas na tradição filosófica clássica e também nos projetos

[1] Este problema foi brevemente abordado no último texto componente da presente *Coletânea de artigos*.

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

atualmente em desenvolvimento na Ciência Cognitiva e Filosofia da Mente, conforme ilustraremos nos capítulos seguintes.

No presente capítulo, discutimos os problemas (2) e (8), enunciados na apresentação, sobre a diferença entre conhecimento e opinião verdadeira e, também, entre conhecimento e sensação. No que diz respeito ao problema (8), Platão desenvolve no diálogo citado uma reflexão sobre a concepção de Protágoras, segundo a qual o conhecimento não seria nada mais do que sensação. Protágoras, diz Sócrates no diálogo com Teeteto, "afirmava que o homem é a medida de todas as coisas, da existência das que existem e da não existência das que não existem." (*Teeteto*, VII, p. 11, 152 a)

Segundo a interpretação platônica do pensamento de Protágoras, sensação e aparência se eqüivalem, o que conduz à análise do conhecimento em termos do fluxo do movimento das coisas, tais como elas nos aparecem. O aspecto relacional do conhecimento, implícito nesta concepção, é, então, objeto de análise da investigação platônica, ilustrada, por exemplo, na seguinte passagem:

> " Sócrates - Considera o assunto, meu caro, do seguinte modo: inicialmente, com relação à vista, o que denominamos cor branca não é algo com existência própria, nem fora de teus olhos nem dentro de teus olhos, nem em qualquer outro local que se lhe assinalarem, pois se assim fosse, ela existiria num determinado lugar em caráter estável, deixando, por conseguinte, de formar-se" (*Teeteto*, X, p.14, 154a).

O caráter singular, relativo ao indivíduo, daquilo que lhe aparece, pode ser inferido a partir do pressuposto do conhecimento entendido como sensação, uma vez que as sensações parecem variar dependendo do estado daqueles que as experienciam. Aquilo que é sentido e percebido, por exemplo, como amargo por alguém pode ser percebido como doce por outrem, dependendo do estado físico de cada um. Nesse sentido, aquilo que é verdadeiro para um será falso para outro.

Ainda que as aparências ou percepções das coisas não possam ser identificadas às próprias coisas, o diálogo platônico conduz à conclusão de que a hipótese inicial, que identifica conhecimento à sensação, nos leva não apenas a inferir a não existência de objetos estáveis, como também a impossibilidade do conhecimento que ultrapassa os limites do indivíduo que vivencia tais sensações. Dado o caráter singular das sensações, a possibilidade de qualificá-las como conhecimento universal deve ser descartada.

unesp

O segundo problema tratado no *Teeteto* que nos interessa particularmente diz respeito à distinção entre conhecimento e opinião verdadeira (problema (2)), o qual está diretamente relacionado ao problema da distinção entre sensação e conhecimento.

No diálogo, Platão argumenta que os seres vivos captam as impressões singulares que atingem a alma através do corpo, mas é trabalho da alma sozinha apreender o que existe de comum às coisas; a essência das coisas não pode ser captada pelo corpo; cabe à alma discernir, julgar, e deliberar sobre os fatos. Assim, sentimos pelo tato a dureza e a moleza, o calor e o frio, mas apenas através do julgamento podemos discernir tal dualidade. A passagem abaixo ilustra esse ponto:

> "Sócrates - ... desde o nascimento, tanto os homens como os animais têm o poder de captar as impressões que atingem a alma por intermédio do corpo. Porém, relacioná-las com a essência e considerar a sua utilidade, é o que só com tempo, trabalho e estudo conseguem os raros a quem é dada semelhante faculdade." (*Teeteto,* XXX, p.47, 186b-c)

Aos raros, a quem se refere Platão, é dada a faculdade de julgar, através da apreensão das essências, e só a eles, que podem discernir entre o verdadeiro e o falso, é dada a capacidade de conhecer. O conhecimento se distinguirá da mera sensação e da opinião pela presença, neste seres, da capacidade de contemplar, através do exercício da dialética (termo sinônimo de Filosofia para Platão), as essências alcançadas pela alma. A passagem do diálogo abaixo ilustra, também, essa postura:

> "Sócrates – E poderá atingir a verdade de alguma coisa quem não alcançar a sua essência?
> Teeteto – Nunca!
> Sócrates – E do que não se alcança a verdade, poder-se á ter conhecimento?
> Teeteto – De que jeito, Sócrates?
> Sócrates – Naquelas impressões, por conseguinte, não é que reside o conhecimento, mas no raciocínio a seu respeito; é o único caminho, ao que parece, para atingir a essência e a verdade; de outra forma é impossível.
> Teeteto – Claro.
> Sócrates – E darás o mesmo nome aos dois processos, já que é tão grande a diferença entre ambos?
> Teeteto – Não seria justo.

unesp

Sócrates – Então, que nome dás ao primeiro, isto é, ao fato de ver, ouvir, cheirar e sentir frio ou calor?

Teeteto – O de sensação. Qual mais poderia ser?

Sócrates- A tudo isso dás o nome de sensação?

Teeteto – Forçosamente.

Sócrates – Ao que, conforme vimos, não é dado atingir a verdade, por isso mesmo que não nos conduz à essência.

Teeteto – Sem dúvida.

Sócrates – Como não atinge o conhecimento.

Teeteto – Não, de fato.

Sócrates – Sendo assim, Teeteto, não poderão ser a mesma coisa sensação e conhecimento.

Teeteto – Parece-me que não, Sócrates. Patenteou-se-nos agora que conhecimento é diferente de sensação. (*Teeteto*, XXX, p. 47, 186c-187a)

Em resumo, dado o caráter singular das sensações, por si sós elas não podem, na visão platônica, proporcionar o conhecimento. A ausência de alma capaz de filosofar, nas criaturas que não se enquadram entre os raros dotados dessa aptidão, condena a grande maioria dos seres vivos a viverem nas trevas, no que se refere à busca de conhecimento. Contudo, como mencionado, Platão acredita que mesmo aqueles que possuem a capacidade reflexiva necessitam de tempo, trabalho e estudo para alcançar a verdade.

É nesse contexto que se situa a análise platônica da distinção entre conhecimento e opinião verdadeira (problema (2)). Na sua inspiradora alegoria da cera, ele supõe que na alma há um tipo de cera: "... numas pessoas, maior, noutras, menor, nalguns casos, de cera limpa, noutros com impurezas, ou mais dura ou mais úmida, conforme o tipo. Sempre que queremos nos lembrar de algo visto ou ouvido, ou mesmo pensado, calcamos a cera mole sobre nossas sensações ou pensamentos e nela gravamos em relevo... Do que fica impresso, temos lembranças e conhecimento enquanto persiste a imagem, o que se apaga ou não pôde ser impresso, esquecemos e ignoramos." (*Teeteto*, XXXIII, p. 54, 191c)

As falsas opiniões ocorreriam principalmente quando confundimos aquilo que é conhecido ou percebido com aquilo que, apesar de ser percebido, não é conhecido; ocorreria um tipo de

unesp

desajuste entre sensação e pensamento. No caso das crenças verdadeiras, ao contrário, existiria um perfeito ajuste entre a opinião verdadeira e aquilo que é conhecido pela alma.

Neste cenário, haveria algum lugar para o problema (2), sobre a diferença entre conhecimento e crença ou opinião verdadeira? Platão argumenta que sim, iniciando uma cadeia de investigações cujos ecos ainda ouvimos atualmente, seja na filosofia clássica seja na filosofia da mente contemporânea.

Grosso modo, o argumento de Platão segundo o qual é possível se ter crenças verdadeiras mas, ainda assim, não se possuir conhecimento se inicia com a distinção entre "possuir" e "ter" conhecimento. A diferença entre "ter" e "possuir" é ilustrada metaforicamente, através do seguinte exemplo: " se alguém compra uma roupa, e, na qualidade de dono dessa roupa, não a usa, não diremos que ele a tem, mas que a possui" (*Teeteto*, XXXVI, p. 60, 197c).

A moral da história é que tem conhecimento aquele que não apenas possui crenças verdadeiras, mas sabe utilizá-las e expressá-las apropriadamente através das explicações racionais. Sem entrar em detalhes sobre o que vem a ser *explicação racional*, é suficiente para os nossos propósitos indicar a conclusão, aparentemente inevitável, da circularidade a que conduz o raciocínio platônico, no que diz respeito ao problema (2): O emprego apropriado das crenças verdadeiras parece, ele próprio, exigir conhecimento. O mesmo ocorre com as explicações racionais, cuja fundamentação parece exigir conhecimento. Admitindo que o conhecimento pode ser entendido como opinião verdadeira aliada à explicação correta (racional) e que esta requer conhecimento para ser qualificada como tal, Platão conclui:

> "Sócrates – Ora, será o cúmulo da simplicidade, estando nós à procura do conhecimento vir alguém dizer-nos que é a opinião certa aliada ao conhecimento, seja da diferença ou do que for. Desse modo, Teeteto, conhecimento não pode ser nem sensação, nem opinião verdadeira, nem a explicação racional acrescentada a essa opinião." (*Teeteto*, XLII, p.76, 209a)

Ao final dessa investigação, que nos deixa sem uma definição do conhecimento, Sócrates teria ido apresentar-se a seus juízes para responder às acusações de corrupção da juventude que lhe eram dirigidas. Seu projeto de definição do conhecimento a partir da busca de verdades, identificadas pela alma, que filtra a informação proveniente das sensações, parece conduzir a impasses

unesp

estabelecidos pela própria alma. O pressuposto platônico de uma ontologia dualista radical – segundo a qual podemos conhecer somente através da contemplação direta das idéias, quando a alma, ao se libertar do seu fardo corpóreo, está preparada para o exercício filosófico - , parece transformar esse exercício em um remédio, uma preparação, para a morte,[1] ocasião jubilosa para Platão, na medida em que a alma, liberta do corpo, volta a conviver com as próprias coisas ou idéias.

Argumentaremos no restante deste trabalho que o legado platônico da filosofia, entendida como um exercício de "purificação" da matéria, encontrou um terreno fértil na concepção funcionalista, vigente na Filosofia da Mente e na Ciência Cognitiva tradicional. Aliada aos recursos tecnológicos de nossa era, e às descobertas da Teoria da Informação, as raízes da filosofia platônica parecem ganhar terreno na biologia, na engenharia genética e nos estudos cognitivos interdisciplinares. As conseqüências dessa parceria nos parecem, no mínimo, objeto de grande preocupação para os amantes da vida. De modo a fundamentar esse argumento retomaremos, na seção seguinte, o problema do *Teeteto*, agora analisado sob o viés contemporâneo da Teoria da Informação desenvolvida por Fred Dretske.

II- O problema do *Teeteto* revisitado por Dretske

O problema platônico apresentado no *Teeteto,* sobre a distinção entre conhecimento e opinião verdadeira, tem sido reformulado atualmente da seguinte maneira: o que teria a mais um homem possuidor de conhecimento sobre um evento quando comparado a outro que possui uma opinião verdadeira, sobre o mesmo evento, mas não tem conhecimento? Ou, em outros termos: o que é que adicionado à opinião verdadeira gera conhecimento?[2] Inicialmente considera-se que o primeiro, mas não o segundo, terá evidência adequada sobre a ocorrência do evento em questão, enquanto que faltaria ao segundo tal evidência. Duas dificuldades fundamentais se colocam, entretanto, nessa estratégia. A primeira reside no estabelecimento daquilo que é evidente; a segunda pressupõe resolvida esta primeira dificuldade e adiciona a ela o problema da determinação do que seria evidência adequada

---

[1] Agradeço à Mariana Cláudia Broens por me fazer compreender que o platonismo com seu dualismo radical e a conseqüente negação da corporeidade pode ser visto como uma filosofia preparatória para a morte.

[2] Uma formulação do problema do *Teeteto* nestes termos é apresentada, por exemplo, por Chilsholm em "*Theory of Knowledge*".

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

para o caso em questão, sem ter que pressupor a existência da alma, que como vimos na seção anterior, teria a função de realizar tal discernimento.

No que diz respeito à primeira dificuldade, uma forma clássica de investigar a noção de evidência consiste em assumi-la como tal e se interrogar, através de uma dúvida metódica, sobre as razões que justificam o pressuposto dessa evidência como evidente. Essa estratégia metodológica, entretanto, pode ser problemática, pois pressupõe que a dúvida em questão expressa uma dúvida real, quando muitas vezes seu estatuto, enquanto evidência, está implícito no raciocínio no qual ela se coloca. Em geral, essa estratégia assume formas diversas do método platônico de explicitar o conhecimento já implícito no espirito do inquiridor.[3]

Admitindo-se que, de alguma forma, a primeira dificuldade na determinação do que seja evidente possa ser resolvida, confrontamo-nos com uma segunda dificuldade que por sua vez pode ser subdividida em três casos, já examinados por Platão no *Teeteto*:
(1) É possível ter evidência adequada sem, no entanto, ter conhecimento, (2) pode-se ter evidência adequada sem saber que se possui tal evidência e (3) o critério de relevância adotado para distinguir as evidências adequadas das evidências inadequadas requer conhecimento.

Há quem argumente, como Descartes, por exemplo, que as dificuldades acima são passíveis de resolução através de um método apropriado de análise que possibilite não apenas alcançar os princípios básicos sob os quais se fundaria o conhecimento, mas, também, que indique as etapas pelas quais o raciocínio explicativo desse conhecimento poderia ser desenvolvido. Em Gonzalez (1989) discutimos o método de análise e síntese, proposto originalmente por Pappus, e amplamente adotado pelos filósofos modernos e contemporâneos que se preocupam com a questão dos fundamentos do conhecimento. Não entraremos em detalhes sobre esse tópico aqui; nos deteremos, antes, na ilustração de uma abordagem contemporânea do problema do *Teeteto*, indicando os impasses a que ele parece conduzir. Nesse contexto, introduzimos a proposta informacional oferecida por Fred Dretske para solucionar esses impasses. Tal proposta servirá de fio condutor para a análise, que desenvolveremos no restante deste trabalho, sobre possíveis conseqüências, no plano do conhecimento comum, das

---

[3] Os diálogos platônicos ilustram amplamente esta situação. O mesmo poderia ser dito, por exemplo, de Descartes em sua busca de fundamentos seguros para a ciência. A evidência do *cogito*, como argumentam vários filósofos, entre eles Oswaldo Porchat Pereira, parece estar implícita nas etapas do raciocínio cartesiano que conduz à sua primeira verdade.

unesp

investigações sobre a natureza do conhecimento a partir da perspectiva informacional vigente no século XXI.

Em "*Knowledge and the Flow of Information*", Dretske propõe uma saída para a dificuldade (3) acima, reformulando o problema do *Teeteto* a partir da Teoria da Informação. Nessa perspectiva , a definição do conhecimento pode ser assim formulada:

S conhece / sabe que F se e somente se :
S possui crença verdadeira (**Cr**) sobre F
e **Cr** é sustentada, fundada, em informação

Inspirado na teoria matemática da comunicação (MTC, do inglês - *Mathematical theory of communication*), desenvolvida por Shanon e Weaver (1949) na área de engenharia, Dretske (1981, 1983) elabora a sua teoria tendo como pressuposto a hipótese de que a informação é um elemento objetivo, "cuja existência (enquanto informação) é (amplamente) independente das atividades interpretativas de agentes conscientes" (Dretske, 1983, p. 55). Contudo, diferentemente da MTC, cujo objetivo central é a transmissão eficiente de sinais, a teoria desenvolvida por Dretske propõe uma explicação para o aspecto significativo da informação presente na comunicação do senso comum.

É justamente a habilidade interpretativa, presente na comunicação significativa do senso comum, que constitui o objeto central de seu projeto naturalista dos processos mentais, responsáveis pelo conhecimento, os quais envolvem informação. No mesmo viés, Fodor (1998, p. 98) sugere que, de modo a evitar a circularidade, uma teoria materialista do significado deveria articular, em termos não semânticos, as condições suficientes para que um "*bit*" do mundo possa expressar (representar, ou ser verdadeiro sobre) um outro "*bit*" do mundo.

Para efetivar seu projeto naturalista, Dretske propõe uma *definição nuclear* de informação segundo a qual: um sinal carrega informação sobre o que ocorre em uma fonte (o seu conteúdo) se ele for capaz de reproduzir factualmente as relações que se estabelecem na fonte, tornando-as acessíveis para qualquer observador que se encontre em condições de recebê-las. Ainda que ele não exija uma correspondência biunívoca entre o sinal e o seu conteúdo (assim, por exemplo, o sinal que carrega a informação sobre um objeto grande não precisa ele próprio ser grande), a definição nuclear de informação pressupõe que os eventos que ocorrem na fonte impõem certas *restrições* ao sinal que carrega informação sobre eles, tornando-os nomologicamente dependentes.

unesp

Enquanto candidato à informação no sentido nuclear, um sinal deve refletir /transmitir fielmente aquilo que ocorre na fonte, de tal forma que um receptor, devidamente equipado, poderá conhecer, com probabilidade condicional = 1, ou vir a aprender o que lá ocorre, apenas observando esse sinal. Tal caracterização nuclear da informação é assim definida por Dretske (1981, p.45): "Um sinal ou estados de coisas contém informação sobre X justamente à medida que um observador *adequadamente situado* pode aprender algo sobre X consultando-os". (grifo nosso) Dessa forma, assegura-se que sinais específicos carreguem conteúdos informacionais específicos para um sujeito devidamente equipado para apreendê-las. (Como procuraremos argumentar na seção III, ainda que Dretske não faça (obviamente) referência à alma no seu projeto naturalista da mente, entendemos que, em certos casos, a posição privilegiada do receptor de informação, tal como caracterizada nesta definição, poderia bem ser igualada à posição dos "raros possuidores de alma reflexiva" aos quais se referia Platão no *Teeteto*).

De modo a garantir a correspondência entre o que ocorre na fonte e o sinal que a transmite Dretske (1981, p. 65) estabelece a seguinte condição: se existe uma correlação necessária entre os eventos X e Y, de tal forma que uma mudança específica em X produz uma mudança, igualmente específica em Y, com probabilidade condicional =1, e se tal mudança é registrada em Y, então Y carrega informação sobre X. A exigência da probabilidade condicional =1 expressa uma correlação nômica, legiforme, entre X e Y; ela constitui uma característica intrínseca das relações informacionais, na visão dretskeana. Essa exigência tem sido objeto de inúmeras críticas por parte dos leitores de Dretske, uma vez que o mundo contingente em que vivemos está repleto de falhas e ruídos, não comportando, assim, sua verificação. (Esse ponto será retomado ao final deste capítulo).

Ainda que a definição nuclear de informação admita a existência de informação que independe de um receptor consciente, a apreensão correta dessa informação envolve um processo de filtragem e solução por parte do receptor, que pode eventualmente se equivocar. Tal processo é inicialmente descrito por Dretske em termos da distinção entre representação analógica e representação digital da informação.

A distinção entre analógico e digital é geralmente descrita na literatura em termos de representações contínuas e discretas, respectivamente. Assim, por exemplo, os termômetros com coluna de mercúrio ilustram mecanismos de representação analógica, uma vez que eles descrevem as mudanças de temperatura ambiental em termos de mudanças contínuas na coluna de mercúrio. Em

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

contraste, o termômetro digital representa a temperatura em termos de seqüências discretas de números em uma certa escala.

Adotando a distinção entre informação analógica e digitalmente representada, Dretske sugere a seguinte definição de informação digital:

> "um sinal (estrutura, evento, estado) carrega informação que **s** é **F** na forma digital se e somente se o sinal não carrega informação adicional sobre **s,** que não esteja já contido no fato de **s** ser **F**. Se o sinal carrega informação adicional sobre **s,** a qual não está imbricada no fato de **s** ser **F,** então direi que o sinal carrega essa informação na forma analógica (Dretske, 1981, p. 137).

Quando essa condição de digitalização se realizar, Dretske entende que um tipo de *conteúdo natural* será transmitido através de sinais ou eventos, como por exemplo no caso em que a fumaça carrega consigo o sinal indicativo de fogo. Essa representação digital pode ser obtida através da filtragem da informação analógica quando parte da informação contida no sinal for descartada, por exemplo, ao se ignorar os detalhes sobre a origem da fumaça – se ela provém de queima de cana, papel, madeira, etc. Assim caracterizada, a conversão analógico/digital estaria na base do processo de classificação/categorização de estímulos que recebemos do mundo.

O seu famoso exemplo do velocímetro ilustra a hipótese acima: imagina-se um velocímetro de um automóvel, com duas partes. Na primeira, temos um ponteiro móvel que registra, analogicamente, a velocidade do automóvel, que varia, digamos, entre 0 e 120 km por hora. Na segunda parte, temos um conversor que transforma velocidades em sinais sonoros, correspondentes aos períodos de 0 a 20 km, 20 a 40 km, 40 a 60 km e 60 a 120 km, os quais indicam o momento da mudança de marcha. A hipótese de Dretske é que essa filtragem, que envolve a digitalização da informação expressa através da codificação de intervalos de velocidade em sons específicos com diferentes tonalidades, permite ao motorista inexperiente identificar, e posteriormente classificar, o momento adequado para a troca de marchas.

No exemplo acima, um engenheiro hipotético teria estabelecido um sistema que inclui um processo de transdução de velocidades em sinais sonoros, o qual seria ativado sempre que estímulos provenientes do movimento das rodas do automóvel chegassem ao velocímetro. Esse mesmo

unesp

processo de digitalização ocorre, segundo Dretske, entre a retina e o córtex, possibilitando a identificação e a caracterização da informação visual analógica disponível no meio ambiente.

No caso dos seres humanos, as células da retina responsáveis pela visão de cores respondem mais fortemente à luz verde azulada, **causando** o disparo automático das mesmas de acordo com uma taxa de variação específica. Contudo, tais células, argumenta Dretske, apesar de responderem causalmente ao estímulo, não carregam informação sobre a cor azul esverdeada. Isso porque elas não distinguem as baixas intensidades do azul esverdeado das altas intensidades do amarelo; tanto uma quanto a outra causam o disparo dessas células com a mesma taxa de variação. Ainda que desempenhem um papel causal na percepção de cores, elas, não carregam informação sobre a cor azul esverdeada devido, possivelmente, à presença de ruídos e ambigüidade; a relação de dependência entre tais células e os sinais que carregam não satisfaz, assim, a condição de probabilidade condicional =1, mencionada acima.

Em resumo, a digitalização de um sinal envolve um conjunto de filtros operando na elaboração de uma representação analógica do mundo, a qual garantiria a especificidade da informação percebida por meio do seu caráter natural. Essa operação estaria presente também, segundo Dretske, nas ações cotidianas. Assim, por exemplo, o som da campainha de uma residência (acionada em condições normais) carrega a informação de que alguém está pressionando a campainha. A percepção desse evento (aparentemente simples) resulta de um processo de filtragem da informação analógica que inclui, entre outros, o dado de que que alguém está diante da porta em questão, que um circuito elétrico foi ativado, que a estrutura fisiológica do ouvido do receptor foi estimulada, etc. Essa rica variedade informacional sofreria um recorte significativo por parte de possíveis ouvintes do sinal (cachorros, gatos ou pessoas), resultando na informação "há alguém lá fora".

Ainda que a definição nuclear de informação proposta por Dretske admita a existência de informação independente de um receptor consciente, a articulação dessa informação por um sujeito, e a atribuição de um significado apropriado à mesma, envolve uma postura intencional relacionada à ação. É o sistema, em sua totalidade, (que inclui uma estrutura específica desenvolvida de acordo com sua história evolutiva, situada em um ambiente que lhe propicia a ação) que aprende e comunica o conteúdo informacional daquilo que é aprendido. No discurso comunicativo, Dretske supõe que, "... afirmações que descrevem a informação carregada por um sinal são descrições *intencionais,* e os fenômenos que elas descrevem podem ser qualificados como fenômenos intencionais" (Dretske, 1981, p. 76). Conforme será explicitado na seção seguinte, a intencionalidade decorre, em parte, da



possibilidade de equívoco na apreensão e articulação do conteúdo natural presente na informação digitalizada.

Essa hipótese dretskeana, segundo a qual a especificidade da informação carrega consigo um conteúdo natural, cuja articulação apropriada de significado requer uma postura intencional, possui vários desdobramentos. O primeiro diz respeito ao papel do cérebro em tal articulação de significado: Dretske argumenta que o cérebro desempenha apenas um papel de veículo de recepção, registro e transmissão de informação; mas, por si só, a atividade neural é destituída de conteúdo significativo. Em outras palavras, as estruturas cerebrais não se identificam às estruturas informacionais articuladas pelo organismo, elas são apenas portadoras da informação selecionada pela estrutura funcional do organismo que atua no meio ambiente.

A distinção estabelecida por Dretske entre "história-veículo" e a "história-conteúdo" nos auxilia a compreender sua perspectiva teórica: num livro de História, o desenrolar dos eventos é contado sob a forma de sentenças (*tokens,* que constituem a "história veículo"). Contudo, os acontecimentos narrados não estão no livro; os acontecimentos, selecionados pelas estruturas cognitivas e descritos através das sentenças, constituem, enquanto tais, a "história conteúdo". Analogamente, Dretske argumenta que o que está no cérebro é uma intensa atividade eletroquímica que carrega informação, mas não o seu conteúdo. Nesse sentido, a mente não está na cabeça assim como a história não está no livro (Dretske [1995, p.38]). Permanece, então, a questão de esclarecer a natureza do conteúdo dos estados mentais (retomaremos esse ponto na seção seguinte).

Para o presente propósito de esclarecer em que sentido a proposta informacional de Dretske poderia indicar uma saída para a circularidade presente nas tentativas de solução do problema do *Teeteto*, parece bastar a suposição de que a informação que sustenta uma determinada crença, independe (enquanto informação) do conhecimento do sujeito que a apreende. Essa proposta de Dretske parece, à primeira vista, quebrar a circularidade identificada por Platão nas sua tentativa de buscar os fundamentos para o conhecimento em termos proposicionais. Conforme argumentaremos na seção seguinte, a proposta informacional de Dretske parece, algumas vezes, substituir a alma platônica, que seleciona os dados obtidos das sensações, pelas estruturas cognitivas funcionais do organismo que seleciona a informação analógica e digital recebida pelos canais de informação. Se for este o caso, então estaremos de volta com a problemática platônica e sua circularidade.

unesp

III – Informação, conhecimento e ação: uma saída para o problema do *Teeteto*?

Na seção anterior deixamos sem resposta a questão de como explicar a natureza do conteúdo dos estados mentais, cuja informação correspondente poderia, segundo Dretske, servir de sustentação para as crenças candidatas à fundamentação do conhecimento. A principal dificuldade dessa tarefa reside no fato de que Dretske atribui uma existência objetiva à informação e, ao mesmo tempo, torna a articulação significativa do seu conteúdo dependente do sistema cognitivo que a seleciona. Não estaríamos aqui diante de uma versão informatizada do platonismo, que atribui realidade objetiva à forma, embora conceba a sua existência dissociada do fluxo da natureza? Como vimos, na alegoria mencionada na seção anterior, Platão supõe que existiria na alma um tipo de cera cuja pureza, volume, umidade e densidade variariam de indivíduo para indivíduo. As formas estariam impressas, "como anéis", nessa cera e seriam mais facilmente lembradas naqueles indivíduos cuja profundeza e umidade da cera, moldada pela reflexão, permitisse pensar com acerto. O oposto ocorreria àqueles que pensam erroneamente. A seguinte passagem do *Teeteto* ilustra essa concepção:

> " Sócrates – ... Quando a cera que se tem na alma é profunda e abundante, branda e suficientemente amassada, tudo o que transmite pelo canal das sensações vai gravar-se no coração da alma, como diz Homero, aludindo à sua semelhança com a cera, saindo puras as impressões aí deixadas, bastante profundas e duradouras; os indivíduos com semelhante disposição aprendem facilmente e de tudo se recordam e sempre formam pensamentos verdadeiros, sem virem jamais a confundir as marcas de suas sensações, sendo nítidas e bem espaçadas todas as impressões, com facilidade põem em relação cada imagem com a correspondente marca, as coisas reais, como lhes chamam. São esses os denominados sábios. Não te parece que está certo?" (*Teeteto*, XXXIV, p.57, 194 c-d)

Dado que as condições da alma do receptor de informação constituem, na alegoria platônica, um elemento fundamental para o discernimento e apreensão de verdades, qual seria o equivalente informacional dretskeano (se é que existe alguma equivalência) de tais condições? Não haveria algum estranho parentesco entre a proposta dretskeana de uma ontologia "objetivista" da informação e a ontologia platônica das formas? Analisemos com cuidado as possíveis semelhanças. Em primeiro lugar, é importante ressaltar que a teoria da informação proposta por Dretske se situa no

unesp

interior do paradigma funcionalista representacional da mente, de acordo com o qual: "... um sistema S, representa uma propriedade, F, se e somente se S tem a *função* de indicar (fornecer informação sobre) a propriedade F de um certo domínio de objetos ..." (Dretske, 1995, p.2)

De acordo com essa concepção funcionalista, S desempenha a sua função informacional ocupando diferentes estados s1, s2, ... sn, que correspondem a valores f1, f2, ... fn de F. No exemplo do velocímetro apresentado na seção anterior, S corresponderia ao velocímetro e F à velocidade do carro. A função do velocímetro, no que diz respeito ao processamento de informação, é indicar a velocidade do carro para o motorista através dos estados - por exemplo, s1 = 30 km/h, s2 = 60 km/h, etc). Tal função existe apenas no contexto sistêmico em que se analisa o velocímetro: é graças a sua estrutura, enquanto velocímetro, que se coloca a função de indicar a velocidade, desta ou daquela maneira.

Essa função, estabelecida convencionalmente ao velocímetro, é contrastada com a *função natural* desempenhada pelos órgãos sensoriais, função essa adquirida através da história evolucionária dos organismos em sua interação com o ambiente. Ambas têm a função de indicar, em decorrência de sua estrutura sistêmica , informações sobre estados do mundo. Contudo, Dretske julga que no caso das funções naturais: " As representações que elas produzem, ao executar suas *funções informacionais,* possuem um conteúdo, algo que elas transmitem ou significam, que não dependem da existência de nossos propósitos ou intenções". (Dretske, 1995, p.8)

Surge assim, evolucionariamente, o conteúdo ou o significado natural das representações ligadas à experiência sensorial, o qual, segundo Dretske, é compartilhado por um grande número de organismos, independente dos nossos estados conscientes. Em relação a ele, temos pouco, ou quase nenhum controle, pois tal significado decorre das propriedades sistêmicas dos organismos; os conteúdos naturais são, de alguma forma, fixados pelas funções biológicas dos sistemas sensoriais a que pertencem:

> "Uma vez que herdamos nossos sistemas sensoriais, uma vez que eles são ..."*hardwired",* não podemos ...mudar o caráter sistêmico representacional de nossa experiência sensorial" (Dretske, 1995, p.15).

O mesmo não acontece, entretanto, com o significado ou o conteúdo das crenças, e do pensamento em geral, pertencentes aos organismos possuidores de estruturas evolucionárias complexas, que são passíveis de alteração através de aprendizagem: " através da aprendizagem, posso

unesp

mudar aquilo que acredito quando vejo *K* , mas não posso mudar o tipo de experiência visual que *K* produz em mim" (Dretske, 1995, p.15). As crenças (e estados mentais em geral) operam apenas inicialmente com o conteúdo natural que chega aos organismos, através das sensações, sustentando-as e distinguindo-as das alucinações.

Um resultado dessa distinção entre experiência sensorial (que opera com representações naturais) e pensamento ou crença (que envolve representações aprendidas) servirá de ponto de partida para a distinção dretskeana entre comportamento instintivo e comportamento inteligente.

No caso do comportamento inteligente, a função da experiência sensorial seria fornecer informação (com um conteúdo natural) para o sistema cognitivo que, dependendo de sua flexibilidade, faria os ajustes necessários para torná-la apropriada ao controle do comportamento. Assim, por exemplo, o aprendizado de uma língua, que requer certo grau de inteligência, envolve a experiência do som (do tato ou da visão), que é transmitida inicialmente em forma analógica pelos órgãos sensoriais. O seu processo de aprendizagem se completa com a classificação e reconhecimento de palavras e sentenças, representadas digitalmente após a eliminação de ruído e redundância. Esse processo de classificação desencadeia outros processos mais complexos, ligados ao controle e ajuste da ação.

Uma vez aprendida uma língua, o seu usuário passará a ignorar grande parte da informação sonora que se apresenta na fala analogicamente; " ainda ouvimos, naturalmente, sons mas, após o aprendizado da língua, ... uma dimensão representacional é adicionada ao processo" (Dretske, 1995, p. 20). Essa nova dimensão informacional, de segunda ordem, será empregada para controlar e ajustar o comportamento no meio ambiente: através da fala disponibilizamos informação por meio da qual criamos normas e regras que afetarão, num processo de causalidade circular, nossas ações num plano coletivo e individual.

É justamente essa nova dimensão da informação - obtida através de mecanismos de aprendizagem- que caracteriza, para Dretske, o comportamento inteligente, distinguindo-o do comportamento instintivo. Ela envolve uma postura intencional por parte do sujeito possuidor de crenças, desejos, e de um sistema cognitivo. Além disso, uma das propriedades fundamentais do comportamento intencional, característico das crenças, é sua propensão ao erro: " crenças ... têm o poder de "dizer" ou "significar" que K é F quando K não é F. Na verdade, elas têm esse poder ... mesmo quando K não existe" (Dretske, 1995, p. 28).

Longe de ser visto como uma deficiência do sistema que processa informação, o erro é entendido como um dos elementos chave para se distinguir o comportamento instintivo do

unesp

comportamento inteligente. Isso porque parece não haver lugar para o erro e sua possibilidade de correção no comportamento instintivo; apenas o comportamento inteligente, argumenta Dretske, pode estar equivocado, possuindo recursos de autocorreção. Em sua opinião, o comportamento instintivo não envolve mecanismos de aprendizagem mas, sim, mecanismos de tropismo:

> "Tropismos são processos simples, mecânicos ou químicos, de *feedback*, ou combinações desses processos, que têm a interessante propriedade de se parecer com o comportamento organizado por motivações" (Dretske, 1992, p. 93).

O exemplo de organismos aparentemente muito simples, que se movem em direção à luz, é fornecido para ilustrar esse tipo de comportamento: em certos contextos, o movimento desses organismos pode assemelhar-se àquele dos organismos mais complexos que guiam o seu comportamento em função da luz, de modo a maximizar a eficiência da busca de alimentos, em especial quando eles se movem para o topo das árvores em busca de alimentos nos brotos dos ramos. Tal comportamento é aparentemente inteligente, uma vez que ele parece estar maximizando as possibilidades de sobrevivência dos organismos, mas, segundo Dretske, trata-se apenas de aparências, pois ele não é suscetível de mudanças: quando submetidos a circunstâncias em que a luz os atrai para lugares em que não existe comida, tais criaturas, segundo o relato de alguns observadores, morrem de fome, sem buscar alternativas para corrigir o erro no seu percurso em busca de comida.

Em síntese, o comportamento desses organismos simples envolveria somente o processamento de informação analógica, que é disponibilizado aos seus órgãos dos sentidos, sem os processos subseqüentes de classificação e articulação dessa informação, que a tornaria passível de erros e acertos Em contraste, os organismos mais complexos são capazes de distinguir as circunstâncias em que seu comportamento foi bem sucedido daquelas em que houve alguma falha. Embora ambos acessem, igualmente, a informação analógica que lhes chega através dos órgãos dos sentidos, somente estes últimos registrariam e classificariam as experiências sensoriais; aprendendo através de erros e acertos eles desenvolveriam informação, de segunda ordem, que lhes permitiria alguma forma de controle sobre o comportamento.

Com essa síntese da proposta informacional de Dretske estamos em condições de responder as duas questões formuladas anteriormente: a primeira consiste em verificar se a teoria informacional proposta por Dretske não poderia ser vista como uma versão informatizada do platonismo, que atribui

unesp

realidade objetiva à forma, embora conceba a sua existência dissociada do fluxo da natureza. A segunda questão, de certo modo relacionada com a primeira, consiste em verificar se essa proposta informacional oferece uma solução para o problema do *Teeteto*.

No que diz respeito à primeira questão, julgamos que o funcionalismo informacional - em especial na sua versão adotada em grande parte da Inteligência Artificial, Filosofia da Mente e Ciência Cognitiva tradicional - pode ser visto como um herdeiro do legado platônico, entendido como uma "purificação" da matéria. A proposta de Dretske, de classificação do comportamento em inteligente e instintivo em função da filtragem dos dados sensoriais, parece partilhar desse legado que, associado aos recursos tecnológicos de nosso século e às descobertas da teoria da informação na biologia e na engenharia genética, vem alterando radicalmente os rumos de nossa sociedade.

Contudo, uma hipótese central de Dretske é que o controle do comportamento só é possível graças à existência de mecanismos internos, indicadores de sucesso ou de fracasso, desenvolvidos através de aprendizagem. Uma pergunta que se coloca imediatamente é: de onde vem esses mecanismos internos que funcionam como indicadores no processo de aprendizagem? A resposta de Dretske é que eles podem ser construídos (como nos casos das redes neurais artificiais que possuem mecanismos de aprendizagem envolvendo a eliminação de erros), aprendidos (principalmente através de processos associativos que se estabelecem entre certas condições ambientais e o comportamento bem sucedido) ou, simplesmente, existirem em decorrência de processos evolutivos.

Se por um lado essa postura aparentemente behaviorista de Dretske o afasta de universo platônico, por outro, ela nos deixa, reconhecidamente, sem uma resposta sobre a origem dos mecanismos de aprendizagem que operam com tais indicadores de relações:

"Através do reforço de certos *output* – que são recompensados quando e apenas quando eles ocorrem em certas condições – indicadores internos dessas condições são recrutados como causas desses *output*.. Como é que eles são recrutados por esse processo pode ser (e para mim é) um mistério completo" (Dretske, 2002, p.98).

Se deixarmos de lado essa dificuldade, referente à origem das crenças e dos indicadores informacionais, ainda nos resta saber em que medida a proposta informacional dretskeana oferece uma saída para o problema da circularidade levantado por Platão na sua tentativa de caracterizar a distinção entre conhecimento e crença verdadeira.

Entendemos que a sugestão proposta por Dretske quanto ao problema do *Teeteto* é a seguinte: Se um organismo tiver acesso a uma fonte de informação e, além da experiência sensorial, ele tiver o

unesp

poder de discernir as condições favoráveis das desfavoráveis ao seu comportamento, então indicadores informacionais se estabelecerão, os quais lhe servirão como instrumento de aprendizagem e controle do comportamento. Tais indicadores constituem informações que serão utilizadas como *razões* explicativas do sucesso ou falha do comportamento inteligente; eles fornecem subsídios para crenças verdadeiras a respeito do mundo. De posse de tais indicadores, um organismo teria acesso ao conhecimento sempre que:

(a) possuir crenças verdadeiras, sustentadas por informação e,

(b) tais crenças devem mostrar-se relevantes para a otimização do seu comportamento no ajuste com o meio ambiente.

A partir de (a) e (b), Dretske caracteriza o **conhecimento** (perceptual) como um estado do sistema que possui um conteúdo informacional correspondente à informação que o produziu.

Como o erro é naturalmente possível, ele teria que eliminar as crenças falsas, que não estariam fundadas em informação e, também, dispor de critérios de relevância que lhe auxiliassem na seleção da informação para que, de fato, ele pudesse realizar a tarefa de harmonizar o seu comportamento com o meio ambiente, maximizando a sua existência nesse meio. Mas, neste caso, quais seriam os critérios de relevância adotados para realizar essa tarefa? Dretske nos remete ao contexto, ao meio ambiente e às condições estruturais do organismo para a solução dessa dificuldade, o que o livra, de certa forma, do legado platônico, mas nos deixa com a semente do problema do *Teeteto* que, inevitavelmente, se perpetuará enquanto continuarmos a buscar verdades.

A nossa sugestão é que abandonemos a busca de verdades, adotando uma postura perspectivista que, sem cair no relativismo, possa estar comprometida com a preservação do meio ambiente e da vida na sua dinâmica criadora (os tópicos da criatividade e do perspectivismo serão objetos de nossa investigação nos capítulos 2 e 3, respectivamente). Esse perspectivismo se diferenciaria do relativismo, em especial no domínio da investigação cientifica e filosófica, à medida que, ao compartilharmos nossos pensamentos em grupos de estudiosos de um certo tema, assumimos como evidentes certos termos primitivos expressos na linguagem natural ou na linguagem técnica em que determinado problema se expressa. Nesse contexto, certas dúvidas céticas não se colocariam, sob pena de violação de regras da linguagem aceitas pelos integrantes dessa comunidade. Os problemas considerados relevantes nestes grupos seriam aqueles que abalariam, de alguma forma, os

unesp

pressupostos básicos sob os quais estão assentados os hábitos constitutivos do pensamento então vigente.

Conforme argumentaremos no capítulo 2, a presença de conflitos, anomalias e rupturas nos hábitos fundamentais, que se mostraram eficientes na manutenção de nossas crenças, oferece a oportunidade para os seres criativos de reavaliar a sua postura em relação ao fluxo da vida. Neste ponto concordamos com a hipótese dretskeana, segundo a qual uma das principais características do comportamento inteligente é a sua capacidade de aprender através de mecanismos de correção de erros. Como veremos no capítulo seguinte, a capacidade de detectar e corrigir erros está na base do pensamento criativo.

Bibliografia

BATESON, Gregory. *Mind and Nature*: A necessary unit. 1.ed. New York: Cambridge Press, 1979.

BATESON, Gregory. *Steps to an Ecology of mind.* New York: Cambridge Press, 1999.

DRETSKE, Fred I. *Explaining Behavior*. Oxford: Basil Blackwell Publisher, 1981.

_____. *Knowledge and Flow of the Information.* 1. ed. Oxford: Basil Blackwell Publisher, 1981.

_____. *Naturalizing the mind.* Cambridge: MIT Press, 1995.

_____. Precis of knowledge and flow of the Information and replies to comments; *Behavioral and Brain Sciences*; 1983.

_____. *Explaining behavior:* Reasons a world of causes. Cambridge: MIT Press, 1988.

_____. *Naturalizing the mind.* Cambridge: MIT Press, 1995.

FODOR, J. A. *Psychosemantics.* Cambridge: MIT Press, 1987.

GONZALEZ, M.E.Q. (1989) *A Cognitive approach to visual perception*; Tese de doutorado, University of Essex, pp.182.

PEREIRA Jr., A.& GONZALEZ, M.E.Q. Informação, organização e linguagem. In: *Espaço e Tempo,* Fátima R.R. Évora (ed.). Coleção CLE – Unicamp. Vol.15, 1995.

PLATÃO. *Teeteto.* Trad. Carlos Alberto Nunes. Digitalização Grupo de Discussão Acrópolis. http:/br.egroups.com/group/acropolis/

PLATON. Teeteto. In: *Obras completas.* Madrid: Aguilar, 1966.

SHANNON, Claude E.; WEAVER, Warren. *The Mathematical Theory of Communication.* rev. e aum. Urbana and Chicago: University of Illinois Press, 1998.

STONIER, Tom. *Information and the internal structure of the universe*: an exploration into information physics. 1.ed. London: Spring-Verlag, 1990.

unesp

Complexidade e criatividade: uma abordagem sistêmica dos processos auto-organizados

*Maria Eunice Quilici Gonzalez**

Investigaremos aqui a noção de criatividade a partir da perspectiva da Teoria Geral dos Sistemas, com ênfase nos processos auto-organizados presentes na ação criativa. Para tanto faremos, na seção I, uma breve introdução aos pressupostos centrais da sistêmica, tal como originalmente formulados por Bertalanffy (1969), situando-os no cenário da teoria da auto-orgnização (Debrun 1996, Haken 1999, Gonzalez e Haselager 2002) e dos sistemas complexos (Jensen 1998). Nesse contexto, e inspirados nas idéias de Peirce (1931-1958) sobre o raciocínio abdutivo, apresentamos, na seção II, uma caracterização da criatividade entendida como um processo auto-organizado que possibilita a expansão de hábitos em sistemas complexos.

Palavras-chave: auto-organização, criatividade, complexidade, emergência, raciocínio abdutivo, parâmetro de ordem, sistema.

Pressupostos da Teoria Geral dos Sistemas

Uma das suposições fundamentais da Teoria Geral dos Sistemas (TGS), tal como formulada por Bertalanffy (1969), Weinberg (1975), Boulding (1964), entre outros, é que a ordem do mundo empírico, que experienciamos em nossas ações no plano macroscópico dos

---

* Da UNESP, gonzalez@marilia.unesp.br

cm 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 unesp

fenômenos, possui uma ordem ainda mais geral, de segundo grau, que unifica a imensa quantidade de elementos que compõem esse mundo. Uma tal ordem unificadora não se resume à somatória das partes de seus constituintes, mas apresenta uma característica global, inovadora, que depende da dinâmica de interação entre as partes individuais, existentes no plano microscópico, de sistemas complexos. Conforme ressalta Jensen (1998, p. 1), um sistema é complexo no sentido de que a sua evolução inclui elementos existentes em várias escalas temporal e espacial. Paradoxalmente, embora a dinâmica de tais sistemas seja complexa, suas propriedades estatísticas podem ser descritas por leis simples que se aplicam a sistemas aparentemente diferentes quando investigados a partir de uma perspectiva microscópica.

Um dos sonhos dos teóricos da sistêmica, como por exemplo Weinberg (1975) reconhece, é encontrar uma lei geral unificadora, de terceira ordem, que governasse o processo de geração de leis (de segunda ordem) no plano físico, químico, biológico, psicológico e até mesmo sociocultural. Apesar de aparentemente ingênua, essa pretensão não surge da ignorância a respeito da imensa complexidade que permeia a variedade do mundo físico. Ao contrário, a complexidade é o ponto de partida das análises sistêmicas da ordem subjacente ao mundo empírico de agentes situados em um ambiente dinâmico, em constante transformação. Foi principalmente na biologia, com a investigação da intrincada dinâmica da vida, que a Teoria Geral dos Sistemas encontrou um lugar natural. Ela surgiu na tentativa de solucionar os impasses criados pela abordagem vitalista e sua rival mecanicista do comportamento dos organismos, propondo uma alternativa holística que acarreta uma reviravolta nos métodos de análise (e síntese) vigentes na pesquisa científica contemporânea. Tal reviravolta, como procuraremos mostrar, abre caminhos para uma visão integradora do organismo no seu meio ambiente.

No contexto dos estudos da criatividade, no qual o presente trabalho se situa, a abordagem sistêmica focaliza os princípios ou leis gerais que supostamente estruturam o raciocínio ou a ação criadora. Entre eles, existem dois princípios básicos que guiam a busca de leis gerais unificadoras na investigação sistêmica. O primeiro, conhecido como *princípio de mutualidade sujeito-ambiente* (Gibson 1979), pressupõe a existência de um plano relacional que molda a identidade dos agentes e do ambiente que os cerca. Nessa perspectiva, a ordem expe-

unesp

rienciada é sempre "a ordem para um agente" que molda e é, simultaneamente, moldado pelo seu ecossistema. Diferentes ecologias são geradas pelos distintos planos relacionais que se delineiam nas junções dos inúmeros nichos encapsulados no complexo sistema da vida. Cada membro de uma espécie, na sua ação cotidiana, reúne diferentes planos da existência, os quais integradamente dão lugar às distintas "visões de mundo".

O segundo princípio, conhecido como *princípio de emergência*, admite a existência de propriedades globais emergentes em sistemas complexos, cujas características específicas só podem ser encontradas no produto da interação global, espontânea, que se estabelece entre os elementos desses sistemas no plano microscópico. Conforme explicitaremos adiante, a *emergência* espontânea é uma propriedade dos sistemas que se auto-organizam, e sua presença nesses sistemas constitui um fator estruturante fundamental das diversas ordens experienciadas no mundo empírico. As propriedades emergentes possuem uma taxonomia própria e são, em geral, classificadas em várias categorias, tais como emergência forte, fraca, de propriedades etc. Contudo, para os nossos propósitos, é suficiente a idéia básica segundo a qual tais propriedades, características de sistemas complexos auto-organizados, não podem ser encontradas nas partes isoladas , mas apenas na totalidade interativa desses sistemas.

A auto-organização, por sua vez, é um processo pelo qual novas formas de organização são espontaneamente geradas sem a interferência direta, ou determinante, de elementos externos ao próprio processo. Em outras palavras, uma organização é "auto" quando produz a si própria em decorrência da interação entre elementos distintos e independentes que se reúnem sem qualquer finalidade preestabelecida ou, idealmente, por obra do acaso (Debrun 1996).

Nos processos de auto-organização simples, também conhecidos como *auto-organização primária*, o encontro entre elementos independentes pode desencadear uma dinâmica interativa que possibilita a formação de padrões simples e bastante instáveis. Dada a fragilidade desses padrões, eles tendem a se desfazer rapidamente com a interferência de elementos externos ao próprio processo. Exemplos desses processos podem ser encontrados nos padrões de nuvens que se formam na atmosfera, nas dunas de areia ou mesmo nos encontros entre indivíduos que se reúnem nas grandes multidões e se dispersam com

unesp

o passar do tempo. Entre os inúmeros padrões que se criam por meio dos processos de auto-organização primária, apenas alguns adquirem estabilidade, em geral graças a mecanismos de ajustes e à variedade dos planos em que eles se situam, dando lugar aos processos mais complexos de auto-organização.

Nos processos de auto-organização mais complexos, também conhecidos como *auto-organização secundária*, a emergência de padrões é acompanhada de mecanismos de aprendizagem que possibilitam o aperfeiçoamento e a estabilização de tendências que são apenas esboçadas nos processos auto-organizados simples. A estabilidade dos padrões gerados nos processos de auto-organização secundária decorre, principalmente, dos mecanismos de ajuste, subjacentes aos mecanismos de aprendizagem, e dos *parâmetros de ordem* existentes em escalas temporais e espaciais distintas.

Conforme ressalta Haken (1999), parâmetros de ordem são formas estruturadoras emergentes da dinâmica de interação entre elementos no plano microscópico, as quais produzem novas características em escalas macroscópicas. Uma propriedade importante dos parâmetros de ordem é que, uma vez formados, eles subjugam e restringem o movimento dos elementos individuais (no plano microscópico) que lhes deram origem, tornando o comportamento dos mesmos, em certo sentido, previsíveis.

Entre os inúmeros exemplos de auto-organização secundária, podemos mencionar o caso daquelas instituições que, uma vez estabelecidas sem qualquer planejamento, e tendo incorporado o resultado de processos de aprendizagem que lhes atribuem uma certa estabilidade, restringem, e algumas vezes determinam, de maneira circular, o comportamento dos indivíduos que lhes deram origem.

Um outro exemplo, mais polêmico, é dado por Debrun (1996), que ilustra o processo de auto-organização secundária por meio do estudo da dinâmica das partidas de futebol. Ainda que o jogo seja regrado, e determinado por fatores externos que se impõem à sua organização enquanto jogo de futebol (e neste sentido ele não é auto-organizado), a dinâmica das partidas (quando auto-organizadas) parece escapar do controle externo, dependendo principalmente da interação entre os membros do time. Nesse sentido, Debrun considera que a dinâmica presente em cada partida de futebol, quando desen-

unesp

volvida honestamente, ilustra por excelência os movimentos do processo de auto-organização secundária.

Tal dinâmica se inicia com o encontro de times de futebol, encontro este que pressupõe a existência de formas estáveis de organização resultantes de um longo processo de aprendizagem em escalas temporais e espaciais distintas. Nessas condições, o resultado da partida, quando auto-organizada, decorre da emergência de parâmetros de ordem, que se estabelecem por meio da dinâmica espontânea das interações entre os membros dos times. Segundo Debrun, a competição e a cooperação, bem como os diversos mecanismos de ajuste que se instalam numa partida, constituem os ingredientes da "copa e cozinha" da auto-organização secundária; eles surgem sem a influência determinante de fatores externos à própria partida.

Um dado curioso, característico dos processos de auto-organização secundária, é que, uma vez estabelecidos os parâmetros de ordem que estruturam a dinâmica de uma partida, eles restringem o comportamento dos jogadores, deixando pouco lugar para a inovação. Neste estágio, inicia-se um mecanismo de cristalização que indica o fechamento do processo e, conseqüentemente, o seu fim enquanto um processo auto-organizado. Em tais circunstâncias, será somente em condições especiais, nas quais instabilidades ou crises abalam os parâmetros de ordem estabelecidos, que o comportamento dos jogadores apresentará características inesperadas ou inovadoras. Conforme argumentaremos na seção II, é principalmente nos momentos de instabilidade e imprevisibilidade que a criatividade pode ganhar espaço na ação dos jogadores, ampliando o grau de complexidade das partidas e abrindo caminhos para a emergência de novidades.

Os exemplos supra ilustram os princípios e processos fundamentais sobre os quais se apóia a visão sistêmica em busca de leis gerais unificadoras do plano macroscópico da experiência: *mutualidade sujeito-ambiente*, emergência, auto-organização. A partir dessa perspectiva, e inspirados nas idéias de Peirce sobre o raciocínio abdutivo, esboçaremos a seguir uma possível aplicação da Teoria Geral dos Sistemas no estudo da criatividade.

unesp

Uma visão sistêmica da criatividade

Situações em que nossas expectativas cotidianas são frustradas nos levam a refletir sobre a presença de estruturas informacionais encapsuladas em nossas ações. Essas estruturas informacionais são, por sua vez, moldadas por uma rede complexa de disposições e hábitos que organizam nosso universo físico, biológico e sociocultural. O mesmo parece ocorrer no contexto das ações criativas, seja no domínio da arte ou da ciência nos quais os hábitos se expressam por meio de processos implícitos de inferências.

Assim, por exemplo, em um estudo sobre a credibilidade de nossas fontes de conhecimento, Soza (2003) questiona como nos sentiríamos se, após observar um espetáculo de dança, viéssemos a descobrir que os movimentos harmoniosos da bailarina nada mais eram do que a sua tentativa de evitar uma queda abrupta produzida por um vento forte, ou por outro acidente qualquer que teria provocado o desequilíbrio de seu corpo. A opinião geral é que em tais circunstâncias sentiríamos uma frustração de nossas expectativas, acompanhada da sensação de desapontamento. O mesmo ocorreria se viéssemos a descobrir que os traços deixados por Picasso no relevo de uma parede, que julgávamos de grande valor artístico, resultaram unicamente do produto de uma rápida ação realizada pelo pintor para tapar um vazamento de água naquele local durante um dia chuvoso que teria produzido uma inundação naquele local.

Quando nos colocamos no papel de espectadores do trabalho criativo, seja na arte ou na ciência, é comum julgarmos que eles resultam de obra do acaso ou, mais provavelmente, do *insight* de gênios ou criaturas exóticas, sem qualquer estrutura lógica subjacente. É nesse sentido, por exemplo, que Popper (1959) argumenta que não existe uma lógica no processo de descoberta científica, ou qualquer método para a geração ou reconstrução de novas idéias, uma vez que, em sua opinião, toda descoberta criadora contém um elemento de irracionalidade e acaso.

Sem negar o papel do acaso, e apoiados nas idéias de Peirce (1931-1958), defenderemos a hipótese segundo a qual o processo criativo envolve, em geral, um movimento de ruptura e ajuste no conjunto de crenças e hábitos bem estabelecidos. Faremos aqui uma breve

discussão dessa hipótese, que foi desenvolvida em detalhes em Gonzalez & Haselager (2002), com o objetivo de indicar um novo cenário metodológico a partir do qual entendemos que a perspectiva sistêmica torna inteligível a dinâmica geradora do pensamento criativo.

Muito antes de se oficializar a Teoria Geral dos Sistemas e a teoria da auto-organização, Peirce (1931-1958) desenvolveu uma belíssima teoria do cosmos a partir do seu sistema filosófico que ficou conhecido como pragmatismo. Nesse contexto, Peirce situa o pensamento criativo como um elemento central auto-estruturador do cosmos. Em particular, ele argumenta que a atividade do pensamento criador constitui o motor da dinâmica de autogeração e expansão dos estados mentais, sendo responsável pela produção de novos hábitos.

Os hábitos, entendidos por Peirce como uma lei fundamental do cosmos que regula as tendências à repetição, podem ser estáveis ou instáveis. Quando estáveis, eles constituem uma crença que, diz Peirce, "enquanto dura, é um hábito forte e, como tal, força o homem a acreditar até que alguma surpresa quebre esse hábito" (CP 5.524). Como os hábitos não existem isoladamente, a compreensão da dinâmica de produção de hábitos é fundamental para se entender os processos gerais que organizam as crenças que dão corpo ao fluxo do pensamento (não necessária ou exclusivamente humano).

De maneira bastante resumida, indicaremos a proposta de organização e expansão do pensamento que julgamos corresponder à perspectiva pragmaticista de Peirce, para, em seguida, indicar como ela pode nos auxiliar na tentativa de explicar a dinâmica do processo criativo na perspectiva da ciência cognitiva e da filosofia da mente contemporânea.

Para Peirce, a produção de hábitos, entendidos como uma tendência à repetição, obedece a uma lei cósmica fundamental à qual todas as mentes estão inevitavelmente sujeitas. Uma vez formado, um hábito dá lugar a outros hábitos, criando, numa perspectiva auto-organizada, uma rede. Na medida em que essa rede se estabiliza, ela permite a emergência de crenças (as quais, como ressaltamos, são, para Peirce, nada mais do que hábitos fortalecidos), com base nas quais experienciamos e agimos no ambiente que nos cerca, alterando-o e sendo por ele alterados.

É justamente no pano de fundo dessas crenças que criamos expectativas, fazemos previsões e antecipamos a ocorrência dos even-

tos em nossas ações. É também em relação a esse pano de fundo de crenças que surgem as novidades e surpresas, tornando-nos muitas vezes instáveis e confusos no entendimento dos eventos que nos dizem respeito. Felizmente, a maioria de nós não vive eternamente em um mar de confusões e surpresas. Estas surgem, em geral, como prenúncios de mudanças em nossa forma de enxergar o mundo: dado o caráter dinâmico de nossa percepção e do meio ambiente, ocorre algumas vezes um descompasso entre eles e os nossos hábitos estáveis.

O descompasso entre a dinâmica das expectativas, guiada pelos hábitos estáveis, e a dinâmica do mundo gera um desconforto na mente, que passa a vivenciar um conflito. Tal conflito, ressalta Peirce, produz na mente um efeito de surpresa, que pode ser ativo ou passivo. As surpresas ativas ocorrem "...quando aquilo que é percebido positivamente se choca com as expectativas". Já as surpresas passivas ocorrem quando, na ausência de expectativas positivas, ou de qualquer suspeita de que algo de novo possa ocorrer, "...alguma coisa bastante inesperada acontece, – tal como um eclipse do sol que não se tinha antecipado" (CP 8.315).

Dúvidas e mais dúvidas surgem em decorrência da percepção de novos fatos (ou de alguma aparente anomalia) que se chocam com as expectativas sustentadas por hábitos sedimentados, ou crenças. O choque ou a surpresa, seguida da dúvida, constituirá a alavanca motriz do processo de investigação e criação de novos hábitos.

A hipótese de Peirce, fundamental para o entendimento de sua lógica da descoberta, é que a mente, motivada pelo desconforto da surpresa e da dúvida, iniciará uma busca de novos hábitos ou conjuntos de crenças que permitam a eliminação, em especial, daquelas dúvidas que a tornam instável. Tal busca apenas será interrompida quando essas novas crenças forem estabelecidas e os fatos surpreendentes ou anômalos puderem ser vistos como não problemáticos, triviais ou como "algo que se poderia esperar".

Uma das grandes contribuições de Peirce para a compreensão do processo criativo, que se caracteriza principalmente pela busca de hábitos estáveis para a mente, reside na sua proposta de formalizar as etapas dessa busca, às quais ele deu o nome de raciocínio abdutivo. Este tipo de raciocínio possui a seguinte forma:



Um fato surpreendente, C, é observado;

Mas se H fosse [uma hipótese] verdadeira, C ocorreria como uma questão [natural] de fato;

Portanto, existem razões para suspeitar que H seja verdadeira. (CP 5.189)

Peirce insiste em que o ponto inicial desse raciocínio, que desencadeia a busca por hipóteses ou crenças verdadeiras, é sempre uma anomalia ou um fato surpreendente. A suspeita de que uma hipótese (H) possa explicitar aquilo que inicialmente parece problemático ou anômalo é bastante para que a mente eleja tal hipótese como uma possível candidata à crença estável. Contudo, deve ficar claro de antemão que o raciocínio abdutivo não fornece garantias de sucesso, mas apenas indica uma tendência ou uma disposição da mente no processo de geração de hábitos de conduta que poderiam livrá-la da dúvida e da instabilidade. Essa disposição, ou faculdade, que muitas vezes nos conduz ao erro, é considerada por Peirce como um tipo de instinto:

"Essa Faculdade é ao mesmo tempo da natureza geral do instinto, assemelhando-se aos instintos dos animais na sua capacidade de ultrapassar as forças gerais da razão e dirigir-nos como se estivéssemos de posse de fatos que estão inteiramente além do alcance de nossos sentidos. Ela se assemelha também ao instinto na sua pouca confiabilidade no que diz respeito ao erro: pois embora esteja mais freqüentemente errada do que certa, ainda assim, a relativa freqüência com que está certa é, na sua totalidade, a coisa mais maravilhosa de nossa constituição". (CP 5.173, tradução livre)

Uma vez sugerida uma hipótese por meio do raciocínio abdutivo, ela será submetida aos procedimentos clássicos de verificação indutiva e, mais freqüentemente no caso da atividade científica, de verificação dedutiva; estes procedimentos determinarão a veracidade da hipótese formulada para resolver os problemas em questão. Escapa do escopo deste pequeno trabalho a análise dos detalhes da elaboração desenvolvida por Peirce sobre os procedimentos de verificação científica de hipóteses. Interessa-nos mais ressaltar um dado curioso da passagem acima que diz respeito à natureza do pensamento e, indiretamente, dos mecanismos de inferência que o estruturam.

Embora Peirce compare a faculdade geradora do raciocínio abdutivo aos instintos animais, ele não a restringe apenas ao cérebro

humano, nem mesmo aos organismos. Em uma das inúmeras passagens de sua obra sobre a natureza dos processos de pensamento, ele ressalta que "...o pensamento não está necessariamente conectado com um cérebro. Ele aparece no trabalho das abelhas, dos cristais, e no mundo puramente físico..." (CP 4.5551)

Também não vamos entrar em discussão aqui a respeito do mérito dessa postura ontológica assumida por Peirce no que concerne à natureza do pensamento, mas vamos ressaltar a relevância de suas idéias sobre a estrutura do raciocínio abdutivo para a perspectiva sistêmica contemporânea. Em particular, julgamos que Peirce nos oferece um recurso profícuo para a elaboração de um método de investigação do processo criativo a ser incorporado no paradigma da sistêmica.

Conforme ressaltamos na seção I deste trabalho, um dos objetivos da sistêmica é buscar leis gerais unificadoras, que governem o processo de geração de leis no plano físico, químico, biológico, psicológico e até mesmo sociocultural. Nesse sentido, parece relevante a contribuição de Peirce na formalização do raciocínio abdutivo entendido como uma lei geral responsável pelo processo de expansão de hábitos.

Parece igualmente relevante ressaltar a natureza dinâmica das leis na concepção pragmaticista de Peirce: nas interações com o cosmos, os organismos estão submetidos a leis gerais que podem modificar-se lentamente ao longo do tempo. Elas apenas indicam tendências ou disposições no processo de geração de hábitos; assemelham-se mais aos parâmetros de ordem, tal como caracterizados na seção I – como propriedades emergentes da dinâmica de interação entre elementos no plano microscópico –, do que às leis físicas imutáveis, tal como concebidas pela física clássica.

Além disso, no caso da investigação sobre a natureza da criatividade, se aceitarmos a sua hipótese de que a criação de novos hábitos surge em decorrência de um mecanismo da mente que busca (como tendência) a estabilidade, num contexto de rupturas, então teremos um poderoso algoritmo para investigar os processos de invenção e descoberta tanto na ciência como na arte.

Um ponto de partida de tal investigação seria o conjunto de hábitos estabelecidos nos contextos biológico, psicológico e sociocultural dos indivíduos situados em seus nichos específicos. Nessa perspectiva, a análise do cruzamento das várias fontes de tensão, que mui-

unesp

tas vezes se interpolam por obra do acaso em planos distintos, poderia nos indicar pistas sobre as rupturas que possivelmente teriam dado lugar ao surgimento espontâneo de novas formas de hábitos que, também por força do hábito, assumiram uma dimensão artística relevante para a comunidade ou o indivíduo em questão.

Na visão supra esboçada, os grandes gênios ou inventores exóticos teriam um papel minimizado em suas invenções na medida em que eles seriam dissolvidos no fluxo auto-organizado das séries de hábitos constitutivas dos vários universos em que habitam. Sem dúvida essa abordagem deixa no horizonte vários problemas para serem resolvidos, em especial o problema do livre-arbítrio na ação criadora. Felizmente, podemos adotar temporariamente o raciocínio abdutivo como uma etapa de um exercício criativo e reiniciar, a partir daqui, uma nova cadeia de hábitos que talvez nos auxiliem na tentativa de compreender a dinâmica dos nossos processos mentais criativos.

Referências bibliográficas

Ashby, W. R. (1962). Principles of the self-organizing system. In H. von Foerster. & G. W. Zopf, Jr. (Eds.), *Principles of self-organization*. Oxford: Pergamon.

Bertalanffy, L. (1969). General systems theory. New York: Braziller.

Boden, M. (1990). *The creative mind*. London: Sphere Books.

Boden, M. (1996). What is creativity? In M. Boden (Ed.), *Dimensions of creativity*. London: MIT Press.

Debrun, M. A. (1996). A idéia de auto-organização. In M. Debrun, M. E. Q. Gonzalez & O. Pessoa Jr. (Eds.) *Auto-organização – estudos interdisciplinares*. Coleção CLE. v. 18. Campinas, SP, Brasil: Unicamp.

Soza, E (2003). Palestra realizada durante o *Encontro de filosofia da mente* realizado pela Sociedade Argentina de Filosofia, no período de 1-3 de novembro, em Buenos Aires.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gonzalez, M. E. Q. (2000). The self-organizing process of distributed information: a way out of the mind-body problem? In *Proc. 5th Brazilian – International Conference on Neural Networks*. Rio de Janeiro, Brasil.

Gonzalez, M. E. Q. & Haselager, W. F. G. (2002). Raciocínio abdutivo, criatividade e auto-organização. *Cognitio – Revista de Filosofia*, n. III, EDUC-Angru, São Paulo, Brasil.

Haken, H. (1983). *Synergetics*. Berlin: Springer Verlag.

Haken, H. (1999). Synergetics and some applications to psychology. In W. Tschacher & P. P. Daualder (Eds.), *Dynamics, synergetics, autonomous agents*. London: World Scientific.

Jensen, H. J.(1998). *Self-organized criticality: emergent complex behavior in physical and biological systems*. Cambridge: Cambridge University Press.

Peirce, C. S. (1931-1958). Collected papers (CP), v. 1-8. In C. Hartshorne, P. Weiss and A. Burks (Eds.). *The collected papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Popper. K (1959). *The logic of scientific discovery*. New York: Basic Books.

Weinberg, G. (1975). *An introduction to general system thinking*. New York, John Wiley & Sons.

Algumas das idéias apresentadas neste capítulo foram desenvolvidas em trabalhos realizados em colaboração com os professores Pim Haselager, da Universidade Católica de Nijmegen, Holanda, e Maria Candida Soares Del-Masso, UNESP, a quem agradeço imensamente pela preciosa cooperação. Agradeço especialmente à professora Mariana Claudia Broens e aos colegas do grupo CLE de auto-organização da UNICAMP pela fonte inesgotável de inspiração e à FAPESP e UNESP por financiarem a pesquisa que tornou possível esta publicação.

AUTO-ORGANIZAÇÃO E PERSPECTIVISMO: ALGUM ACRÉSCIMO À CIÊNCIA COGNITIVA?[1]

Maria Eunice Quilici GONZALES[2]

Introdução

A jovem e controversa história da ciência cognitiva nos apresenta duas tendências centrais, aparentemente antagônicas, no estudo da natureza dos estados mentais responsáveis pelo comportamento inteligente. A primeira, predominante na **Inteligência Artificial (IA),** focaliza as *regras abstratas e os algoritmos* no estudo dos elementos estruturadores das representações mentais, que por hipótese constituiriam os pilares do nosso universo cognitivo. A segunda, predominante no **Conexionismo**, ou Redes Neurais Artificiais **(RNA**) enfatiza as *leis naturais e os mecanismos físicos* atuantes no processo de formação, organização e desenvolvimento de tais representações. Apesar de distintas, essas duas tendências possuem em comum pressupostos básicos fundamentais que direcionam os seus programas de pesquisa através da elaboração de modelos computacionais para a explicação da natureza e estrutura dos estados mentais.

Inúmeras críticas têm sido dirigidas ao projeto de explicação da ciência cognitiva, através do emprego de modelos computacionais, elaborados a partir de algoritmos e símbolos destituídos de significado. Assim, por exemplo, Searle, tendo debatido por anos o seu argumento do quarto chinês contra tal projeto, faz o seguinte desabafo em relação à postura objetivista da ciência cognitiva:

> ...A ciência cognitiva prometia um rompimento com a tradição behaviorista na psicologia, porque pretendia penetrar na caixa-preta da mente e examinar o seu

---

[1] Texto da comunicação apresentado no III Encontro Brasileiro/Internacional de Ciência Cognitiva em 16.04.98.

[2] Departamento de Filosofia – Faculdade de Filosofia e Ciências – UNESP – 17525-900 – Marília – SP. gonzales@marilia.unesp.br

> funcionamento interno. Infelizmente, porém, a maioria dos cientistas cognitivos da corrente principal simplesmente repetiu os piores erros dos behavioristas: insistiu em estudar apenas fenômenos objetivamente observáveis, ignorando dessa forma, as características essenciais da mente. Portanto, quando esses cientistas abriram a grande caixa-preta, só encontraram lá dentro uma porção de caixas-pretas. (Searle, 1997, p. 3)

As pequenas caixas-pretas a que Searle se refere podem ser entendidas como os inúmeros modelos computacionais elaborados na ciência cognitiva para simular, e assim explicar, a estrutura e funcionamento do pensamento inteligente, nas suas várias modalidades. Esses modelos foram inicialmente propostos por Turing na década de 50, através do conhecido *teste de Turing*, que pode ser resumido, grosso modo, na seguinte proposta: se for possível programar uma máquina seqüencial de estados discretos (conhecida como máquina de Turing), de tal forma que o seu comportamento, na realização de uma tarefa, seja indistinguível daquele apresentado por um ser humano inteligente, então esta máquina teria passado no teste de Turing. Nesta circunstância, Turing argumenta que tal máquina simularia a atividade inteligente em questão e que o seu programa constituiria um bom modelo explicativo dos estados (mentais?) do sujeito que geraram o seu comportamento inteligente.

A máquina de Turing é uma máquina abstrata, que pode ser instanciada, materializada, em uma máquina física. Contudo, o seu aspecto físico seria secundário e irrelevante para a proposta de Turing, cujo núcleo fundamental reside na hipótese segundo a qual é possível, e esclarecedor, o estudo da inteligência humana através das estruturas simbólico-funcionais, decomponíveis em unidades básicas de informação.

É justamente contra a proposta de Turing que se dirigem as críticas de Searle e de outros críticos da IA. No que se segue, discutiremos, na seção I, alguns dos pressupostos metodológicos subjacentes à abordagem cognitivista do mental aos quais estão dirigidas as referidas críticas. Em seguida, respondendo à pergunta-título do presente trabalho, argumentaremos, na seção II, que a atual metodologia adotada nos programas de pesquisa da ciência cognitiva poderia ser enriquecida através da teoria da auto-organização, em particular na sua aplicação ao estudo da noção de *sujeito*. Finalmente, indicaremos como essas idéias, em conjunto

com o Perspectivismo proposto por Peterson (1996), poderiam ampliar o universo metodológico da atual ciência cognitiva, tornando-a conceitualmente mais equipada para o estudo do pensamento inteligente.

O quadro geral da metodologia dos programas de pesquisas da ciência cognitiva

Conforme mencionamos na introdução, as duas vertentes da ciência cognitiva, IA e RNA, apesar de aparentemente antagônicas, compartilham pressupostos básicos, tais como as hipóteses representacionista e mecanicista da natureza dos processos mentais, bem como o ideal da unidade metodológica de explicação científica.

A hipótese representacionista, segundo a qual o nosso conhecimento do mundo é sempre mediado por representações mentais, abre um campo fértil para o desenvolvimento e teste da hipótese mecanicista, que reduz os estados mentais a estados cerebrais, regidos por leis físicas. À medida em que tais leis podem ser descritas através de regras abstratas e operadores funcionais, elas passam a ser objeto de simulação dos modelos computacionais que constituem os instrumentos explicativos fundamentais da ciência cognitiva, na sua versão funcionalista computacional. Sem tal pressuposto representacionista dificilmente o programa funcionalista de pesquisa da IA, fundado nas técnicas de processamento seqüencial de informação dos computadores digitais, poderia ter sido desenvolvido. Da mesma forma, o conexionismo, apoiando-se nas técnicas de processamento de informação em paralelo, teria seu programa de estudos bastante alterado, pois as RNAs são essencialmente treinadas para o reconhecimento de *inputs* informacionais, uma vez estabelecidas as suas *representações* através dos padrões de conectividade que se estabelecem em suas unidades.

Tendo como ponto de partida para o estudo do pensamento inteligente a hipótese representacionista da mente, os dois programas de pesquisa da ciência cognitiva, IA e RNA, vão diferenciar-se principalmente na forma de caracterizar a natureza das representações mentais. Enquanto a IA as entende como entidades abstratas, passíveis de descrição lógico-funcional, a RNA as caracteriza

como padrões de atividade que se estabelecem entre as suas unidades neurônio símile.

Ainda que estas diferentes caracterizações acerca da natureza das representações mentais conduzam a abordagens distintas da natureza dos estados mentais, ambas situam-se em um contexto científico que tem como ideal a **tese da unidade metodológica de explicação científica.** Esta tese, bastante conhecida dos filósofos da ciência, reconhece legitimidade científica apenas aos métodos de explicação característicos das *ciências naturais*, que são passíveis de ***testes empíricos*** e inferência lógica - representando um ideal de objetividade e precisão desejadas nas explicações científicas dos fenômenos, sejam eles físicos, biológicos ou sociais.

Conforme ressaltamos, críticas não têm faltado ao projeto cognitivista, com o seu ideal naturalista, subjacente às propostas de explicação dos estados mentais, em especial no que tange ao estudo dos fenômenos mentais humanos que, segundo os seus críticos, não se ajustam aos moldes explicativos das ciências naturais. Uma das críticas merecedoras de destaque provém de Putnam (1995), na sua fase recente de rejeição das suas próprias crenças acerca da eficiência da visão mecanicista da mente. Na condição de proponente central do funcionalismo computacional, que estruturou a ciência cognitiva desde a década de 60 até os nossos dias, Putnam declara:

> ...era o materialismo científico que me dominava nas décadas de 50 e 60. Eu acreditava que tudo o que existe pode ser explicado e descrito por uma única teoria... Eu acreditava que a melhor metafísica era aquela que os positivistas chamavam de 'ciência unificada'... pela aplicação de leis fundamentais da física. (Putnam, 1995, p. 2)

Rejeitando radicalmente a sua antiga busca de unificação das ciências através do espelho da física, Putnam defende atualmente um modelo de explicação dos eventos mentais fundado em um ***pluralismo metodológico***. Ele não esconde o seu pessimismo diante da crença, vigente na ciência cognitiva atual, sobre a suposta eficiência explicativa dos modelos computacionais da mente.

A postura atual de Putnam frente à visão exclusivamente mecanicista da ciência cognitiva encontra eco também entre os defensores de uma visão humanista,

anti-mecanicista, dos estados mentais, incompatível com a abordagem que eles denominam de *cientificista*. Assim, por exemplo, Button, Coulter, Lee & Sharrock (1998) desenvolvem uma investigação crítica dos principais problemas e hipóteses vigentes na ciência cognitiva sobre a natureza dos processos mentais, classificando-as como produto da confusão conceitual dominante nos estudos cientificistas da mente. Em particular, eles argumentam que a hipótese cognitivista sobre a natureza *mecânica* da mente teria suas raízes na visão cientificista do mundo sugerida por Galileo e Descartes, no interior da qual não parece restar um lugar natural para a mente no universo.

Seguindo as trilhas de Wittgenstein e Ryle na rejeição radical do cientificismo aplicado às teorias psicológicas e filosóficas da mente, Button, Coulter, Lee e Sharrock criticam os usos e abusos no emprego dos chamados "métodos experimentais" (que envolvem a simulação computacional) na avaliação das teorias da mente. Segundo os autores, a proposta de uma psicologia científica, tal como desenvolvida pelos cognitivistas, que considera a conduta humana como o produto (externo) da atividade mental (interna) modelável através do computador, ignora a relação intrínseca entre linguagem e pensamento, bem como " a relevância do fato de que a linguagem é uma instituição social." Tal concepção equivocada da natureza da mente reuniria, segundo os autores, uma série de pseudo-problemas, entre os quais se incluem a controvérsia *eliminativismo materialista versus representacionismo* sobre a natureza do mental; o problema da possível identidade entre estados mentais e estados cerebrais; o debate sobre a interioridade de uma linguagem do pensamento *versus* exterioridade dos estados mentais, manifestos através do comportamento público, e tantos outros que constituem a agenda atual dos cientistas cognitivos.

Empenhados em "...combater a concepção de filosofia que dá origem ao fetichismo do computador e do papel da pretensa racionalidade científica que nos forneceu essa máquina realmente impressionante" (p. 20), os autores enfatizam a importância da perspectiva humanista, que concebe o homem e alguns dos seus semelhantes como os ***únicos*** detentores da capacidade de pensar. Esta postura humanista, bastante comum entre os opositores da ciência cognitiva, desempenha um papel crítico fundamental no andamento do projeto cognitivista. Contudo, por vezes, ela torna difícil a compreensão da chamada ***revolução***

copernicana, supostamente realizada por esta ciência no entendimento da estrutura e natureza do pensamento inteligente. Tal revolução, que comentaremos a seguir, estabelece-se no projeto de desconstrução da metafísica da subjetividade (Dupuy, 1997). Sua proposta fundamental é o desenvolvimento de uma teoria da mente **sem** um sujeito transcendental ou absoluto, possuidor de uma alma imortal.

A noção de sujeito cognitivo que restou dessa suposta revolução está diretamente ligada àquela de ***auto-organização***, que caracteriza, por sua vez, os processos complexos cuja dinâmica de interação entre suas partes constituintes possibilita o estabelecimento de organizações, independente de regras fixas preestabelecidas e de um centro organizador único.

Os teóricos da Auto-organização insistem na necessidade de se considerar a complexidade dos estados mentais em suas investigações sobre o sujeito, levando-se em consideração a dimensão do "fluxo da vida", físico-temporal e cultural, entre outras. O ponto nevrálgico desta proposta consiste em saber **como** investigar a complexidade dos fenômenos humanos, no estudo da natureza e estrutura dos estados mentais, sem perder de vista o seu caráter **científico**. Conforme indicaremos a seguir, a grande novidade do **Perspectivismo** e da **Teoria da Auto-organização** consiste justamente em apontar caminhos para tal empreendimento complexo.

Auto-organização e Perspectivismo na análise da noção de Sujeito

A teoria da Auto-organização (**TAO**), cujo foco de análise reside na ***dinâmica*** dos processos de formação e desenvolvimento dos sistemas complexos, tem pouco a dizer à concepção funcionalista computacional da IA. Por outro lado, o seu parentesco é orgânico e fundamental com o conexionismo à medida em que este propõe uma abordagem essencialmente **bottom-up** da atividade desenvolvida nas RNAs e também investiga o processo de formação espontâneo de padrões na memória, sem um centro organizador único. Apesar deste elo orgânico, a TAO possui um domínio de atuação mais amplo do que aquele da ciência cognitiva na sua vertente da RNA, uma vez que o seu domínio não se resume ao universo cognitivo do sujeito humano. Conforme ressalta Debrun (1997):

.... nas modalidades da auto-organização em que a presença do

> ***sujeito é** importante, (e nem sempre é este o caso, uma vez que temos modalidades da auto-organização* que dispensam completamente a noção de sujeito), o que interessa é *menos a sua atividade cognitiva* em si do que a contribuição que ela pode trazer para a constituição de uma *forma*. Esta *forma*, conforme o caso, pode ser uma nova aptidão ou tendência individual - obtida por aprendizagem, ou um ajuste entre as atividades econômicas, políticas, esportivas etc. de *múltiplos* sujeitos. O único caso de auto-organização em que a atividade cognitiva interessa em si mesma é ... aquele de automontagem ou autotransformação, de uma rede cognitiva ou de sua base neural. (Debrun, 1997, p. 28)

Mas que sujeito é esse de que fala Debrun, cuja atividade cognitiva só interessa à auto-organização à medida em que ela se caracteriza como uma nova aptidão, disposição ou tendência individual - obtida através de um processo de aprendizagem ou de um ajuste entre as atividades de múltiplos sujeitos?

A resposta oferecida por Debrun a essa pergunta pode fornecer uma contribuição significativa à ciência cognitiva, no seu projeto aparentemente revolucionário, mas pouco fundamentado, de análise da mente dissociada de um sujeito cognitivo absoluto. Esse ***sujeito***, devidamente simplificado, parece ser justamente aquele ***sujeito*** que restou da suposta ***revolução copernicana*** realizada pela ciência cognitiva. Esta possui como diretriz dois eixos fundamentais: **(1)** ***o ataque 'a concepção humanista*** - que concebe o homem como dotado de uma alma imortal, distinta do corpo, que lhe garantiria o status de detentor único da capacidade de pensar e, **(2)** a proposta de ***desconstrução da metafísica da subjetividade***, que elege o sujeito cognitivo, nos moldes cartesianos, como palco de introspeção e lugar de acesso às verdades absolutas, claras e distintas.

Uma vez instaurada tal revolução, o que resta do sujeito cognitivo humano neste contexto? Que imagem teremos desse novo homem?

A proposta neomecanicista da ciência cognitiva, na sua vertente conexionista, consiste em situar esse novo homem no universo físico, onde se desenvolvem as tramas informacionais dos processos naturais. Retirado do centro do universo cognitivo, esse sujeito (um entre outros inúmeros sistemas complexos

de processamento de informação) terá sua identidade descrita em termos de processos auto-organizados que se estabelecem através do complexo cruzamento de redes causais e informacionais que moldam a sua memória e a sua forma de perceber o mundo.

A este cenário, a TAO adiciona, na sua proposta de estudo do sujeito, a análise dos processos orgânico-biológicos e sócio-culturais. Sem defender uma concepção materialista reducionista, a TAO desenvolvida por Debrun e outros teóricos da auto-organização focaliza a análise do sujeito na dinâmica do fluxo da vida. Este situa-se no espaço de interação entre elementos que, independentes de um núcleo organizador fixo, ou preestabelecido, atuam em um sujeito, agente corporificado.

Situado em um ambiente efetivo, este sujeito possui uma história que, como ressalta Debrun, ele mesmo não pode, em princípio, dominar, nem tão pouco conhecer por completo. Essa concepção *orgânica* do sujeito que, situado no mundo, constrói a sua identidade à medida em que age no meio ambiente, contrasta-se com a visão de um sujeito transcendental ou absoluto. Nas palavras de Debrun:

> Um sujeito absoluto ou quase absoluto é incompatível com a auto-organização. Pois, se fosse possível um tal sujeito, e caso ele quisesse e pudesse se programar de ponta a ponta, sua operação seria na verdade um ato de hetero-organização. É que pela radicalidade de sua exteriorização em relação a ele próprio, ele realizaria a programação ou a reprogramação não dele, mas de um outro, de um simples objeto. Ele se dividiria assim entre um ego transcendental e um ego empírico. Ao contrário, o sujeito real, efetivo, quando se relaciona com seu próprio corpo (mais o passado, o futuro e o ambiente imediato), só consegue agir, e agir sobre seu corpo (que ele tem de experimentar como tal, e como um corpo anônimo), na medida em que não o sobrevoa. (Debrun, 1997, p. 28-9)

Esta concepção geral de ***sujeito*** (efetivo) concebida por Debrun, ao distanciar-se daquela de sujeito transcendental ou absoluto, distancia-se consequentemente da tradicional noção de sujeito cognitivo racional, possuidor de uma mente imaterial estruturada de acordo com regras lógicas cognoscíveis pelo recurso à introspeção. Ela se afasta também da visão cognitivista, proposta pela IA, que concebe o homem como um exemplar da máquina de Turing, na qual não existe

qualquer possibilidade de auto-organização. O seu parentesco é com a concepção de sujeito cognitivo esboçada pelas RNAs, na sua ênfase aos mecanismos de *cooperação, competição, neutralização e ajuste,* que estabelecem a dinâmica de interação entre os elementos constituintes dos processos de auto-organização estruturadores da **memória** e, conseqüentemente, da identidade do sujeito.

No estudo do sujeito cognitivo, assim entendido, abandona-se qualquer pretensão a uma busca de verdades absolutas, buscando-se, ao invés, elementos para a sua *adequação* a uma certa perspectiva ou visão de mundo. Neste sentido, o Perspectivismo de Peterson (1996) vem complementar a proposta da TAO, uma vez que ele volta-se para a análise dos processos de adequação e ajuste de uma perspectiva adotada por um sujeito em seu contexto específico.

De acordo com o Perspectivismo de Peterson (1996), *a adequação* de uma perspectiva é sempre relativa, mas ela não se estabelece de modo arbitrário. Na escolha de uma perspectiva devemos levar em conta a sua **adequação ao usuário**, a uma **situação** determinada, a um **processo**, a uma **ontologia** e a uma **meta** específica. É justamente a análise de cada um destes ingredientes que se pretende desenvolver no estudo dos processos de aprendizagem que atuam, por exemplo, na memória dos sujeitos cognitivos. Em particular, interessa à TAO e ao Perspectivismo o entendimento da arte de ajustes e desajustes de sujeitos anônimos às situações e eventos que constituem o seu universo cognitivo.

Assim, é fácil compreender por que essa abordagem do sujeito difere da concepção proposta pela IA, segundo a qual os sujeitos cognitivos são entendidos como manipuladores de regras preestabelecidas, responsáveis pela composicionalidade de suas representações mentais, mas não da concepção conexionista, cuja ênfase reside nos mecanismos de *cooperação, competição, neutralização e ajuste* que se estabelecem no processo de auto-organização das RNAs.

Apesar dessa proximidade com a ciência cognitiva, a TAO proposta por Debrum para o estudo do sujeito supera aquelas atualmente desenvolvidas na ciência cognitiva na medida em que ela está enriquecida pela sua ênfase nos processos de **ajuste** do sujeito ao seu meio ambiente sociocultural. É o seu olhar direcionado para a trama complexa das relações sociais, principalmente naquelas estabelecidas espontaneamente, sem a presença de um líder ou de um centro organizador fixo,



que aponta para a relevância das interações circulares (de feedback) que se estabelecem entre os indivíduos. É para esse mesmo olhar que a atividade cognitiva do sujeito só interessa

> ... à medida que ela se caracteriza como uma rede de disposições ou tendências individuais - obtidas através de aprendizagem ou de ajustes entre as atividades econômicas, políticas, esportivas etc. de múltiplos sujeitos. (Debrun, 1997, p. 28).

O entendimento desse sujeito cognitivo orgânico requer um estudo interdisciplinar, que resguarde a complexidade e a riqueza informacional que o constitui. É neste sentido que a abordagem da TAO, aliada ao Perspectivismo - que propõe vários níveis complementares de análise da cognição -, poderia acrescentar uma nova e enriquecedora dimensão explicativa à ciência cognitiva, possibilitando uma análise menos simplificada do que as disponíveis atualmente nesta área. Contudo é importante ressaltar que esse estudo, além de introdutório, é necessariamente incompleto; ele se faz através da observação das tendências que se formam nos indivíduos a partir de sua experiência com a dinâmica orgânico-ambiental. Entendidos como sistemas auto-organizados, tais indivíduos mesclam-se, ao mesmo tempo que se diferenciam através da complexidade que marca os sistemas cognitivos, no fluxo auto-organizante da vida.

Referência Bibliográfica

BUTTON, G., COULTER, J., LEE, J., SHARROCK, W. *Computadores, mentes e conduta.* São Paulo: UNESP, 1998.

DEBRUN, M. Auto-organização e ciências cognitivas. In: GONZALES, M. E. Q., LUNGARZO, C., MILIDONI, C. B., PEREIRA JÚNIOR, O., WRIGLEY, M. (Org.) *Encontro com as Ciências Cognitivas.* Marília: Faculdade de Filosofia e Ciência, 1996. v. 1.

DEBRUN, M., GONZALES, M.E,Q., PESSOA JÚNIOR, O. *Auto-organização: estudos interdisciplinares.* Campinas: UNICAMP, 1997. (Coleção, CLE).

DUPUY, J. P. *Nas origens da Ciências Cognitivas.* São Paulo: UNESP, 1997.

GONZALES, M. E. Q., LUNGARZO, C., MILIDONI, C. B., PEREIRA JÚNIOR, O.,

WRIGLEY, M. *Encontro com as Ciências Cognitivas.* Marília: Faculdade de Filosofia e Ciência, 1996. v. 1.

GONZALES, M. E. Q., ALLEN, A. Self organization and the the mind brain problem. *Ciência Hoje*, v. 50, p. 30, 1998.

PETERSON, D. (Ed.) *Forms of representantion.* Exeter: Intellect books, 1996.

PUTNAM, H. *Renewing philosophy.* Harvard: Harvard University Press, 1995.

SEARLE, J. *A redescoberta da mente.* São Paulo: Martins Fontes, 1997.

PARTE II

Informação e Cognição nos modelos de Redes Neurais

unesp

25

O Nascimento da Ciência Cognitiva e suas Raízes na Física do Século XIX

Maria Eunice Quilici Gonzales

A Ciência Cognitiva é uma área de investigação interdisciplinar, cujo objeto de estudos é o processo cognitivo de sistemas que manipulam informação inteligentemente. O seu reconhecimento oficial enquanto área de investigação ocorreu, segundo Gardner (1986), no ano de 1956 quando alguns filósofos, lingüistas, físicos, matemáticos, neurofisiólogos e psicólogos se reuniram no MIT para discutir a adequação do uso da Teoria Matemática da Comunicação, MTC *(Mathematical Theory of Communication)*, no estudo dos processos e estados cognitivos humanos.

A MTC foi originalmente formulada por engenheiros, sendo o trabalho de Shannon e Weaver (1949) um dos seus principais representantes. Ainda que Shannon tenha, explicitamente, declarado que sua teoria de processamento e transmissão de informação era puramente *quantitativa*, sem qualquer preocupação com noções semânticas ligadas à concepção ordinária de informação, Weaver, na mesma obra, afirma que a MTC tal como formalizada por Shannon, fornece uma chave para o entendimento da noção de 'significado' presente na comunicação humana.

Como veremos no final deste trabalho, a obra de Shannon e Weaver se distancia do senso comum ao propor uma análise do conceito de *informação* que *não* está, em princípio, associada à atividade humana interpretativa, consciente. Contudo, essa mesma obra tem motivado a elaboração de teorias naturalistas do significado e intencionalidade presentes na comunicação humana (veja-se, por exemplo, os trabalhos de Dretske (1984), Sayre (1986) e grande parte dos modelos elaborados na Ciência Cognitiva).

O conceito de informação, tal como desenvolvido na teoria de Shannon e Weaver, foi inspirado nas idéias de Boltzmann. Em particular, a sua abordagem da segunda lei da termodinâmica serviu como uma fonte de inferência analógica para a equação de Shannon, que determina a quantidade média de informação gerada por uma fonte de informação.

No que se segue indicaremos, em três etapas, alguns aspectos metodológicos da Física no século XIX que, acreditamos, desempenharam um papel fundamental no nascimento da moderna Ciência Cognitiva, os quais estão ligados ao desenvolvimento de métodos experimentais e matemáticos para explicar o desempenho do sistema cognitivo humano.

Na Seção 1, faremos uma breve introdução à Ciência Cognitiva, enfatizando as duas áreas centrais que a constituem, as quais são conhecidas por "Inteligência Artificial" e "Redes Neurais". Na Seção 2, apresentaremos uma igualmente breve introdução às principais idéias concernentes ao estudo das Redes Neurais Artificiais. Finalmente, na Seção 3, ilustraremos algumas aplicações das hipóteses de Shannon e Weaver concernentes ao conceito de *informação* nos modelos de redes neurais.

1. Introdução à Ciência Cognitiva

A *Ciência Cognitiva* pode ser caracterizada como "...um esforço contemporâneo, empiricamente baseado, para responder antigas questões epistemológicas – particularmente aquelas concernentes à natureza do conhecimento humano, seus componentes, fontes, desenvolvimento, etc..." (GARDNER, p. 6).

O caráter empírico da Ciência Cognitiva baseia-se na suposição de que os computadores constituem instrumentos úteis para se ana-



lisar o comportamento inteligente. Eles são usados como modelos para *simular* aspectos da atividade humana inteligente e para *testar* as hipóteses e teorias acerca do funcionamento da mente. Esta suposição reflete uma crença, comum na Ciência Cognitiva, de que se for possível descrever com precisão o comportamento, ou o processo de pensamento, de um organismo, então pode-se programar um computador que operará de forma semelhante a esse organismo (GARDNER, p. 18). O sucesso de um modelo computacional, que simula o comportamento de um organismo, é considerado como boa evidência para a plausibilidade das hipóteses e teorias que este modelo descreve.

Ainda que seja bastante controversa esta suposição de que os computadores constituem instrumentos adequados para se analisar a atividade humana inteligente, o que mais interessa aos cognitivistas é que a prática de elaboração dos modelos computacionais sugere um *critério científico* para testar as teorias psicológicas. Além disso, pode-se fazer *previsões* acerca do comportamento do sistema modelado através de sua simulação computacional.

Uma vez que é possível enumerar várias diferenças relevantes entre organismos inteligentes e computadores, não é difícil apresentar críticas à hipótese da adequação dos modelos computacionais para o estudo do comportamento inteligente. Contudo, interessa-nos discutir aqui a tese funcionalista, subjacente a esta suposição. Uma versão informal do funcionalismo nos diz que, apesar das máquinas e seres humanos serem fisicamente diferentes, ambos possuem mecanismos de *processamento de informações,* cujas operações mentais/computacionais podem ser entendidas em termos de suas *organizações funcionais* (GONZALES, 1989; 1991). Funcionalistas como Putnam (1967), Fodor (1988), entre outros, sustentam que o comportamento dos organismos pode ser propriamente caracterizado em função dos seus *input/output* e das interações entre os estados internos deste organismo.

Atualmente, existem dois ramos principais do funcionalismo na Ciência Cognitiva, os quais denominamos: (i) *funcionalismo lógico-computacional* (FLC) e: (ii) *funcionalismo neuro-computacional* (FNC) (GONZALES, 1989, p. 11-20). O primeiro, FLC estuda os processos mentais como se eles fossem apenas *computações abstratas,* independentemente de suas características físicas e do seu meio

ambiente (cf. FODOR, 1988, PUTNAM, 1967, PYLYSHYN, 1984. O foco de análise aqui é o conjunto de *regras* abstratas possivelmente responsável pela estrutura funcional dos estados cognitivos, que são concebidos como elementos de um sistema de símbolos, semelhante àquele da linguagem.

Os funcionalistas *neuro*-computacionais, FNC, por outro lado, enfatizam a relevância de considerações acerca dos substratos *neurofisiológicos* e *ambientais* do sistema para o estudo e compreensão da estrutura funcional dos seus processos cognitivos. O foco de análise aqui não é em sistemas de símbolos ou regras abstratas, mas, sim, nas *leis físicas* que, possivelmente, estruturam o comportamento inteligente.

Correspondentemente às versões (i) e (ii) do funcionalismo, os cognitivistas se dividem em duas subáreas principais denominadas *Inteligência Artificial* e *Redes Neurais* ou *Conexionismo*. Os pesquisadores da *Inteligência Artificial* (AI, do inglês, *Artificial Intelligence*) (cf. NEWELL, SHAW & SIMON, 1972) são, em geral, funcionalistas lógico-computacionais. Eles julgam que os estados mentais são entidades abstratas que não apenas podem, mas devem ser estudadas independentemente dos elementos físicos que as instanciam. Segundo esta perspectiva, a capacidade de manipular e codificar *símbolos* constitui a essência da atividade computacional e representacional da mente (BICKHARD & RICHARD, 1983, p. 12). Sem a codificação de símbolos não há computação; sem as regras abstratas para ordenar os estados internos não há como estruturá-los.

Os Conexionistas, por sua vez, podem ser classificados como funcionalistas neuro-computacionais, no sentido que eles consideram as informações sobre os estados físicos/cerebrais do organismo, e a sua interação com o meio ambiente, como sendo de extrema importância para o estudo do sistema congnitivo. Ao invés de investigar os procedimentos heurísticos de manipulação de regras e símbolos, como fazem os pesquisadores da *AI* (veja-se, por exemplo, Newell, Shaw e Simon, 1972), a maior parte dos conexionistas situam seus estudos num nível mais básico de descrição. Este é conhecido como nível *subsimbólico* (SMOLENSKY, 1986) porque ele descreve estados físicos do sistema cognitivo, os quais consituem, por hipótese, o substrato de níveis de abstração hierarquicamente superiores deste sistema. Suas

unidades primitivas de análise são aquelas denominadas "neurônio-símile", que possuem propriedades semelhantes às do cérebro (retomaremos adiante esta introdução aos modelos conexionistas).

Apesar das diferenças teóricas e metodológicas existentes entre AI e Conexionistas, ambos se utilizam de recursos matemáticos para analisar o sistema cognitivo, enfatizando as relações funcionais entre *input* e *output* de máquinas ideais ou reais. O sistema cognitivo é concebido, nestes dois ramos da Ciência Cognitiva, como um autômato cujos estados podem ser descritos por regras abstratas (no caso da AI) e/ou leis matemáticas, semelhantes àquelas que descrevem as leis da natureza (no caso do Conexionismo).

Através da prática de elaboração de modelos computacionais para simular o comportamento inteligente, os cognitivistas incorporam o ideal de cientificidade que já triunfava no início do século XIX, quando o programa newtoniano imperava no universo científico. No nosso século, a reunião de técnica computacional, teoria matemática e experimentação prolonga este ideal de cientificidade, situando o projeto cognitivista no âmbito da ciência moderna. O preço a se pagar por uma tal posição no domínio científico é a extrema simplificação da natureza do mental, em função das hipóteses que a experimentação, através dos modelos computacionais, permite.

Conforme veremos, os conexionistas levaram adiante os ideais da Física do século XIX ao transportarem algumas das hipóteses da termodinâmica e da mecânica estatística para os seus modelos das redes neurais artificiais. No que se segue faremos uma introdução ao conexionismo (uma versão mais elaborada, da qual a Seção 2 é apenas um resumo, pode ser encontrada em GONZALES, 1989 E 1991).

2. Introdução ao Conexionismo

O *Conexionismo,* também conhecido como PDP *(Parallel Distributed Processing)* Processamento em Paralelo de Informação Distribuída, ou *Redes Neurais (Neural Networks)*, é aquele ramo da Ciência Cognitiva que acabamos de associar ao funcionalismo neurocomputacional. Os modelos elaborados nesta área, para simular a atividade do sistema cognitivo, são inspirados no funcionamento de sistemas dinâmicos, com características semelhantes àquelas do cérebro humano. Dadas as simplificações que tais modelos computacionais

incorporam, apenas uma vaga semelhança permanece entre os seus mecanismos e aqueles que supostamente atuam no cérebro e são responsáveis pela organização do sistema congnitivo humano.

Conforme mencionamos na Seção 1, as unidades neurônio-símile *(neuron-like units)* constituem os elementos primitivos de análise dos modelos conexionistas. Eles possuem um nível de ativação que varia em função de influências externas e das suas conexões (sinápticas) com outras unidades neurônio-símiles. A Figura 1 (extraída de GONZALES, 1991, p. 99) ilustra o funcionamento de uma tal unidade.

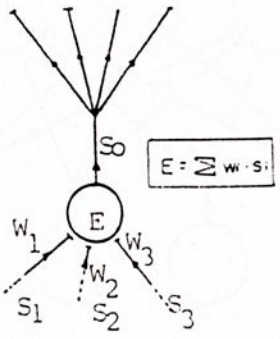


Figura 1: Esquema simplificado de uma unidade neurônio-símile

Si = intensidade do *input*
Wi = peso da sinápse
So = intensidade do *output*
E = *input* total

Estas unidades, ao serem ativadas, enviam sinais às outras unidades que, por sua vez, dependendo da intensidade do sinal recebido, transmitem um outro sinal a outras unidades, e assim sucessivamente. Populações de unidades neurônio-símiles formam redes de camadas estruturadas, contendo unidades de entrada ou *input units*; unidades de resposta, ou *output units* e, finalmente, zero ou mais camadas intermediárias entre *input* e *output*, ou *hidden units*. A Figura 2 (extraída de GONZALES, 1991, p. 100) ilustra um segmento de rede

neural com suas várias camadas.

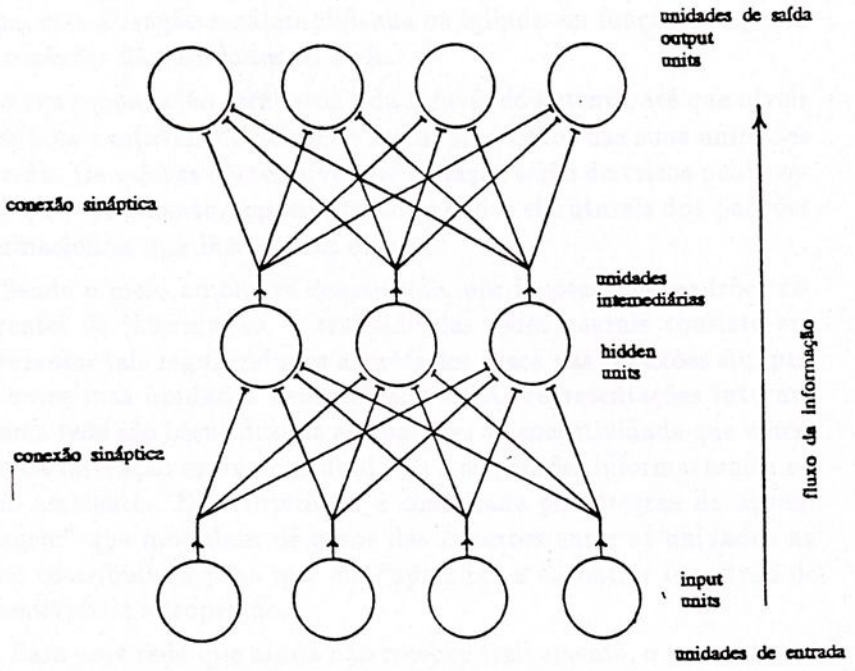


Figura 2: Ilustração das camadas de uma rede neural artificial.

O *estado informacional* de cada unidade é descrito por uma função *output*, que soma todas as cargas das ativações das unidades intermediárias e de entrada do sistema, as quais são amplificadas ou inibidas em função dos pesos das conexões entre estas unidades e de alguma função limite *(threshold)* que possa existir na rede.

O meio ambiente de uma rede artificial é constituído de padrões recorrentes de informação, os quais chamaremos de *padrões informacionais*. Estes padrões reúnem conjuntos de relações relativamente fixas entre elementos que caracterizam objetos, eventos ou situações específicas. Eles incluem, por exemplo, as freqüências e intensidades específicas dos raios luminosos que são refletidos dos objetos no campo visual do sistema modelado. Esses padrões informacionais são apresentados às redes na forma de vetores, que representam a ordenação, posição, etc. dos seus elementos.

Ao entrarem em contato com tais padrões de *input*, as unidades de entrada da rede geram níveis de ativação, os quais são propagados

quando ultrapassam um valor limite *(threshold)*. Conforme indicado acima, essa ativação será amplificada ou inibida em função dos pesos das conexões das unidades da rede.

A sua propagação será estendida através do sistema, até que níveis específicos e estáveis de ativação sejam produzidos nas suas unidades de saída. Os valores destes níveis de ativação serão descritos pelos valores que, idealmente, representarão os dados estruturais dos padrões informacionais que lhes deram origem.

Sendo o meio ambiente constituído, por hipótese, de padrões recorrentes de informação, o trabalho das redes neurais consiste em *representar* tais regularidades através dos pesos das conexões sinápticas entre suas unidades neurônio-símile. As representações internas de uma rede são identificadas aos padrões de conectividade que emergem da interação entre suas unidades e os padrões informacionais no meio ambiente. Esta interação é controlada por "regras de aprendizagem" que modulam os pesos das conexões entre as unidades na rede, contribuindo para que ela "aprenda" a encontrar o padrão de conectividade apropriado.

Para uma rede que ainda não recebeu treinamento, o peso de suas conexões poderá ter um valor inicial arbitrário, estabelecendo-se, conseqüentemente, correlações arbitrárias entre os padrões de conexões aleatórios que se formam na rede e os padrões informacionais no meio ambiente. Nesta situação, dado que não existe uma dependência entre os padrões informacionais e os padrões de conectividade, a rede não terá como evitar erros na identificação de padrões informacionais específicos. Contudo, através de treino, onde se empregam "regras" de aprendizagem, é possível ajustar gradativamente os pesos das conexões até que eles convirjam a um *padrão estável* que representará as regularidades no meio ambiente, idealmente sem ambigüidades.

À medida que elas funcionam como um indicador de padrões informacionais específicos, tais *estruturas estáveis* – emergentes da interação entre os padrões de atividade das unidades neurônio-símile e dos padrões de informação no meio ambiente – adquirem o estatuto de *representações internas* do sistema modelado. Uma vez formadas, elas possibilitam a identificação dos padrões informacionais que lhes deram origem. Além disso, o sistema em questão adquire a capacidade de realizar generalizações, no sentido que ele pode fazer



antecipações e reconhecer objetos no meio ambiente, ainda que estes apresentem pequenas variações em relação ao objeto original, para cujo reconhecimento ele foi treinado.

Um ponto relevante para o presente trabalho é que os padrões estáveis de conectividade, que representam padrões informacionais específicos, se comportam como se estivessem *minimizando* uma quantidade que desempenha o papel de energia física na rede, sendo que seu *output* é o estado do sistema em equilíbrio. Esta situação assemelha-se àquela das máquinas térmicas concebidas no século XIX nos estudos de termodinâmica. A analogia utilizada pelos conexionistas entre situações/eventos no domínio da termodinâmica e mecânica estatística caracteriza e distingue as teorias conexionistas das demais teorias psicológicas e filosóficas tradicionais do sistema cognitivo.

Para concluir, vamos investigar algumas extensões do emprego desta analogia entre os domínios da termodinâmica e do conexionismo através do estudo da noção de informação, tal como sugerida por Shannon e Weaver, sendo posteriormente desenvolvida por Drestske e outros.

3. Aplicações de hipóteses da termodinâmica nos modelos conexionistas

Grandes inovações foram introduzidas por Boltzmann na Física do século XIX, através do uso do conceito de probabilidade na sua interpretação da segunda lei da termodinâmica, proposta por Rudolf Clausius.

Resumidamente, podemos expressar o bem conhecido problema enfrentado por Boltzmann da seguinte forma:

O comportamento da energia de um sistema fechado foi descrito por Clausius em duas leis. A primeira, conhecida como "primeira lei da termodinâmica", nos diz que a energia sempre se conserva, isto é, ela não pode ser criada ou destruída. A segunda lei nos diz que ainda que a energia se conserve sempre em quantidade, ela pode perder qualidade. A medida desta perda de qualidade Clausius denominou *Entropia*, palavra de origem grega significando *transformação* (CAMPBELL, 1983).

Além disso, pelos postulados de Clausius pode-se concluir que a entropia do Universo tende irreversivelmente para um valor máximo.

Se interpretarmos a entropia como um aumento na desordem entre as partículas de um sistema, podemos dizer que a tendência dos sistemas físicos é de se tornarem cada vez mais desorganizados. Existe um estado de *equilíbrio irreversível* em relação ao qual todos eles tendem.

O problema que intrigou Boltzmann consistia em saber por que a energia obedece a esse percurso irreversível de perda de qualidade. Seguindo os passos de Maxwell, ele procurou explicar a segunda lei utilizando-se da Teoria da Probabilidade. A entropia foi caracterizada por Boltzmann como uma propriedade estatística de um grande número de partículas no microcosmo da matéria. Em particular, ele interpretou o crescimento irreversível de entropia em sistemas fechados como uma "...expressão do crescimento da desordem molecular, do esquecimento progressivo de toda dissimetria inicial, porque toda dissimetria é *improvável* em relação ao estado correspondente do número máximo de complexões" (PRIGOGINE & STENGERS, 1984, p. 99-100).

Assim, o progresso irreversível na direção do aumento de entropia é apenas *provável*. Seria possível, em princípio, que as partículas invertessem essa direção, de modo a diminuir, ao invés de aumentar, a desordem do sistema. Contudo, a probabilidade de ocorrência de um tal evento é tão pequena que, na prática, ela se torna desprezível. Pode-se prever, com alta probabilidade, que o sistema tenderá para um estado irreversível onde a entropia é máxima.

A fórmula da Segunda Lei da Termodinâmica na interpretação dada por Boltzmann é:

$$S = K.\log P$$

onde S é a entropia do sistema, K é uma constante conhecida como constante de Boltzmann e $P = 1/W$, onde W é o número total de microestados equiprováveis do sistema.

A entropia S terá um valor máximo quando a desorganização do sistema é tal que todos os arranjos dos seus microestados tornam-se tão misturados que não temos qualquer evidência para esperar que um estado particular, e não outro, ocorra.

Boltzmann observou que quanto maior for a entropia de um sistema, *menos informação* teremos sobre as suas partes constituintes (CAMPBELL, 1983). No limite não teremos como saber qual é o estado atual das mesmas.

O processo realizado por Boltzmann para propor esta interpretação da segunda lei é bastante conhecido pelos cientistas e filósofos da Ciência. A principal razão para retomar essa descrição aqui consiste na possibilidade de se investigar o uso recente das idéias de Boltzmann nos modelos conexionistas.

O elo inicial da corrente que conecta a termodinâmica às redes neurais parece residir na formulação de Shannon e Weaver da Teoria Matemática da Comunicação (MTC), à qual nos referimos na Introdução.

Inspirando-se, inicialmente, nas idéias de Boltzmann, Shannon formulou sua teoria matemática da informação em termos de probabilidade. Conforme mencionamos na introdução, esta teoria difere daquelas do senso comum ao tratar a informação como uma entidade objetiva, cuja existência independe de qualquer intérprete consciente.

Ainda que Shannon e Weaver não definam explicitamente o conceito de informação, eles afirmam que esta denomina "... uma medida da liberdade que se dispõe na seleção de uma mensagem" (SHANNON & WEAVER, 1949, p. 9). Uma *mensagem*, por sua vez, é uma seqüência de eventos (pontos, sinais, letras, etc.) transmissíveis ao longo do tempo. Quando certos eventos, *entre outros possíveis*, ocorrem num meio qualquer, M, dando origem à ocorrência de mensagens, M pode ser considerado como uma fonte geradora de informações.

Um ponto importante a ser ressaltado aqui é que, de acordo com a concepção de Shannon e Weaver, não faz sentido falar de uma única mensagem, ou item de informação, isoladamente. Devido ao seu caráter probabilístico (não necessariamente causal), ela não pode ser entendida senão em relação ao conjunto de mensagens possíveis, ao qual a mensagem atual pertence.

Assim, dado um conjunto de mensagens possíveis, a MTC fornece uma forma de se medir, em *bits*, a quantidade média de informação associada à ocorrência atual de uma dessas mensagens, através da seguinte fórmula, conhecida como *principle of maximal missing information*:

$$I(s) = -\Sigma P(s).\log P(s)$$

onde $P(s)$ é a probabilidade associada à ocorrência do evento S, e a função *log*, na base 2, possibilita medir $I(s)$ em bits.

Esta equação permite calcular o número de decisões binárias en-

volvidas no processo de redução de n estados possíveis de um sistema para um único estado s. Neste sentido, a informação indica uma quantidade que representa o número de decisões realizadas para *reduzir a incerteza* na escolha de um evento/mensagem. Quanto maior o número de estados possíveis, maior a incerteza e, conseqüentemente, maior a quantidade de *missing information* associada àquele sistema.

Para ilustrar este ponto, consideremos, por exemplo, uma situação onde existem oito presos em uma cela, sendo que apenas um deles conseguirá escapar em uma situação de fuga. Dadas as condições em que se encontram, eles concordam em escolher o possível candidato à fuga através do emprego de uma moeda. Adotando-se o procedimento de cara/coroa, o grupo é inicialmente dividido em dois subgrupos de quatro elementos. A moeda é novamente atirada de forma a se determinar o próximo subgrupo, agora de 2 elementos, no qual a última decisão recairá. Finalmente a moeda é jogada pela terceira vez no subgrupo de dois elementos, e um indivíduo específico, digamos Marcos, é escolhido.

De acordo com a equação de Shannon, a quantidade de informação associada ao fato de que Marcos foi escolhido é 3 *bits*, isto é,

$$I(s) = -[1/8.\log(1/8).8] = 3\text{bits}.$$

O resultado expressa o número de decisões binárias tomadas para se reduzir a incerteza de 8 para 1. A incerteza acerca de quem seria o candidato à fuga foi reduzida de 8 possibilidades para uma única, e a quantidade de informação associada a esta redução é 3 bits.

Contudo, se inicialmente não houvesse qualquer incerteza sobre uma tal situação, se tivesse sido determinado de antemão que Marcos seria o tal candidato, então a quantidade de informação associada a este evento seria $I(s) = 0$.

A equação de Shannon para medir a quantidade média de informação associada à ocorrência de um certo evento possui uma óbvia semelhança com a equação de Boltzmann, apresentada anteriormente. Tanto Shannon como Weaver fazem referência a Boltzmann nos seus trabalhos de 1949. Em particular, Weaver afirma, nesta obra, que a quantidade que satisfaz as condições por eles impostas à noção de informação é exatamente aquela conhecida na termodinâmica como entropia.

unesp

Existe na literatura uma ampla controvérsia sobre a exata (inexata) correlação entre as noções de entropia e informação. A opinião mais corrente é que existe apenas uma correlação formal entre as equações de Boltzmann e Shannon. Contudo, há quem acredite que o estudo desta correlação entre entropia e informação poderá conduzir a descobertas significantes. Weaver, na seguinte passagem, parece compartilhar desta suspeita ao afirmar que:

> ...quando se encontra o conceito de entropia na teoria da comunicação tem-se ...o direito de suspeitar que se encontrou algo que pode vir a manifestar-se básico e importante. Que a informação seja medida pela entropia é, afinal de contas, natural quando lembramos que, na teoria da comunicação, ela está associada à quantidade de liberdade de escolha na construção de mensagens. Assim, pode-se dizer de uma fonte de informação exatamente o que se diria de um aparato termodinâmico. "Esta situação é altamente organizada, ela não é caracterizada por alto grau de aleatoriedade ou de escolha – o que equivale a dizer que a informação (entropia) é baixa" (SHANNON & WEAVER, 1949, p. 13).

Para aumentar o grau de "entropia" a respeito da possível correlação entre entropia e informação, vale a pena lembrar a observação de Campbell, mencionada anteriormente, de que Boltzmann teria enfatizado que quanto maior a entropia de um sistema, menos informação teremos sobre as suas partes constituintes. De acordo com esta perspectiva, a alta entropia significaria falta de informação, ao contrário do que sugerem as palavras de Weaver.

Esta aparente contradição parece residir em um problema essencialmente lingüístico. Pode-se argumentar que o termo "informação", tal como possivelmente utilizado por Boltzmann, situa-se ao nível do senso comum, enquanto que na concepção de Shannon e Weaver ela está associada à liberdade de escolha na construção de mensagens. Uma vez estabelecidos corretamente os termos do vocabulário utilizado por ambos, a contradição provavelmente desapareceria. Existem outras complicações mais sérias do que esta na tentativa de se estabelecer paralelos entre as noções de informação e entropia. Não entraremos em detalhes sobre elas aqui (para uma discussão enfática sobre este tópico ver CAMPBELL, 1982 e STONIER, 1990, entre outros). No que se segue, investigaremos algumas aplicações das hipóteses da MTC e da termodinâmica nos modelos conexionistas.

Antes de investigarmos tais aplicações convém comentar uma di-

ficuldade concernente ao emprego da MTC no estudo do sistema cognitivo. A dificuldade reside no fato que a MTC é puramente quantitativa; ela fornece apenas meios de se medir a quantidade de informação associada à ocorrência de eventos interrelacionados. Contudo, quando estudamos a informação no contexto do sistema cognitivo estamos principalmente preocupados com os seus aspectos semânticos, que não se reduzem a fatores puramente quantitativos. Esta dificuldade é discutida por Dretske (1981) que, partindo das idéias de Shannon, desenvolve uma teoria naturalista do significado, envolvendo a noção de conteúdo semântico de um estado informacional (uma análise das idéias de Dretske sobre este tópico foi desenvolvida em GONZALES, 1989).

Se por um lado a análise quantitativa de Shannon apresenta uma dificuldade para o seu emprego no estudo do sistema cognitivo, por outro lado a sua hipótese de que a informação é uma entidade objetiva, que pode ser analisada independentemente do conteúdo específico das mensagens, tem sido amplamente utilizada pelos cognitivistas. Os modelos elaborados na Inteligência Artificial, por exemplo, que manipulam estruturas de símbolos, não envolvem especificações do conteúdo informacional/semântico dos mesmos. Entretanto, é óbvio que uma análise menos simplificada dos estados cognitivos não pode ser realizada sem a consideração do seu conteúdo informacional. Retomaremos esta discussão mais adiante. Agora vamos investigar, finalmente, como se aplicam nas redes neurais os princípios da MTC e da termodinâmica que acabamos de apresentar.

Uma ponte entre as hipóteses da MTC e da termodinâmica foi sugerida por Smolensky (1986) para o estudo do sistema cognitivo. Ele esboçou uma teoria conexionista das representações mentais, no domínio do conhecimento perceptual, conhecida como *Teoria Harmônica (Harmony Theory)*. Uma hipótese geral desta teoria, presente nas demais abordagens de redes neurais, é que os pesos das conexões entre as unidades neurônio-símile de uma rede neural artificial codificam a freqüência com que os padrões informacionais invariantes ocorrem no meio ambiente. Quanto mais freqüentemente um invariante for detectado mais alto será o valor dos pesos das conexões dos padrões que o registram.

Uma função central do treino nas redes neurais, através do qual

mecanismos de aprendizagem são utilizados para ajustar os pesos de suas conexões sinápticas, é justamente reduzir o valor correspondente ao grau de entropia na rede. Nos modelos sugeridos pela Teoria Harmônica esta redução é feita através de um procedimento conhecido como *simulated annealing* onde a temperatura computacional, T, da rede é inicialmente fixada em um valor bastante alto, sendo gradulmente diminuído à medida em que a rede vai sendo treinada. Quando o valor de T atinge um mínimo, estruturas estáveis são formadas entre os seus padrões de conectividade. Estas estruturas constituem as representações internas da rede. Conforme mencionamos na Seção 2, elas se comportam como se estivessem minimizando uma certa quantidade que desempenha o papel de energia física da rede.

Os padrões de conectividade que registram regularidades no meio ambiente são denominados, na Teoria Harmônica, "átomos de conhecimento". Conjuntos de átomos de conhecimento que possuem um certo grau de coerência formam os *esquemas*. Este grau de coerência, por sua vez, é medido através de uma função, *H(r,a)*, denominada *função harmonia*, que estima o número de concordâncias entre os elementos do vetor de *input* r e os elementos do vetor do átomo de conhecimento "a". Quanto maior for este número de concordância, maior será o valor da sua função harmonia. Os átomos de conhecimento cujos valores da função harmonia são máximos constituem os principais candidatos aos esquemas. A sua existência reduz a entropia da rede uma vez que o estabelecimento de um esquema inibe, ou desativa, vários outros possíveis padrões de conectividade que podem ter sido inicialmente ativados quando um padrão informacional é apresentado às suas unidades de *input*.

Uma vez estabelecidos os esquemas, as redes podem efetuar inferências, no sentido que elas identificam e completam padrões informacionais. Este reconhecimento é feito através de mecanismos que comparam os padrões informacionais de *input* com os esquemas criados naqueles padrões de conectividade que incorporam as estruturas gerais, das quais o caso em questão é apenas uma instância particular.

A tese fundamental de Smolensky na Teoria Harmônica afirma que o sistema cognitivo (modelado através de redes neurais com apenas duas camadas) tende a selecionar aquelas freqüências de ocorrências dos invariantes ambientais que se harmonizam com as freqüências



aprendidas através do treino e registradas nos esquemas. Em outras palavras, quando novos estímulos chegam ao sistema cognitivo ele tende a detectar aqueles elementos que concordam ou se harmonizam com os esquemas que se encontram nos seus padrões de conectividade. Estímulos não compatíveis com tais esquemas aumentam o grau de entropia do sistema, diminuindo o valor correspondente à medida da função harmonia, $H(r, a)$. Em tais circunstâncias, a Teoria Harmônica estabelece que novos conjuntos de estímulos sejam examinados e que elementos mais compatíveis com os esquemas já adquiridos sejam selecionados.

A relação entre a Teoria Harmônica e a Termodinâmica dispensa, acreditamos, maiores comentários no momento. Resta agora saber qual é a sua correlação com a MTC.

Como vimos, a teoria proposta por Smolensky sugere uma descrição da dinâmica do processo cognitivo, na esfera do conhecimento perceptual, especialmente no que se refere à representação de eventos que ocorrem no meio ambiente. De acordo com esta teoria, os atos perceptuais podem ser caracterizados em termos da aquisição de esquemas perceptuais consistentes.

A hipótese subjacente a esta concepção da dinâmica do sistema perceptual pode ser expressa através de um princípio que denominamos "princípio de relevância" (GONZALES, 1991). De acordo com este princípio, sistemas de processamento de informação semelhantes ao cérebro humano maximizam a eficiência de seu comportamento através da seleção coerente de informação disponível no meio ambiente, sendo que tal seleção deve envolver um mínimo de energia útil por parte do sistema. Este princípio é chamado 'princípio de relevância" porque, dado um conjunto de padrões informacionais alternativos, ele indica um critério de seleção dos padrões cuja representação se ajusta, consistentemente, ao conjunto das representações existentes no sistema.

Colocado em termos procedurais, este princípio nos diz que para um sistema selecionar o padrão de *input* mais apropriado, dados os seus esquemas disponíveis e a observação da freqüência com que tais padrões se repetem no meio ambiente, ele deve:

(a) Escolher o padrão de *input* que seja mais consistente com a informação já disponível;



(b) Não superestimar o grau de ordem existente no meio ambiente.

A condição (a) pode ser realizada utilizando-se a função harmonia, $H(r, a)$, para determinar o grau de consistência dos padrões disponíveis. Para a realização da condição (b) Smolensky sugere que se utilize o princípio *maximal missing information* de Shannon. Esta correlação entre a Teoria Harmônica e o Princípio de Informação de Shannon pode ser ilustrada através do seguinte exemplo oferecido por Smolensky:

Considera-se um meio ambiente muito simples que comporta apenas quatro eventos E1, E2, E3 e E4, sendo que os dois primeiros eventos ocorrem com probabilidades iguais de 0.4. Os eventos E3 e E4 ocorrem com probabilidades 0.7 e 0.1, respectivamente. De acordo com o princípio de *maximal missing information*, a quantidade de informação gerada pela ocorrência do primeiro conjunto de eventos (mais homogêneo que o segundo conjunto de eventos) é:

$$I(E1, E2) = (0.4.1n(0.4) + 0.4.In(0.4)) = 0.73 \ .$$

A ocorrência do segundo conjunto de eventos (menos homogêneo que o primeiro) é:

$$I(E3, E4) = (0.7.In(0.7) + 0.1.In(0.1)) = 0.48 \ .$$

Admitindo-se que em um determinado momento o sistema modelado possui apenas a informação de que o evento E1 ocorre com uma freqüência de 40%, como ele estimaria a probabilidade de ocorrência dos outros eventos? Smolensky sugere que o sistema estime uma tal probabilidade sem superestimar o grau de ordem do meio ambiente. Assim, ele deverá escolher a distribuição de probabilidades mais homogênea, à qual está associado um maior grau de incerteza. Ao escolher a distribuição mais homogênea o sistema estará menos comprometido com qualquer particular estimativa da freqüência de ocorrência de eventos desconhecidos. Dado algum conhecimento prévio, este pressuposto diminui a possibilidade de inconsistências na representação da freqüência dos eventos desconhecidos, contribuindo para a satisfação da condição (a).

O trabalho de Smolensky deixa claro que existe uma relação exponencial entre a função harmonia, $H(r, a)$, e a probabilidade que

um certo padrão de *input* seja selecionado no meio ambiente, dado um conjunto de átomos de conhecimento. Esta correlação é expressa pelo seu Teorema da Competência *(Competence Theorem)*

$$p(r,a) \ \alpha \ e^{H(r,a)},$$

onde r é o vetor correspondente ao padrão de *input* e a é o vetor do átomo de conhecimento pré-existente na rede.

Smolensky demonstra que a correlação acima pode ser derivada do princípio de "maximal missing information" de Shannon, estabelecendo assim uma conexão entre os princípios da Termodinâmica, que estão presentes na sua teoria, e aqueles da MTC.

Como os modelos da Teoria Harmônica não são determinísticos, o teorema da competência se aplica apenas de forma probabilística. Algumas vezes, padrões informacionais, ou aspectos destes padrões, são representados tornando-se inconsistentes em relação aos esquemas pré-existentes. Em tal situação, a tendência da rede será aquela de reorganizar os seus átomos de conhecimento ou "ignorar" a nova representação.

Este esboço de teoria (como a caracteriza o seu autor) reconhecidamente necessita de um maior desenvolvimento. Obviamente, a complexidade do sistema cognitivo é muito maior que esta breve investigação pode sugerir. Em particular, o sistema cognitivo humano se desenvolve em um meio informacional extremamente complexo, o qual inclui, entre outras coisas, informações a respeito de outras mentes.

A esperança dos estudiosos desta área repousa na expectativa de que a Ciência Cognitiva, que está ainda engatinhando, consiga ultrapassar o presente estágio, em que grandes simplificações se impõem às suas análises. Da nossa perspectiva, esperamos que as raízes da Física do século XIX, que procuramos situar no interior da Ciência Cognitiva, possam desenvolver-se num solo fértil, senão para explicar, mas para fornecer alguns subsídios à pesquisa do sistema cognitivo humano.

Bibliografia

ARMSTRONG, D.M. *A Materialist Theory of the Mind.* New York : Humanities Press, 1968.

BAKER, L.R. *Saving Beliefs: A Critique of Physicalism.* New Jersey : Princeton University Press, 1987.

BICHARD, M. & RICHARD, D. On the Nature of Mental Representations: a Case Study of James Gibson's. *Theory of Perception.* London : Praeger, 1983.

CAMPBELL, J. *Grammatical Man.* Suffolk : The Chaucer Press, 1982.

CHURCHLAND, P.M. Cognitive Activity in Artificial Neural Networks. In: OSHERSON, D.N. & SMITH, E.E. (eds.). *Thinking.* Cambridge : MIT Press, 1990. V.3, p. 199-227.

DREYFUS, H. *What Computers Can't Do.* New York : Harper and Row, 1972.

DRETSKE, F. *Knowledge and the Flow of Information.* Cambridge : Bradford Books, 1981.

FODOR, J.A. & PYLYSHYN Z.W. Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, n. 28, p. 3-71, 1988.

GARDNER, H. *The Mind's New Science: A History of Cognitive Revolution.* New York : Basic Books, 1987.

GIBSON, J.J. *The Ecological Approach to Visual Perception.* Boston : Houghton Mifflin, 1979.

GONZALES, M.E.Q. *Approach to Visual Perception.* Colchester : Essex University, 1989. (Tese de Doutoramento).

GONZALES, M.E.Q., FRENCH, T. & TREFFNER, P. A Naturalistic Approach to Mental Representation. In: GILHOLY, K., KEANE, M. & ERDOS, G. (eds.). *Lines of Thinking: Reflections on the Psychology of Thought.* London : John Willey & Sons, 1990, p. 57-69.

GONZALES, M.E.Q. Redes Neurais e Representação Mental: Um Ensaio Sobre a Harmonia e a Racionalidade. *Trans/Form/Ação,* v. 14, p. 93-108, 1991.

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

GOSCHKE, T. & KOPPELBERG, D. The Concept of Representation and the Representation of Concepts in Connectionist Models. In: RAMSEY, W., STICH, S. & RUMELHART, D. (ed.). *Philosophy and Connectionist Theory.* New Jersey : Lawrence Erlbaum Associates, Hillsdale, 1991.

HEBB, D.O. *Organization of Behavior.* New York : John Wiley & Sons, 1949.

NEWELL, A., SHAW, J. & SIMON, H. *Human Problem Solving.* New Jersey : Englewood Cliffs, 1972.

PUTNAM, H. The Mental Life of Some Machines. In: CASTAÑEDA, H.N. (ed.). *Intensionality, Minds and Perception;* Detroit : Wayne State University Press, 1967.

PYLYSHYN, Z.W. *Computation and Cognition.* Cambridge (Ma) : MIT Press, 1984.

PRIGOGINE, I. & STENGES, I. *A Nova Aliança.* Trad. de Miguel Faria & Maria Joaquina M. Trincheira. Brasília : Ed. Universidade de Brasília, 1984.

RUMELHART, D. & MCCLELLAND, J. *Explorations in Parallel Distributed Processing.* Cambridge : MIT Press, 1986. V.1.

SAYRE, K. Intentionality and Information Processing: An Alternative Model for Cognitive Science. *Behavioral and Brain Sciences*, v. 9, p. 121-166, 1986.

SCHEERER, E. Towards a History of Cognitive Science. *Cognitive Science*, n. 115, p. 7-9, 1988.

SHANNON, C. & WEAVER, W. *The Mathematical Theory of Communication;* Illinois : The University of Illinois Press, 1949.

SMOLENSKY, P. Information Processing in Dynamical Systems: Foundations of Harmony Theory. In: RUMELHART, D. & MACCLELLAND, J., 1986, p. 390-431.

STONIER, T. *Information and the Internal Structure of the Universe;* London : Springer Verlag, 1990.

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

WIENER, N. *Cybernetics, Thought and Reality.* Cambridge (Ma) : MIT Press, 1961.

UM ESTUDO COGNITIVO-INFORMACIONAL DAS REPRESENTAÇÕES MENTAIS

Maria Eunice Quilici Gonzales
Departamento de Filosofia
UNESP, Marília, SP

A palavra informação aparece com muita freqüência nos trabalhos desenvolvidos atualmente na área de Ciência Cognitiva. Em particular, a hipótese de que somos um tipo especial de sistema de processamento de informação possibilita aos cognitivistas utilizarem computadores para estudar (simular) o comportamento humano inteligente. Apesar de seu uso constante, o conceito de informação é raramente explicitado nessa área, que tem deixado para o futuro o seu possível esclarecimento.

Conforme foi sugerido por Küppers (1990), vivemos na era da informação, mas estamos em situação quase similar àquela do homem da Era do Ferro que, apesar de manipular e viver cercado por instrumentos de ferro, não dispunha de instrumentos conceituais apro- priados para explicar a natureza química ou física desse elemento. Seria função das gerações posteriores elaborar uma teoria que permitisse caracterizar o ferro a nível molecular, com precisão, distinguindo-o do alumínio, do cobre e de outros metais.

Embora existam algumas obras na literatura que procuram explicitar a natureza da informação, ainda estamos longe de um consenso a seu respeito. No momento, a noção mais difundida de informação é aquela do senso comum, de acordo com a qual a informação é um produto abstrato da mente humana consciente. Essa noção difere daquelas tais como a de "matéria", de "energia", etc., que admitem uma existência objetiva, independente de qualquer ser consciente.

Introduziremos aqui uma hipótese que foi proposta por Gibson (1966), sendo retomada de tempos em tempos por filósofos e cientistas, segundo a qual a informação deve ser entendida como uma "enti-

Publicado em: Paulo Abrantes (organizador). Epistemologia e Cognição. Editora da Universidade de Brasília, 1994

dade" que existe independentemente da existência de qualquer atividade consciente, interpretativa, por parte de um sujeito.

Essa hipótese será investigada na área de Ciência Cognitiva, sugerindo-se algumas possíveis contribuições que o seu emprego poderia trazer ao estudo das representações mentais, especificamente relacionadas à percepção visual, tal como caracterizada nos modelos conexionistas ou de redes neurais artificiais.

Conforme será explicitado a seguir, as representações mentais são caracterizadas nos modelos conexionistas em termos das propriedades emergentes de unidades neurônio-símile (*neuron like units*) que podem ser ativadas ou inibidas, e cujas conexões são moduladas de acordo com princípios físicos que atuam sobre a rede.

Nossa proposta de estudar os modelos conexionistas de representação mental, a partir da noção de informação desenvolvida por Gibson, necessita de uma breve justificativa. Isso porque a posição anti-representacionalista (eliminativista por excelência) de Gibson é amplamente difundida na literatura, em particular naquela da escola da percepção direta (*direct perception*). Sabidamente, Gibson recusou-se terminantemente a admitir a existência de representações mentais (entendidas como entidades abstratas) na percepção ou ação humana, desenvolvendo vários argumentos em defesa de sua postura eliminativista. Contudo, acreditamos que ele estaria disposto a considerar a possibilidade da existência das representações mentais, entendidas no sentido conexionista, a ser explicado mais adiante. Entretanto, independentemente de sua aceitação, interessa-nos no momento investigar a plausibilidade do emprego, nos modelos conexionistas, do conceito de informação por ele desenvolvido, o qual não pressupõe qualquer intérprete consciente para manipulá-la. Esta noção "objetiva" de informação, que introduziremos mais adiante, apóia-se também nos trabalhos de Shannon e Weaver (1949), Dretske (1981), Sayre (1986), Küppers (1990) e Stonier (1990), bem como na Teoria da Harmonia (*Harmony Theory*) formulada por Smolensky (1986).

Como a Ciência Cognitiva ainda está pouco difundida no Brasil, faremos uma breve discussão dos seus principais pressupostos, de forma a poder situar o presente estudo nessa área (para uma versão mais detalhada deste tópico ver Gonzales, 1989). Esse estudo será desenvolvido nas seguintes etapas:

1) Principais pressupostos da Ciência Cognitiva.

2) O problema da natureza das representações mentais na Ciência Cognitiva.



3) Contribuições da concepção gibsoniana de informação para o estudo dos modelos conexionistas.

PRINCIPAIS PRESSUPOSTOS DA CIÊNCIA COGNITIVA

A Ciência Cognitiva pode ser caracterizada como uma área multidisciplinar de estudos cujo objetivo central é a elaboração de modelos e teorias científicas dos processos cognitivos humanos. Ela existe há mais ou menos trinta anos e, embora o termo Ciência Cognitiva tenha sido oficializado em 1973 (Scheerer, 1988, p.7), suas raízes filosóficas podem ser encontradas nos antigos debates gregos sobre a natureza do conhecimento humano.

Uma diferença óbvia entre a abordagem dos antigos gregos e a dos modernos cognitivistas reside no uso recente de técnicas computacionais para estudar o comportamento inteligente e para testar as teorias propostas sobre a estrutura dos processos mentais.

Vários são os pressupostos básicos encontrados nos manuais, nos artigos e nos modelos produzidos na Ciência Cognitiva. Entre eles podemos enumerar os seguintes:

1) O estudo do conhecimento humano requer uma investigação das capacidades representacionais e computacionais da mente (Scheerer, 1988).

2) O estudo científico da mente deve ser desenvolvido a partir de uma perspectiva interdisciplinar.

3) Os computadores, apesar de diferirem materialmente dos organismos, fornecem bons modelos para o estudo do sistema cognitivo humano.

O pressuposto nº 1, que enfatiza o estudo dos aspectos representacionais da mente, não é, obviamente, uma propriedade exclusiva dos estudos dos modernos cognitivistas. Podemos voltar pelo menos até Descartes que, no século XVII, desenvolveu um estudo sistemático das idéias ou representações mentais. Em particular, ele defendeu a hipótese de que o nosso conhecimento do mundo é sempre mediado por tais entidades (abstratas) denominadas idéias, que são "como as imagens das coisas".

Para Descartes, as representações mentais não são entidades físicas; elas não ocupam lugar no espaço nem possuem qualquer das propriedades características dos corpos materiais. Discussões infindáveis sobre a natureza enigmática dessas entidades, que funcionam como mediadores entre o sujeito e o mundo, constituem um dos núcleos fundamentais da Ciência Cognitiva. Mais adiante resumiremos as duas principais tendências existentes atualmente no tratamento desse tópico.

A segunda pressuposição, sobre a interdisciplinaridade na Ciência Cognitiva, é uma decorrência do entendimento de que qualquer estudo relativo ao sistema cognitivo humano é extremamente complexo. Em particular, existem dificuldades decorrentes do fato de que, nessa área de estudos, nos confrontamos com a bizarra situação em que a própria mente tenta entender e explicar a si própria.

Esse problema, provavelmente sem solução (que já foi apontado por Kant), deixa qualquer estudante de Ciência Cognitiva receoso, pois, se não podemos nos distanciar do nosso objeto de estudos, como poderemos analisá-lo cientificamente? Diante de um problema de tal complexidade, os cognitivistas, provenientes princi- palmente da Filosofia, Física, Matemática, Lingüística, Psicologia e Neurofisiologia, procuram unir forças. Eles objetivam a criação de um campo unificado de pesquisa que, idealmente, estabelecerá um vocabulário comum entre os seus pesquisadores.

Finalmente, o terceiro pressuposto, que os computadores constituem objetos úteis para o estudo do sistema cognitivo humano, reflete a crença, fundamental na Ciência Cognitiva, de que os computadores podem servir como modelos adequados para simular o desempenho da mente humana e para testar a precisão das hipóteses concernentes aos seus mecanismos de funcionamento.

Esse último pressuposto é muito controverso, particularmente entre filósofos e psicólogos que contestam a validade da hipótese a ele subjacente, segundo a qual, se for possível descrever com precisão os mecanismos internos responsáveis pelo comportamento inteligente, então deve ser possível programar um computador que simulará o comportamento em questão. Não entraremos em detalhes sobre tal controvérsia. Uma discussão mais detalhada pode ser encontrada em Baker (1987), Dreyfus (1972) e Gonzales (1989), entre outros.

Para o presente estudo dos processos cognitivo-informacionais possivelmente envolvidos na formação das representações mentais,

unesp

interessa-nos principalmente esclarecer a tese funcionalista, que sustenta o pressuposto nº 3. Essa tese nos diz que, apesar da diferença material existente entre os seres humanos e as máquinas, ambos possuem um ponto em comum que reside na capacidade de processar informação. Além disso, interessa ao cognitivista-funcionalista estudar a estrutura dos sistemas de processamento de informação, de modo a explicar como as partes desses sistemas se organizam na produção de estados cognitivos.

O funcionalismo possui duas vertentes centrais na Ciência Cognitiva, as quais denominamos "funcionalismo lógico-computacional" (FLC) e "funcionalismo neurocomputacional" (FNC) (Gonzales, 1989, 1990).

De acordo com o FLC, os estados cognitivos podem ser apropriadamente caracterizados como entidades abstratas e, como tal, devem ser estudados independentemente do seu substrato físico. Trata-se aqui de desenvolver exclusivamente uma análise das regras de composição e funcionamento do sistema cognitivo.

Já o FNC assume que considerações acerca da estrutura e organização dos substratos físicos dos sistemas de processamento de informação, bem como de sua interação com o meio ambiente, são fundamentais no estudo da cognição.

Tanto o FLC como o FNC estão interessados na forma de organização dos sistemas cognitivo-informacionais. Contudo, enquanto o primeiro enfatiza o aspecto lógico, abstrato, desses sistemas, o segundo dá ênfase aos aspectos físicos responsáveis pela sua organização.

De acordo com essas duas vertentes do funcionalismo, temos duas tradições centrais na Ciência Cognitiva. Uma delas é a conhecida Inteligência Artificial, que estuda os aspectos racionais da cognição, com ênfase nos processos de resolução de problemas, métodos de demonstração de teoremas e regras abstratas de manipulação de símbolos.

A outra tradição, correspondente à vertente do FNC, é conhecida como Conexionismo, Redes Neurais ou Sistemas de Processamento de Informação em Paralelo (PDP). A ênfase aqui não é tanto nos aspectos racionais dos sistemas cognitivos, mas, sim, nos seus aspectos adaptativos.

O nível explicativo dos modelos dessa segunda vertente difere daquele da Inteligência Artificial tradicional em pelo menos um

aspecto fundamental. Pode-se considerar que enquanto os modelos da Inteligência Artificial são elaborados levando-se em conta os elementos simbólicos e abstratos do sistema cognitivo, os modelos conexionistas lidam essencialmente com os seus elementos físicos, também conhecidos como subsimbólicos.

O termo subsimbólico foi sugerido por Smolensky (1986) para caracterizar aqueles estados e processos físicos dos sistemas naturais e artificiais que funcionam como substratos dos estados simbólicos abstratos. A hipótese de Smolensky é que os subsímbolos constituem os elementos básicos fundamentais dos processos simbólicos que são hierarquicamente menos primitivos na escala de abstração do sistema cognitivo. A análise dos subsímbolos remete-nos ao nível microscópico da cognição, que é particularmente relevante para os modelos conexionistas. Em contraste, interessa particularmente à Inteligência Artificial o nível macroscópico de análise, que lida com estruturas simbólicas e suas regras de composição.

No caso específico do estudo das representações mentais, os modelos conexionistas são desenvolvidos admitindo-se que, antes de mais nada, se trata de explicar a relação entre a macro e a micro estrutura de sistemas físicos que se adaptam às mais variadas situações do meio ambiente. É justamente esse aspecto adaptativo, de ajuste de estados físicos do sistema cognitivo aos estados do meio ambiente, que será o objeto das nossas investigações feitas a partir da teoria da informação.

O PROBLEMA DA NATUREZA DAS REPRESENTAÇÕES MENTAIS NA CIÊNCIA COGNITIVA

As representações mentais são analisadas nos modelos conexionistas através de recursos computacionais inspirados principalmente na Neurofisiologia e na Física. Da primeira foram adaptados conceitos para descrever a estrutura adaptativa das representações em termos de propriedades emergentes de unidades neurônio-símile, que podem ser ativadas ou inibidas em função das suas conexões ("sinápticas") com outras unidades neurônio-símile e com o meio ambiente.



Da Física foram emprestados conceitos, principalmente da Termodinâmica e da Mecânica Estatística, relacionados às noções de energia, entropia, organização, estados de equilíbrio, entre outros. Tais conceitos possibilitam uma análise do sistema cognitivo, no nível microscópico, ou subsimbólico. Conforme foi mencionado no início deste capítulo, investiga-se nesse nível os estados físicos do sistema que constituem, por hipótese, o substrato dos seus níveis de abstração hierarquicamente superiores.

É principalmente nesse nível subsimbólico que as representações mentais são analisadas nos modelos conexionistas. No que se segue faremos uma breve introdução às propriedades centrais desses modelos, tal como apresentadas em Rumelhart e MacClelland (1986), e resumidas em Gonzales (1989, 1990).

Qualquer modelo conexionista das representações mentais deve possuir um conjunto de unidades neurônio-símile, que são os seus elementos constitutivos básicos. Dada a sua simplicidade, elas guardam apenas uma vaga semelhança com os neurônios reais, mas preservam uma de suas propriedades centrais que consiste na capacidade de, quando ativadas, transmitirem sinais às outras unidades em função de suas conexões sinápticas.

As unidades neurônio-símile, quando ativadas, formam redes constituídas por camadas estruturadas: uma camada de entrada ou *input units*; uma camada de unidades de resposta, ou *output units* e, finalmente, zero ou mais camadas intermediárias entre *input* e *output*, ou *hidden units*. A figura 1 ilustra o esquema de uma rede neural artificial com essas três camadas.

O estado informacional das unidades em uma rede é descrito por uma função *output*, que soma todas as cargas de ativação das suas unidades intermediárias e de *input*. Tais cargas podem ser ampliadas ou inibidas em função dos pesos das conexões existentes entre as unidades.

No caso de modelos da percepção visual, as unidades de *input* são ativadas pelos estímulos provenientes do meio ambiente, os quais são, por hipótese, estruturados na forma de padrões luminosos. Esses padrões serão aqui denominados padrões informacionais; eles contêm invariantes que caracterizam classes de objetos devido à sua consti-

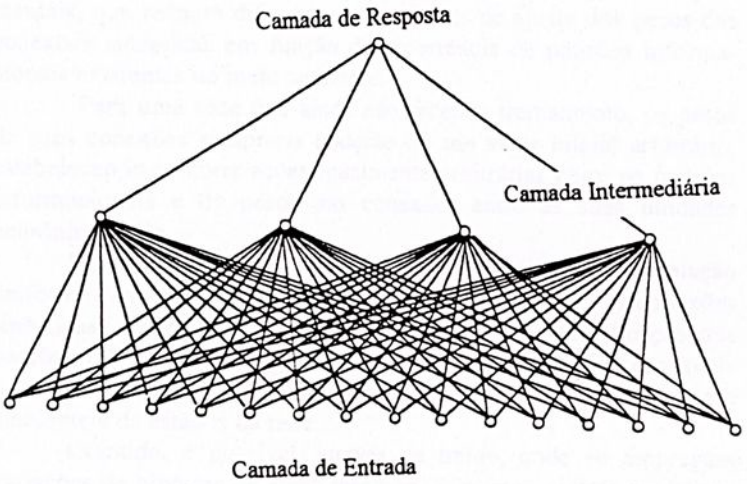


Figura 1 – Ilustração de uma rede neural artificial simplificada que possui as camadas de entrada (*input*), resposta (*output*) e intermediária (*hidden layer*)

tuição físico-química, localização espacial, forma, etc. Idealmente, as propriedades determinantes de um padrão informacional podem ser descritas na forma de vetores, que descrevem, por exemplo, as freqüências e intensidades específicas dos raios luminosos que são refletidos dos objetos com uma determinada regularidade.

Uma função central das redes neurais será, então, registrar tais regularidades nos pesos de suas conexões sinápticas, de tal forma a incorporar a hipótese segundo a qual quanto maior for o peso de determinadas conexões tanto maior será a freqüência de ocorrência de determinadas regularidades no meio ambiente.

Essa hipótese, originalmente proposta por Hebb (1949), diz que: a) estímulos sensoriais, provenientes do meio ambiente, deixam um traço na memória, cuja intensidade depende da intensidade das conexões sinápticas; b) a interação entre unidades altamente ativadas resulta em um acréscimo no peso de suas conexões sinápticas.

Adaptações dessas idéias de Hebb foram realizadas pelos conexionistas na elaboração dos seus modelos das representações mentais, que reúnem diferentes mecanismos de ajuste dos pesos das conexões sinápticas em função da recorrência de padrões informacionais existentes no meio ambiente.

Para uma rede que ainda não recebeu treinamento, os pesos de suas conexões sinápticas poderão ter um valor inicial arbitrário, estabelecendo-se correlações igualmente arbitrárias entre os padrões informacionais e os pesos das conexões entre as suas unidades neurônio-símile.

Nessas condições, dado que não existe qualquer correlação uniforme entre os padrões informacionais e os pesos das conexões sinápticas, a rede não terá como evitar erros na identificação dos padrões informacionais. Não se falará aqui de representações internas, mas apenas de correlações casuais entre os estados do meio ambiente e os estados da rede.

Contudo, é possível, através de treino, onde se empregam variações da hipótese de Hebb mencionada acima, ajustar gradativamente os pesos das conexões entre as unidades neurônio-símile, que foram ativadas por um padrão informacional, até que elas possam convergir para uma estrutura estável que, finalmente, representará o padrão informacional em questão.

Tais estruturas estáveis, que resultam da interação recorrente entre as unidades neurônio-símile e os padrões informacionais no meio ambiente, constituem as representações internas do sistema. Uma vez criadas, elas condicionam as probabilidades de escolha dos dados disponíveis no meio ambiente, em função de sua compatibilidade com o padrão de conectividades mais ativado. A partir daí, a rede condicionará de maneira crescente o seu próprio desenvolvimento, no que diz respeito à seleção dos possíveis elementos que ativarão as suas unidades. Além disso, as representações, uma vez estabelecidas, permitem à rede neural identificar os padrões informacionais que lhes deram origem ou, ainda, reconhecer no ambiente situações ou elementos semelhantes aos padrões aprendidos.



Essa caracterização das representações estabelecidas nos modelos conexionistas — em termos dos padrões de conectividade emergentes das unidades neurônio-símile, que possibilitam a identificação de padrões no meio ambiente — parece situar-se entre duas vertentes filosóficas antagônicas, conhecidas como representacionalismo e eliminativismo.

De acordo com a concepção representacionalista, à qual se ajusta o FLC (e em particular a Inteligência Artificial), as representações mentais desempenham o papel de mediador entre o meio ambiente e o sujeito, possibilitando a este identificar os objetos ou padrões informacionais com os quais ele interage.

Ainda que a identificação de padrões seja uma tarefa que parece envolver representações, tanto nos modelos conexionistas como nos representacionalistas tradicionais, a sua natureza é distinta em cada um destes, pois as representações concebidas pelos representacionalistas são entidades abstratas que não apenas podem, mas devem ser analisadas independentemente dos elementos físicos que as instanciam. Nos modelos conexionistas, em contraste, as representações não podem ser caracterizadas independentemente das unidades neurônio-símile que lhes deram origem, uma vez que a informação do meio ambiente é representada nos padrões de conecti- vidade que se formam entre estas unidades.

Além disso, o conexionismo caracteriza as representações mentais no nível subsimbólico, enquanto o representacionalismo, na sua forma mais radical, pressupõe que a capacidade de manipular símbolos constitui a essência da atividade representacional da mente. Tal manipulação é feita através de algum tipo de "linguagem do pensamento" que possua estrutura sintática e semântica. Problemas relacionados ao processo de interpretação desses símbolos são familiares aos cognitivistas. A exigência, por exemplo, de um intér- prete que decodifica e atribui significado aos símbolos é problemá- tica, pois ela, obviamente, leva a um regresso infinito.

Defensores atuais da vertente representacionalista, tais como Fodor e Pylyshyn (1988), ressaltam dois aspectos centrais das representações mentais:

1) São entidades abstratas que possuem uma "estrutura". Como veremos mais adiante, a noção de estrutura constitui um dos focos de críticas dos representacionalistas tradicionais aos modelos conexionistas das representações.



2) Possuem regras abstratas para a combinação dos seus elementos. Tais regras seriam, supostamente, responsáveis pela estrutura sintática e semântica das representações permitindo, segundo Fodor e Pylyshyn, a realização de inferências. Assim, por exemplo, numa representação onde *p* e *q* fossem consideradas proposições verdadeiras, poderíamos, graças às regras de combinação, inferir que a conjunção (*p* e *q*) seria verdadeira, independentemente das características físicas, locais, de *p* ou de *q*, uma vez que tais regras são supostamente universais.

Por outro lado, os conexionistas argumentam que as redes neurais podem realizar inferências, por exemplo, completando padrões, sem que tais regras abstratas sejam estabelecidas. O sistema comporta-se "como se" possuísse tais regras, contudo nada mais que regularidades físicas estão envolvidas na realização de inferências relativas ao domínio das representações perceptuais.

Apesar de desempenharem algumas funções similares, as representações mentais, tal como caracterizadas pelos conexionistas e representacionalistas, parecem ser incompatíveis. Se admitirmos com os últimos que as representações são entidades abstratas, então, ao estudá-las a partir de elementos físicos ou neurológicos, podemos estar incorrendo num erro categorial, que consistiria em confundir entidades abstratas com elementos físicos. De acordo com a visão representacionalista tradicional, não só estaríamos introduzindo elementos desnecessários, mas, também, inadequados ao estudo das representações mentais.

Os representacionalistas certamente não negam o fato de que necessitamos de um cérebro para instanciar as representações e suas regras de combinação. O ponto controverso que eles defendem é que o estudo dessas representações deve ocorrer num nível puramente lógico. A posição conexionista é, assim, pelo menos problemática em relação a essa perspectiva dos representacionalistas tradicionais.

A posição conexionista também parece ser problemática em relação à segunda vertente filosófica, conhecida por eliminativismo. Este é principalmente representado na Psicologia pela escola de Gibson (1966) e por vários materialistas reducionistas.

O eliminativismo sustenta que as teorias científicas da cognição deveriam dispensar a noção de representação mental, no sentido defendido pelos representacionalistas tradicionais, em favor de uma análise comportamental ou neurológica.



Os eliminativistas mais radicais, preocupados com a cognição, acreditam que o apelo à noção de representação mental espelha o atual estado de ignorância acerca do nosso aparato cognitivo. Eles acreditam que, à medida que a Ciência (em particular a Neuro- fisiologia) se desenvolver vamos conseguir, finalmente, eliminar o termo "representação mental" das teorias científicas.

A certeza de que não existe essa camada de entidades abstratas, intermediária entre o sujeito e o meio ambiente, é expli- citamente enunciada na escola gibsoniana. Gibson insiste que há uma relação direta, não mediada, entre o organismo e o seu meio, sendo a percepção diretamente ligada aos estados do organismo (considerado como um todo) e às invariantes existentes no meio ambiente.

A posição conexionista parece ser, à primeira vista, problemática também em relação ao eliminativismo. O problema surge se consideramos que ela introduz a noção de representação mental que, para os eliminativistas, não é apenas desnecessária, mas, além de tudo, obscurece as tentativas de entendimento da percepção e da ação inteligentes. Entretanto, esse problema se torna menos evidente se admitirmos que a noção conexionista de representação mental difere daquela concebida pelos representacionalistas tradicionais.

Além disso, a ênfase dada pelos eliminativistas na caracterização do nosso aparato cognitivo como um sistema dinâmico, cuja organização depende da interação do sujeito com o seu meio, constitui um elemento comum entre o eliminativismo e o conexio- nismo. Acreditamos que esse elemento permite uma redefinição do conceito tradicional de representação mental (entendida como uma mera abstração), de modo a torná-lo compatível com o paradigma da escola gibsoniana. Em especial, acreditamos que a noção de informação, concebida como uma "entidade" relacional, pode explicitar os possíveis elos existentes entre as hipóteses conexionistas e as gibsonianas sobre a percepção visual.

No que se segue, faremos uma breve introdução às idéias de Gibson concernentes à noção de informação, de modo a poder avaliar algumas de suas possíveis contribuições para o desenvolvimento dos modelos conexionistas das representações mentais.

unesp

CONTRIBUIÇÕES DA CONCEPÇÃO GIBSONIANA DE INFORMAÇÃO PARA O ESTUDO DOS MODELOS CONEXIONISTAS

Pelo menos duas hipóteses defendidas pela escola gibsoniana da percepção direta são relevantes para o presente estudo das redes neurais. Elas dizem respeito às noções interligadas de informação e invariante.

O conceito de informação, proposto por Gibson (1966, 1979), possui um caráter dual. Em primeiro lugar, pode-se caracterizar a informação como um elemento objetivo que existe no meio ambiente em decorrência de suas propriedades relativamente permanentes, que constituem os invariantes estruturais e transformacionais.

Consideremos, por exemplo, um ambiente contendo pedaços de carvão, de papel, de espelho e de mármore. Devido à sua constituição físico-química, cada um desses elementos absorverá e refletirá luz de forma específica e invariante. Ainda que eles se façam em pedaços, ou mudem de posição, poderemos identificá-los como carvão, papel, espelho, etc., desde que suas propriedades físico-químicas sejam preservadas. A figura 2 ilustra alguns aspectos invariantes da informação, tal como concebida por Gibson.

Tais invariantes estruturais, cujas propriedades fundamentais se preservam apesar das possíveis mudanças ao seu redor, correspondem, acreditamos, aos padrões informacionais mencionados anteriormente. Estes incluem também os invariantes transformacionais, que constituem estilos de mudanças que permanecem inalterados em um grande número de situações.

Experimentos realizados com indivíduos numa sala escura, onde marcas luminosas são fixadas às juntas do corpo, ilustram esse tipo de invariante. Enquanto o indivíduo está parado, observadores externos não identificam nada além de luzes dispersas no espaço. Contudo, quando o indivíduo começa a se mover, os observadores percebem, imediatamente, que se trata de um ser humano. Esse reconhecimento se deve, na opinião de Gibson, à existência de invariantes transformacionais presentes nas mudanças.

A figura 3 ilustra um caso particular de invariante transformacional.



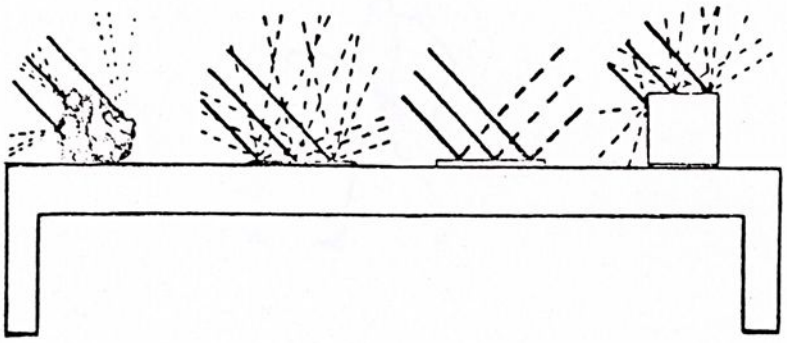


Figura 2 – Uma ilustração esquemática de possíveis invariantes informacionais presentes na informação visual

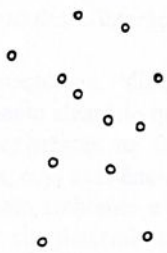
Os invariantes transformacionais e estruturais constituem o núcleo daquilo que Gibson chama de informação sobre algo (*information about x*), e que diz respeito às estruturas invariantes, sejam químicas ou físicas, que existem no meio ambiente independentemente de qualquer intérprete consciente.

A segunda noção de informação, conhecida como informação para alguém (*information for* y), é relacional, e diz respeito àqueles padrões que o organismo está preparado para captar, dado o seu estado evolutivo. Assim, um pássaro pode estar preparado para captar um certo tipo de estrutura, uma cobra um outro tipo, um ser humano ainda um outro, etc.

Essa segunda noção de informação, sendo relacional, não pode ser entendida senão em consideração a um sujeito ativo que, devido ao seu processo evolutivo, está em condições de captar aqueles invariantes que contribuirão para a sua sobrevivência. Esse sujei



a.



b.

Figura 3 – Ilustração de um tipo de invariante transformacional presente na informação disponível sobre o movimento do corpo humano.

to, entretanto, não desempenha qualquer atividade intelectual, interpre- tativa, no estabelecimento desse tipo de relação informacional. A sugestão dos gibsonianos é que, em tal circunstância, um tipo de co-variância é estabelecido entre estados específicos do organismo e eventos específicos no meio ambiente.

A noção de co-variância pode ser entendida aqui como uma relação (de segundo grau) que se estabelece entre conjuntos de

relações (causais ou não) em uma fonte de informação e conjuntos de relações existentes nos estados disposicionais de um receptor ativo. Além disso, sempre que mudanças estruturais ocorrerem na fonte, elas serão, necessariamente, seguidas por mudanças estruturais correspondentes no receptor ativo (uma visão mais detalhada desse tópico é desenvolvida em Gonzales, 1989).

Para concluir, vamos indicar algumas possíveis contribuições que esse conceito dual de informação pode trazer para o desenvolvimento de estruturas auto-organizadas nos modelos conexionistas das representações mentais.

Em primeiro lugar, é preciso ficar claro que a hipótese a ser aqui proposta não fornece, nem pretende fornecer, subsídios para uma teoria reducionista dos estados mentais. As representações mentais a que nos referimos não se reduzem a entidades materiais, ainda que estas constituam uma condição necessária para a existência daquelas.

Como insistimos anteriormente, as representações podem ser entendidas como padrões de conectividade emergentes da interação entre populações de neurônios e padrões informacionais existentes no meio ambiente. A natureza dessa interação não é causal, mas, sim, informacional.

As relações informacionais, diferentemente das relações causais, envolvem um elemento aleatório que opera sobre um conjunto de situações possíveis existentes na fonte e no receptor. Estas várias situações alternativas, cuja existência depende do curso evolucionário percorrido pelo meio ambiente e pelo sistema de representações, introduz um caráter não determinístico à relação informacional. Além disso, as relações informacionais são relações de segundo grau, pois envolvem co-variância entre conjuntos de estados do meio ambiente e conjuntos de estados disposicionais no organis- mo. A ocorrência de co-variância, quando regida por mecanismos adaptativos que possibilitam ao sistema se auto-organizar em função de mudanças no meio ambiente, sugere uma distinção entre representações internas, entendidas como produtos de relações informacionais, e os meros registros realizados por sistemas, tais como os termômetros ou velocímetros, que operam exclusivamente com relações causais simples.

Uma distinção entre os sistemas de representações mentais, do tipo humano, e os meros registros de variações ambientais parece

ser que os primeiros, mas não os últimos, possuem uma estrutura interna que se auto-organiza em função de alterações no meio e nos próprios sistemas.

No que diz respeito às redes neurais artificiais, ainda que essas sejam muito eficientes para trabalhar com situações invariantes, relativas ao reconhecimento de padrões informacionais, elas ainda não lidam satisfatoriamente com situações que envolvem o fluxo contínuo de eventos estruturados no meio ambiente.

Fodor e Pylyshyn (1988), nas suas críticas ao conexionismo, insistem que os modelos sugeridos pelos conexionistas não podem dar conta do aspecto estrutural das representações mentais. Isso porque, dizem eles, tais modelos operam essencialmente com relações causais que atuam nas unidades da rede, cada vez que essa interage com uma particular situação no mundo.

Um ponto relevante da crítica de Fodor e Pylyshyn diz respeito a um antigo problema filosófico que consiste em estabelecer os limites e o alcance das relações causais. O problema surge quando admitimos que as relações causais existentes entre eventos limitam as nossas representações do mundo ao conjunto das experiências *particulares* que vivenciamos através do hábito. A explicação do caráter *geral* das estruturas das representações mentais não parece ser redutível às explicações do tipo puramente causal.

Várias tentativas foram realizadas por filósofos para solucionar esse problema da passagem, ao nível das representações, do particular ao geral. Algumas versões modernas dessas tentativas têm sido sugeridas por filósofos (Dretske, 1983) e cognitivistas (Smolensky, 1986), as quais envolvem a noção de relação informacional. Contudo, tais tentativas ainda se encontram em fase embrionária, uma vez que, conforme mencionamos na abertura deste capítulo, ainda não temos um consenso sobre o conceito de informação e, conseqüentemente, sobre a exata distinção entre *relações causais* e *relações informacionais*.

Ainda que Gibson jamais tenha considerado a possibilidade de explicar qualquer noção relacionada à idéia de representação mental, acreditamos que a sua concepção de informação indica uma possível forma de abordar o problema das estruturas das representações mentais, no sentido conexionista do termo. Em particular, a hipótese da existência de padrões invariantes no meio ambiente sugere que o problema da estrutura das repesentações pode ser pensado em relação

à noção de regularidade, a qual seria determinada pelo registro da freqüência de ocorrência de tais padrões por parte do organismo ativo.

No estudo das redes neurais artificiais, se considerarmos, como Smolensky (1986), que o ambiente de uma rede é composto por padrões invariantes, cuja freqüência de ocorrência é registrada nas conexões existentes entre as suas unidades neurônio-símile, então podemos supor que uma tal rede estruturaria suas representações em função das estruturas preexistentes no meio ambiente. Em outras palavras, a estrutura geral de suas representações refletiria as estruturas dos invariantes no meio.

A interação da rede com os invariantes resultaria, inicialmente, na formação de simples traços/impressões na memória. A recorrência dessa interação com invariantes específicos poderia dar lugar às representações desses invariantes, dependendo da intensidade das conexões entre as unidades que registram suas freqüências de ocorrência. Quanto mais freqüente a ocorrência de um invariante, maior será o valor do peso dessas conexões. As conexões com um valor maximal constituirão as representações mentais, caso ocorra co-variância entre elas e os invariantes no meio.

Em resumo, dadas as pressuposições que: 1) o meio ambiente possui estruturas/padrões invariantes; e 2) a freqüência de ocorrência de tais padrões é registrada nas unidades neurônio-símile da rede; as representações corresponderiam àquelas estruturas estáveis, emergentes da interação das unidades neurônio-símile com os padrões invariantes, sempre que ocorrer entre eles o fenômeno de co-variância, estabelecendo-se entre eles uma relação informacional. Essa última, por sua vez, deve ser regida por regularidades funcionais adaptativas que permitam ao sistema se auto-organizar em função de mudanças no meio e no próprio sistema.

O aspecto estrutural das representações mentais decorreria, assim, dessa relação informacional, que não é puramente causal, uma vez que essa envolve: a) relações entre padrões invariantes; e b) um elemento aleatório no estabelecimento dessas relações, devido ao estado evolucionário do sistema.

Essa proposta de abordagem do problema da estrutura das representações mentais necessita, obviamente, de um maior desenvolvimento, sendo sugerida aqui como uma possível semente para o desenvolvimento de modelos conexionistas, levando-se em consideração o conceito de relação informacional.



Restam ainda inúmeros problemas a serem resolvidos no estudo das representações. Entre eles, o que nos ocupa atualmente diz respeito aos critérios de correspondência entre o conteúdo informacional das representações e os objetos no mundo externo. Problema esse que a tradição filosófica deixou em aberto e que começa agora a ser repensado pelos novos cognitivistas.

REFERÊNCIAS BIBLIOGRÁFICAS

BAKER, L. R. *Saving belief: a critique of physicalism*. Princeton, Princeton University Press, 1987.

DRETSKE, F. *Knowledge and the flow of information*. Oxford, Basil Blackwell, 1981.

_____. "Precis of knowledge and the flow of information". *Behavioral and Brain Sciences*, v. 6, pp. 55-89.

DREYFUS, H. *What computers can't do*. New York, Harper & Row, 1972.

FODOR, J. A. *Representation*. Cambridge (Massachusetts), The MIT Press, Bradford Books, 1981.

FODOR, J. A. e PYLYSHYN, Z. "Connectionism and cognitive architecture: a critical analysis". *Cognition*, v. 28, pp. 3-71.

GIBSON, J. J. *The senses considered as perceptual systems*. Boston, Houghton-Mifflin, 1966.

_____. *The ecological approach to visual perception*. Boston, Houghton-Mifflin, 1979.

GONZALES, M. E. Q. *A cognitive approach to visual perception*. Tese não publicada. Essex University, Colchester, U.K., 1989.

GONZALES, M. E. Q., FRENCH, T. e TREFFNER, P. *A naturalistic approach to mental representation*. Em: GILHOTLY, J. K., KEANE, M.T.C., ERDOS, R. H. (eds.). *Lines of thinking*. Nova York, John Wiley & Sons, 1990.

HEBB, D. *Organization of behavior*. New York, John Wiley, 1949.

KÜPPERS, B. *Information and the origin of life*. Cambridge (Massachusetts), The MIT Press, 1990.

RUMELHART, D. E. e MACCLELLAND, J. L. *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, The MIT Press, 1986, 2 v.

SAYRE, K. "Intentionality and information processing: an alternative model for Cognitive Science". *Behavioral and Brain Sciences*, v. 9, 1986, pp. 121-166.

SCHEERER, E. Towards a history of Cognitive Science. *Cognitive Science*, v. 115, pp. 7-19.

SHANNON, C. e WEAVER, W. *The mathematical theory of communication*. Illinois, The University of Illinois Press.

STONIER, T. *Information and the internal structure of the universe*. London, Springer-Verlag, 1990.

Auto-Organização, org. por M. Debrun, M.E.Q. Gonzales & O. Pessoa Jr., *Coleção CLE 18*, Campinas, 1996, pp. 273-297.

10

Ação, Causalidade e Ruído nas Redes Neurais Auto-Organizadas

MARIA EUNICE QUILICI GONZALES

Depto. de Filosofia, Faculdade de Filosofia e Ciências, UNESP, Marília, SP 17525-900
gonzales@rcb000.uesp.ansp.br

"Eu considero o pensamento tão pouco incompatível com a matéria organizada que, tal como a eletricidade, a faculdade motora, a impenetrabilidade ou a extensão, etc., ele parece-me mais constituir uma propriedade sua" (LA METTRIE, [1747] 1982, p. 96).

1. Introdução

Em sua obra *O Homem Máquina* (1747), JULIEN DE LA METTRIE foi um dos primeiros a iniciar um ousado debate acerca do enfoque mecanicista da estrutura do pensamento humano, o qual veio contribuir para os recentes estudos desenvolvidos pela vertente materialista da moderna Ciência Cognitiva.

Negando, radicalmente, o dualismo estabelecido por DESCARTES (1641) entre mente e corpo, La Mettrie defende a hipótese de que o pensamento bem como o corpo humano devem ser entendidos como manifestações da matéria organizada. O estudo da *dinâmica da matéria* era considerado por ele a chave para o entendimento dos estados mentais humanos. Apesar de admitir que a passagem do estado da matéria inerte para aquele da matéria animada estava além do entendimento humano, La

Mettrie insistia que a matéria organizada está dotada de um *princípio motor*, o único que a diferencia da matéria desorganizada.

Para La Mettrie, os homens e os demais animais diferenciam-se uns dos outros apenas pela variação, quanto ao grau de complexidade e organização, das partes que constituem a maravilhosa máquina do corpo.

Na visão *monista* de La Mettrie, os organismos são máquinas vivas dotadas de auto-locomoção. O homem é uma máquina onde o princípio de movimento desenvolveu, por excelência, a capacidade de aprendizagem e auto-organização. Através do cérebro, peça central de sua maquinaria, ele pode apreender e estocar informação do meio ambiente, ajustando e aprimorando a sua própria organização.

O esboço de teoria iniciado por La Mettrie no século XVIII vem se desenvolvendo desde então, tomando uma nova característica no nosso século, através do emprego de computadores que são utilizados, na Ciência Cognitiva, para estudar o comportamento inteligente.

No presente trabalho, introduzimos uma vertente da Ciência Cognitiva conhecida como *Neoconexionismo*, na qual máquinas, supostamente com capacidade de *aprendizagem* e *auto-organização*, são utilizadas como modelos explicativos da atividade inteligente. O nosso objetivo é investigar em que medida o conceito de "auto-organização" se aplica a tais máquinas. Argumentamos que apenas algumas propriedades da auto-organização podem ser a elas atribuídas, posto que ainda lhes falta o princípio motor acima mencionado, o que as torna pouco "sólidas" enquanto modelos de explicação da atividade inteligente.

2. Inteligência Artificial e Conexionismo

A história da Ciência Cognitiva tem mostrado várias tendências no estudo da cognição. Contudo, duas delas se destacam, as quais são conhecidas como Inteligência Artificial e Conexionismo. Ambas estudam o comportamento inteligente através do emprego de modelos computacionais que, quando devidamente organizados, simulam os estados mentais supostamente subjacentes ao comportamento inteligente. A estratégia de simulação dos estados mentais através dos modelos computacionais é



utilizada, então, para o estabelecimento de hipóteses e teorias acerca do funcionamento e estrutura destes estados mentais. Ainda que estas duas vertentes da ciência cognitiva se preocupem com questões relativas à *organização* e *estrutura* dos estados mentais, e ambas estudem estes estados colocando "entre parênteses" os aspectos emocionais e sociais que possivelmente lhes deram origem, elas diferem quanto à forma de abordagem ao seu objeto de estudo.

Assim, a *Inteligência Artificial*, ao conceber os estados mentais como abstrações, irá desenvolver teorias sobre os mecanismos lógicos e as *regras abstratas* de composição dos sistemas cognitivo-informacionais, enquanto que o Conexionismo irá colocar em evidência os *aspectos físico-estruturais* responsáveis pela sua organização. Na Inteligência Artificial, as pesquisas iniciadas por TURING (1950) foram desenvolvidas, principalmente nas décadas de 70-80, por cientistas como NEWELL & SIMON (1972), através da elaboração de modelos computacionais para a simulação de aspectos racionais do sistema cognitivo, que enfatizam os métodos de resolução de problemas e as regras abstratas de demonstração de teoremas e manipulação de símbolos.

Já o *Conexionismo*, também conhecido como Sistemas de Processamento de Informação em Paralelo (*PDP*, de "Parallel Distributed Processing") ou *Redes Neurais* (devido à sua inspiração na neurofisiologia), desenvolve modelos para a simulação, não tanto dos aspectos racionais dos sistemas cognitivos, mas, sim, dos seus aspectos dinâmico-adaptativos, ligados principalmente à atividade perceptual de reconhecimento de padrões.

Na sua origem, esta segunda vertente da ciência cognitiva centralizou suas linhas de pesquisa nos aspectos racionais da cognição, em especial com os trabalhos sobre máquinas lógicas de MCCULLOCH & PITTS (1943). Contudo, os *neoconexionistas*, seguindo as trilhas de WIENER (1961), se preocupam atualmente em explicar a *dinâmica* dos seres que se *auto-organizam* em função dos seus estados internos e dos eventos que ocorrem no meio ambiente. O estudo das propriedades de auto-organização de alguns modelos conexionistas constitui o principal objeto da presente investigação.

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

Quanto aos modelos desenvolvidos na Inteligência Artificial, acreditamos que eles não constituem genuínos processos de auto-organização (que discutiremos a seguir), porque tais modelos envolvem, em geral, sistemas fechados, no sentido que não existe uma interação entre a máquina (que lida apenas com símbolos) e o meio ambiente. Além disso, tais modelos, estruturados na forma de programas computacionais, não possuem autonomia, uma vez que eles são pré-elaborados pelo programador, sem qualquer possibilidade de alteração significativa das suas regras de funcionamento.

3. Redes Neurais Artificiais

No que se segue, apresentaremos brevemente algumas das principais propriedades dos modelos conexionistas que simulam a atividade de formação de padrões informacionais, também conhecidos como *representações mentais*, para, então, discutirmos as suas possíveis características auto-organizadoras.

Nos modelos conexionistas, a formação e o aperfeiçoamento de padrões informacionais podem ser caracterizados através das conexões que se estabelecem entre conjuntos de *unidades neurônio-símile* (*UNS's*), que compõem a arquitetura de uma rede neural artificial. As UNS's recebem e transmitem estímulos através de suas conexões, conhecidas como "sinapses". O processo de transmissão e recepção de estímulos em uma UNS pode variar de acordo com os diferentes modelos de redes neurais. Contudo, de modo geral, ele ocorrerá de acordo com as três etapas resumidas em CAUDILL & BUTLER (1992, pp. 5-8).

Na *primeira etapa*, os estímulos que chegam a uma UNS são somados, obtendo-se o valor total, ou "net input", que é, normalmente, calculado através da função

$$I_i = \sum_{j=1}^{n} W_{ij} X_j \ , \qquad (1)$$

onde I_i é o valor da soma de todos os sinais recebidos pela unidade i; X_j é o valor do sinal proveniente da unidade j que chega até a unidade i; finalmente, W_{ij} designa o peso ou intensidade da conexão existente entre i e j.

Conforme ressaltam Caudill & Butler, uma vez computado o valor de I_i, perde-se a informação sobre a intensidade dos sinais individuais recebidos por i, pois um sinal forte chegando a uma conexão fraca pode ter menor impacto do que um sinal fraco que chega a uma conexão bastante fortalecida. O que é relevante em I_i é a informação sobre o *comportamento global das unidade* e não de um único sinal que chega a ele.

A *segunda etapa* do processo de transmissão de estímulos consiste em converter o "net input" em um nível de ativação da UNS (o que idealmente corresponderia ao nível de excitação de um neurônio). Esta transformação é geralmente descrita nas redes neurais, por conveniência, através da função sigmóide:

$$f(I) = \frac{1}{1 + e^{-I}} \quad , \tag{2}$$

cujos valores de mínimo e máximo são respectivamente 0 e 1. Este procedimento de conversão do valor total dos inputs em um nível de ativação é comum para todas as UNS's, que terão, assim, valores limites de ativação a elas associados, embora esses possam ser alterados em decorrência da dinâmica de outras funções de ativação que eventualmente estejam incorporadas na rede específica.

Finalmente, a *terceira etapa* do processo de transmissão de estímulos consiste em converter o nível de ativação das UNS's para um sinal de saída destas UNS's. Novamente, essa etapa pode variar de acordo com os diferentes modelos, mas, conforme sugerem Caudill & Butler, ela é, geralmente, estabelecida de acordo com a seguinte expressão:

$$Y_i \begin{cases} = f(I), & \text{se } f(I) > T \\ = 0, & \text{caso contrário,} \end{cases} \tag{3}$$

onde T é um valor limite ou "*threshold*", fixado de acordo com os objetivos a que a rede se destina, e Y_i é o valor do sinal de saída da unidade i. Assim, esta expressão nos diz que o valor de saída ("*output*") de uma UNS será igual ao seu valor de ativação sempre que ele exceder ao valor limite T, e será nulo quando este valor for abaixo do valor de T.

Com estas três etapas, os sinais internos e aqueles provenientes do meio ambiente são processados nas UNS's, que constituirão os átomos dos modelos conexionistas para a simulação de determinados estados mentais. Além disso, uma hipótese fundamental do conexionismo é que o estudo dos estados mentais, determinantes do comportamento inteligente, deve ser feito a partir da análise dos mecanismos, princípios ou processos que conectam, modificam e estruturam *grupos de UNS's*. Estes formam redes com camadas interligadas: uma camada de unidades de entrada, ou "*input layer*", outra de resposta, ou "*output layer*", e zero ou mais camadas intermediárias, ou "*hidden layers*". Essas camadas se comunicam entre si produzindo um *efeito global*, que é, na maioria dos casos, distinto daquele presente em cada uma das unidades consideradas separadamente. É precisamente este nível global de interação entre as UNS's que interessa ao presente estudo dos princípios de auto-organização, pois aqui, supostamente, ocorre o surgimento de *padrões informacionais*, através de mecanismos de *treinamento* e *aprendizagem*.

Quando uma rede interage com padrões de input, que caracterizam os elementos básicos constituintes dos objetos ou estados do meio ambiente, grande parte das UNS's pertencentes à camada de input recebem estímulos deste padrão. Tais estímulos serão transmitidos às unidades da(s) camada(s) intermediária(s) através das conexões existentes entre elas e as UNS's da camada de entrada. Dependendo do peso inicial destas conexões, que podem variar *aleatoriamente*, alguns estímulos serão ampliados e outros inibidos (os pesos funcionam como canais ampliadores ou redutores de estímulos). Conseqüentemente, as UNS's da(s) camada(s) intermediária(s) produzirão uma grande variedade de respostas aos estímulos inicialmente apresentados à rede. Da mesma forma, o peso das conexões, existentes entre as UNS's das camadas intermediárias e de output, exercerá influência variada na quantidade de ativação/inibição das

unesp

UNS's de respostas. Estas, quando convertidas nos sinais de saída da rede, fornecerão uma resposta *global* da rede ao padrão de input original.

Nos casos em que as variáveis essenciais do padrão de entrada são preservadas no padrão de saída, considera-se que a rede representou adequadamente a informação original, *associando* o padrão de entrada a um padrão correspondente de saída. Entretanto, a grande variedade dos pesos das conexões das UNS's, que modulam a intensidade dos estímulos que chegam até elas, praticamente impede um registro imediato, fidedigno, de informação. Para que isto ocorra, a rede terá que ser *treinada*, através de algum mecanismo de aprendizagem, de modo a "aprender" a encontrar o correto peso das conexões "sinápticas" que lhe permitirá representar, sem distorções, a informação existente no meio ambiente. Neste sentido, diz-se que:

> [...] redes neurais aprendem a resolver um problema; elas não são programadas para resolvê-lo (CAUDILL & BUTLER, 1992, p. 7).

Em geral, os mecanismos de aprendizagem constituem uma variante da *lei de Hebb*, segundo a qual estímulos sensoriais deixam um traço na memória, cuja intensidade depende da intensidade das conexões sinápticas (HEBB, 1949). De acordo com a versão sugerida por RUMELHART & MCCLELLAND (1986) da lei de Hebb, pode-se dizer que sempre que uma unidade neuronal estimular, repetidamente, outra unidade, e ambas estiverem altamente ativas, então a conexão sináptica que se estabelece entre elas será fortalecida. Através desta lei pode-se modificar os pesos das conexões existentes fazendo com que determinadas unidades estimulem (ou inibam) umas as outras através de treinamento.

4. Treinamento Não Supervisionado de Redes

Existem três tipos básicos de treinamento nos modelos de redes neurais artificiais: (a) treinamento supervisionado; (b) treinamento por reforço; e (c) treinamento não supervisionado ou por *auto-organização*.

No *treinamento supervisionado* (caso a), um padrão de entrada juntamente com um padrão de saída são fornecidos à rede. Espera-se que,

através de algum mecanismo de aprendizagem, a rede encontre os pesos adequados das conexões que lhe possibilite *associar* o padrão de entrada ao padrão de saída estabelecido por um supervisor.

No *treinamento por reforço* (caso b), não se estabelece um padrão específico de saída mas, dado um padrão de entrada, espera-se que a rede aprenda a identificá-lo, associando-o a um outro padrão, através de sinais de "reforço" fornecidos por um supervisor. Tais sinais servirão para orientar o processo de ajuste adequado dos pesos das UNS's.

Dependendo dos diferentes mecanismos de ajuste dos pesos, ou "regras" de aprendizagem, diferentes modelos de redes neurais serão obtidos, seja no caso dos modelos com supervisão direta ou por reforço. Devido à dependência, direta ou indireta, de um supervisor que controla as etapas do processo de estruturação e desenvolvimento destes modelos, não consideramos que eles se auto-organizam.

Finalmente, *no treinamento não supervisionado ou por auto-organização* (caso c), nenhum padrão específico de entrada ou de saída é fornecido de antemão à rede, que terá que classificar, criar ou detectar um padrão, através do reconhecimento de regularidades na informação disponível no meio ambiente, sem a ajuda de um supervisor. Um dos mecanismos mais simples de ajuste de pesos utilizado no treinamento não supervisionado, ou por auto-organização, é o conhecido princípio de *aprendizagem por competição* (ver KOHONEN, 1984, para inúmeras ilustrações deste mecanismo). Segundo este princípio, uma UNS, quando ativada, pode influenciar o comportamento de suas UNS's vizinhas, inibindo ou ativando-as, desencadeando assim uma reação global que determinará o estado final da rede.

Para entender este mecanismo, consideremos a Figura 1, que ilustra o estado inicial de uma rede que possui n camadas de UNS's, e cujas conexões "sinápticas" podem ser descritas através de uma matriz. Conforme ressaltamos anteriormente, os valores iniciais dos pesos das conexões entre as UNS's podem ser arbitrariamente estabelecidos, ou determinados segundo os objetivos a que a rede se destina. Tais valores podem ser ordenados na forma de vetores que formarão a *matriz peso*. Esta servirá como base para as operações que se realizarão entre a rede e os vetores de input.



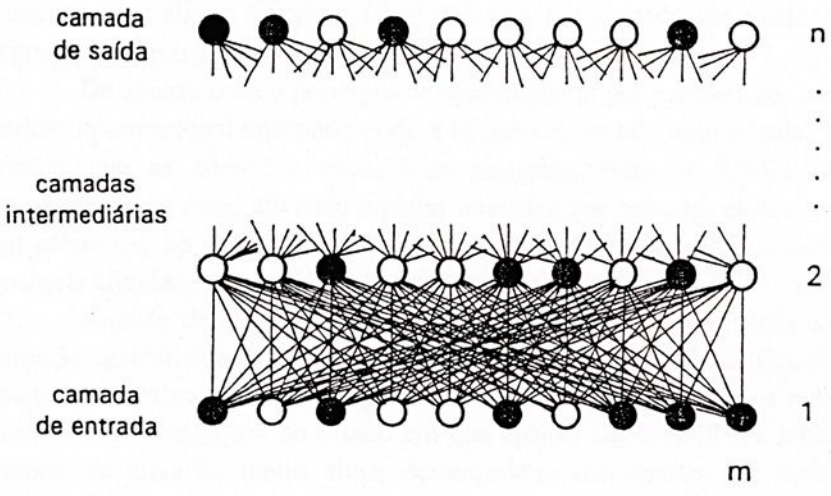


Figura 1. Ilustração do estado inicial de uma rede neural artificial.

Seja $U = \{(a_1,a_2,...,a_m), (b_1,b_2,...,b_m), ... (x_1, x_2,...,x_m)\}$ o conjunto de vetores de input que descrevem, de maneira simplificada, as características de objetos ou de possíveis propriedades invariantes do meio ambiente, e A^* a matriz peso, que descreve o estado inicial de uma rede neural, caracterizado pelo peso das conexões iniciais de suas unidades:

$$A^* = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad (4)$$

As UNS's da rede, que inicialmente constituem apenas potencialidades, ao interagirem com o conjunto de estados descritos pelos vetores de U podem dar lugar a uma *atualidade*, que se apresenta na

forma de um padrão informacional. Este padrão – resultante do efeito global da interação entre as UNS's e os invariantes do meio ambiente (descritos por algum vetor de U) – terá que ser ajustado, de modo a representar adequadamente alguma estrutura invariante.

De acordo com o princípio de aprendizagem por competição, um padrão informacional adequado poderá se formar, em tal circunstância, à medida que as conexões excitatórias existentes entre as UNS's se propaguem pela rede, ativando aquelas unidades que possuem elementos em comum e, ao mesmo tempo, inibindo as demais que não possuem qualquer afinidade com as UNS's inicialmente ativadas.

Através de vários ciclos de iterações, as unidades ativadas se tornarão gradativamente mais ativadas e as UNS's inibidas ficarão gradativamente menos ativas. Após um certo número de iterações, a rede tenderá a se *estabilizar* no estado em que apenas algumas UNS's terão valores de ativação muito altos, destacando-se das demais que terão valores de ativação muito baixos. Quando este estado de estabilidade ocorre, diz-se que uma representação de um invariante, descrito através de algum vetor de U, estabelece-se na *estrutura emergente* das UNS's ativadas, e que a rede adquiriu uma forma de padrão informacional.

Conforme mencionado, o mecanismo de aprendizagem por competição constitui um dos mais simples entre aqueles disponíveis nos modelos conexionistas que envolvem *aprendizagem não supervisionada*. Uma rede atuando segundo este mecanismo modificará o peso de suas conexões com base apenas no seu estado inicial e nas características das regularidades existentes no meio ambiente, dispensando a atuação de um supervisor. Devido à este relativo grau de autonomia, diz-se, em geral, que a rede se *auto-organiza*.

Argumentaremos a seguir que, apesar do emprego comum do termo *auto-organização* nas redes neurais do terceiro tipo acima referido, apenas algumas propriedades da auto-organização podem ser atribuídas às redes neurais, tal como elas se apresentam no momento. Julgamos que a razão principal desta limitação deve-se ao fato que ainda lhes falta o princípio motor, referido por La Mettrie como o elemento fundamental para a caracterização da matéria viva organizada, em contraste com a matéria desorganizada.

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

5. Auto-Organização nas Redes Neurais

O termo *auto-organização* tem sido amplamente empregado na literatura associada aos modelos conexionistas de redes neurais. Isto porque os conexionistas, num movimento de reação à metodologia de pesquisa da Inteligência Artificial tradicional – que emprega essencialmente regras abstratas ou programas na simulação do comportamento inteligente –, desenvolveram modelos de redes neurais cujo funcionamento depende, em princípio, apenas de leis físicas. Embora alguns destes modelos se comportem como se estivessem seguindo regras abstratas pré-estabelecidas, nenhuma regra é explicitamente incorporada a eles. Os mecanismos de aprendizagem, do tipo "regra de Hebb", são considerados *descrições* de processos físicos que supostamente ocorrem no organismo, e não convenções ou conjuntos de instruções que determinariam o curso de ação deste organismo. Conforme vimos na seção anterior, espera-se que as redes possam ser *treinadas* através destes mecanismos, de modo a aprender a encontrar o peso correto das conexões "sinápticas", que lhes permitiria resolver uma determinada tarefa, independente da existência de um programa que especifique cada um dos passos para resolver a tarefa.

Além disso, nas redes neurais com treinamento não supervisionado, nenhum centro organizador se impõe na formação ou classificação dos padrões de conectividade que correspondem as suas representações. Idealmente, tais representações resultam da dinâmica que se estabelece na interação entre a totalidade das UNS's e o meio ambiente.

Assim, neste contexto, o conceito de auto-organização é normalmente entendido em termos de *ausências*, seja de regras, ou programas que direcionem cada etapa do processo de organização das redes, seja de um centro organizador que se lhes impõem externamente. Estas características constituem, entretanto, apenas algumas das propriedades da auto-organização entendida em um contexto mais amplo, tais como nas acepções de ASHBY (1962), ATLAN ([1979] 1992), DEBRUN (1996), entre outros.

De acordo com a concepção de ASHBY (1962) e DEBRUN (1996), por exemplo, o termo *auto-organização* pode ser entendido a partir de duas perpectivas. A primeira, também conhecida como *auto-organização*

primária, se refere a situações em que elementos quaisquer independentes interagem, de tal forma a criar uma relação de dependência mútua entre eles, o que terá como resultado o estabelecimento, criação ou emergência de alguma estrutura, forma ou sistema:

> Tal sistema é "auto-organizante" [*self-organizing*] no sentido de que ele passa de "partes separadas" para "partes unidas" (ASHBY, 1962, p. 266).

Um exemplo de auto-organização primária pode ser concebido em uma rede neural cujas UNS's encontram-se, inicialmente conectadas de forma aleatória, sem qualquer relação de dependência existente entre elas. À medida que a rede interage com invariantes no meio ambiente, valores de ativação serão estabelecidos nas suas unidades de entrada, e propagados às unidades vizinhas. Com o treino não supervisionado, a rede eventualmente convergirá para um estado tal em que os pesos das conexões entre as suas UNS's se manterão estáveis, dando lugar a um padrão de conectividades, que será entendido pelos conexionistas como um modelo de *representação mental* do objeto ou invariante que lhe deu origem.

No exemplo acima, podemos considerar que a rede se auto-organizou primariamente, uma vez que uma estrutura foi criada a partir de elementos inicialmente independentes, sem a intervenção de um centro organizador. O padrão de conectividades em questão resultou da dinâmica do meio ambiente, com suas regularidades e aleatoriedades, em conjunto com os mecanismos de aprendizagem disponíveis na rede.

O segundo sentido de auto-organização referido por ASHBY (1962, p. 267) e DEBRUN (1996), conhecido como *auto-organização secundária*, pressupõe a existência de um sistema primariamente organizado adicionando-lhe mecanismos de evolução. Estes incluem "*feedback*" (retroalimentação) e aprendizagem através da correção de erros, bem como critérios de relevância para distinguir as situações boas/adequadas daquelas más/inadequadas para o desenvolvimento do sistema.

Uma característica fundamental da auto-organização secundária, na concepção de Ashby, consiste na capacidade dos sistemas de evoluir de

unesp

um estado de organização para outro mais aprimorado ou, o que é ainda mais importante, melhor adequado à sua sobrevivência. Neste sentido, apenas os organismos vivos poderiam se auto-organizar secundariamente, pois só eles possuem a capacidade de desenvolver *critérios de relevância* que lhes permita distinguir o adequado do inadequado, o bom do mal, ainda que num sentido estritamente relativo à sua própria estrutura. Ashby acredita que nenhuma máquina pode, *em princípio*, se auto-organizar secundariamente, posto que faltaria a elas esta capacidade básica, presente nos sistemas vivos, de ajustar os seus estados de equilíbrio para a preservação da vida.

Admitindo com Ashby que as máquinas não dispõem da característica fundamental da auto-organização secundária, que se manifesta através dos *critérios de relevância* na seleção de informação para a preservação da vida, resta saber se elas poderiam apresentar alguma outra propriedade que as caracterizasse como um sistema de auto-organização secundária em um sentido fraco. Em particular, se admitirmos a possibilidade de *aprendizagem* nas redes neurais, então, pelo menos em princípio, poderíamos admitir também uma das propriedades da auto-organização secundária, associada à habilidade de desenvolvimento e aprimoração dos seus estados representacionais. Mas em que sentido se diz que uma rede neural artificial aprimora os seus estados?

Vimos nas seções anteriores que as redes neurais artificiais incorporam mecanismos de *aprendizagem*, do tipo *competição* ou variantes da "regra de Hebb", os quais atuam na dinâmica do ajuste de pesos das conexões entre as suas UNS's. Com estes mecanismos, e dependendo da dinâmica do meio ambiente, as redes podem se corrigir na realização de tarefas, em especial naquelas relativas à recognição de padrões, aprimorando, assim, os seus estados internos. Contudo, resta saber se a aprendizagem localizada, presente nas redes neurais, constitui condição suficiente para sua possível evolução, uma vez que esta última requer uma história de longa duração.

Duas são as situações em que podemos considerar a possibilidade de "evolução" ou aprimoramento em uma rede neural artificial: (i) as que requerem *correção através de "feedback"* e (ii) as que empregam o *princípio da ordem a partir do ruído*.



Nas primeiras, (i), os padrões de conectividades, emergentes da dinâmica das UNS's, podem apresentar inicialmente um conteúdo informacional impreciso em relação aos invariantes do meio ambiente que eles representam. Neste caso, considera-se que tais padrões representam erradamente os invariantes que lhes deram origem. A intervenção do acaso no estabelecimento inicial de valores aleatórios aos pesos das conexões das UNS's leva a rede, em geral, a responder aos invariantes de forma inadequada. Após exposição reiterada aos mesmos invariantes por um tempo prolongado, e com a correção dos erros, via ajuste de pesos das conexões, os padrões de conectividades estáveis passarão eventualmente a representar, de forma adequada, os invariantes especificados pelo meio ambiente. Como já dissemos, esta rede – que inicialmente fornecia respostas inapropriadas aos invariantes e, posteriormente, passa a identificá-los corretamente – *aprendeu* a corrigir e aprimorar os seus estados representacionais.

Nesta perspectiva elementar, de *ajuste e correção* dos padrões de conectividades para a classificação ou recognição de padrões de entrada, poderíamos conceber que as redes neurais compartilham uma das propriedades da auto-organização secundária, no sentido fraco. Aparentemente, apenas o *hábito* – estabelecido com base na repetição de *relações causais* atuando no meio ambiente e nas UNS's artificiais, em conjunto com mecanismos elementares de "feedback" –, seria necessário para a realização e aprimoramento de tarefas de reconhecimento e de repetição de padrões realizadas pelas redes neurais artificiais.

A segunda situação, (ii), em que se admite a possibilidade de evolução em uma rede neural artificial envolve o que VON FOERSTER (1960) denominou *princípio de ordem a partir do ruído*. De acordo com este princípio, sistemas abertos submetidos a vários tipos de perturbações, incluindo agressões do meio ambiente, podem vir a desenvolver mecanismos de adaptação, na forma de novas organizações, que os auxiliariam no processo de adaptação. Assim, a presença do ruído em certos sistemas, ao invés de ser caracterizada como algo inerentemente mau, recebe aqui um papel construtivo.

Este aspecto positivo das perturbações foi amplamente investigado por ASHBY ([1956] 1970, pp. 238-251) e ATLAN (1992, pp. 38-45), entre

outros, para quem o ruído, quando adequadamente dosado, pode colaborar para o enriquecimento dos sistemas abertos, à medida em que ele introduz *variedade* às suas formas fixas e bem estabelecidas de organização. Eles consideram que as perturbações, quando não carregam em si um poder devastador absoluto, possibilitam maior flexibilidade a tais sistemas.

No contexto dos seres vivos, o ruído – entendido aqui como uma forma de interferência, algumas vezes destrutiva, na atividade normal dos estados funcionais do organismo – parece desempenhar um papel paradoxal. Sua presença nos organismos frágeis pode ter consequências fatais, contribuindo para a desorientação ou até a extinção dos mesmos. Contudo, nos organismos menos frágeis, seu efeito desestabilizador é em geral temporário e muitas vezes benéfico. As formas fixas de comportamento (algumas das quais poderiam torná-los inadaptáveis à um meio ambiente em constante mudança) são abaladas, mas tal abalo pode dar lugar a novos padrões mais aperfeiçoados de comportamento, tornando os organismos mais fortalecidos e aptos para as mudanças presentes num meio ambiente dinâmico.

6. A Teoria da Harmonia

Admitindo esta característica auto-organizadora do ruído nos sistemas vivos, restaria investigar sua plausibilidade nas redes neurais artificiais. Em que categoria iremos classificá-las – entre os sistemas mais frágeis ou menos frágeis?

Ainda que a última opção pareça ser a resposta imediata, alguns modelos conexionistas, do tipo "*Boltzmann machine*" (máquina de Boltzmann, HINTON & SEJNOWSKI, 1986), sugerem que uma das grandes vantagens das redes neurais artificiais, em relação aos modelos tradicionais da inteligência artificial, é a sua capacidade de lidar com o ruído. Assim, por exemplo, SMOLENSKY (1986), na sua *Teoria da Harmonia* ("*Harmony Theory*"), utiliza-se da noção de "ruído térmico" como um recurso para alterar os estados fixos, indesejáveis, do seu modelo conexionista de representações mentais. Neste modelo, onde se faz uma fusão entre a Teoria Matemática da Informação, a Termodinâmica e a

Mecânica Estatística, o peso das conexões entre as UNS's codifica, como nas redes neurais artificiais, a freqüência com que os invariantes ocorrem no meio ambiente: quanto mais freqüente for o registro de uma regularidade no meio ambiente, mais alto será o valor do peso das conexões das UNS's que a detectam.

Na Teoria da Harmonia, os pontos de conexão entre as UNS's que registram regularidades no meio ambiente são denominados *átomos de conhecimento*. Conjuntos de átomos de conhecimento, que possuem um certo grau de coerência (medido pela similaridade dos caracteres registrados), dão lugar às estruturas estáveis conhecidas como *esquemas*. Uma vez formados os esquemas, a rede pode movimentar-se de um esquema para o outro em função (a) das tarefas de reconhecimento de padrões que lhe forem atribuídas, (b) de mudanças no meio ambiente e (c) do grau de harmonia, ou coerência, nas suas representações. Quando novos estímulos chegam à rede ela *tende a* detectar aqueles conjuntos invariantes de caracteres que concordam ou se harmonizam com os seus esquemas aprendidos. Estímulos não compatíveis com tais esquemas introduzem ruído no sistema, aumentando o seu grau de entropia e, conseqüentemente, diminuem o valor correspondente à sua medida de harmonia. Esta medida é dada por uma função, $H(r,a)$, onde a expressa algum *átomo de conhecimento* e r algum estado, aspecto ou padrão informacional dos "inputs" ambientais.

O princípio estruturador fundamental do modelo de Smolensky é aquele denominado *princípio de harmonia*. Este estabelece que, na criação de esquemas, a rede adote, como uma tendência geral, os caminhos de máxima harmonia, onde os valores de $H(r,a)$ são máximos. Os padrões de conectividades da UNS's, que caracterizam os esquemas segundo este princípio, se comportam como se estivessem minimizando uma quantidade que desempenha o papel de *energia física*.[1]

---

[1] Esta analogia tem inspiração em uma versão do *teorema H* de Boltzmann, onde a função H traduz os efeitos das colisões que em cada instante modificam as posições e velocidades das partículas de um sistema de forma a alcançar, ao longo do tempo, um valor mínimo para H. Aí a distribuição da velocidade e posição das partículas não se modifica, tornando-se assim estável.



Uma vez criados, os esquemas condicionam as probabilidades de escolha, por parte da rede, dos dados disponíveis no meio ambiente, em função do *princípio de harmonia* ou de *maior coerência entre esquemas e estímulos do meio*. A partir daí, a rede condicionará, de maneira crescente, o seu próprio desenvolvimento, no que diz respeito à seleção dos possíveis estímulos que ativarão as suas UNS's. Esta é mais uma propriedade do modelo conexionista que parece aproximá-lo dos sistemas auto-organizados, no que se refere ao *processo de endogenização*.

Conforme foi ressaltado por Debrun, o processo auto-organizado também se caracteriza por uma *endogenização* crescente. Entre outras propriedades da auto-organização secundária, destaca-se aquela em que:

> O processo é cada vez mais "responsável" por seu próprio desenrolar. O que significa [...] que a participação do acaso, que pode ser importante no começo, é progressivamente "absorvida". (DEBRUN, 1996, p. 33).

No caso do modelo de Smolensky, à medida que o peso inicial das conexões é estabelecido aleatoriamente, o *acaso* desempenha um papel relevante no processo inicial de organização da rede, pois, dependendo do valor que se atribui às conexões iniciais entre suas UNS's, diferentes resultados parciais serão obtidos no reconhecimento de um padrão específico. Contudo, o acaso terá um papel cada vez menor uma vez que os esquemas vão se formando e condicionando o futuro desenvolvimento da rede, até que ela chegue a um estado de relativa estabilidade. Com esquemas estáveis, a rede passará a reconhecer os invariantes do meio que lhe deram origem, ignorando a novidade, e algumas vezes, representando-as erroneamente. É nesta situação que, o ruído pode vir a introduzir flexibilidade no sistema, desorganizando parcial ou totalmente os esquemas existentes.

Quando, através de meios específicos, o ruído é introduzido na rede, cria-se uma situação de instabilidade temporária, com baixos valores de *H*. Nesta circunstância, a Teoria da Harmonia determina que novos conjuntos de estímulos sejam examinados pelo modelo e que novos esquemas sejam formados. A seguir, reinicia-se o processo de seleção daqueles estímulos que são compatíveis com os novos esquemas,



resultantes dos antigos padrões adicionados pelo ruído. *O princípio da harmonia deverá então novamente entrar em vigor*, mesmo que seja apenas como uma *tendência* diretiva do sistema.

Desta forma, o ruído desempenha um papel criador no modelo esboçado por Smolensky, que sugere uma descrição da dinâmica do processo de formação de representações mentais perceptuais, a qual inclui a detecção de eventos e mudanças no meio ambiente. O *princípio de harmonia*, ou de *coerência*, parece funcionar neste modelo como um mecanismo de auto-organização que, deixando espaço para o ruído, permite à rede se reorganizar, localmente, diante das mudanças e eventuais perturbações do meio ambiente.

A questão que se coloca aqui é se seria adequado ver nesta concepção da dinâmica de organização do modelo de Smolensky um *critério de relevância,* de acordo com o qual a rede poderia maximizar a eficiência de seu comportamento através da seleção da informação disponível no meio ambiente. Em caso afirmativo, poder-se-ia considerar que dado um conjunto de padrões informacionais alternativos, o princípio de harmonia indicaria um critério flexível de seleção dos padrões cuja representação se ajustaria, consistentemente, aos esquemas já aprendidos, preparando a rede para a interação com a diversidade, presente no ambiente na forma de ruído. Deveríamos neste caso concluir que estamos diante de um sistema artificial que, afinal, não é tão frágil como nos parecia à primeira vista?

Apesar da relativa capacidade deste e de outros modelos de redes neurais artificiais de lidar com o ruído, e da inovação apresentada pelo princípio da harmonia, julgamos que eles ainda devem ser classificados entre os sistemas frágeis, no que diz respeito à sua interação com o ruído. Isto porque o ruído, ao introduzir variabilidade e diversidade tanto nas redes como nos padrões informacionais disponíveis, deixa a elas a difícil tarefa de selecionar apropriadamente aqueles elementos que auxiliariam no seu desenvolvimento. A experiência nos mostra que nem sempre o princípio de harmonia ou de coerência se aplica à situações complexas onde o ruído se instala. Aliás, conforme nos lembra ASHBY ([1956] 1970), o próprio *ruído* não é distinguível, de modo intrínseco, de qualquer outra

forma de variedade. Ele ressalta, no contexto da Teoria da Informação, que:

> Apenas quando é dado algum receptor, que estabelecerá qual dos dois é importante para ele, será possível a distinção entre mensagem e ruído (ASHBY, 1970, p. 219).

Num canal, por exemplo, que transmita ao mesmo tempo uma conversa e alguns efeitos sonoros pouco usuais, estes últimos poderão ser classificados como *ruído* para aquele que está interessado em ouvir a conversa. Contudo, para o pesquisador interessado em medir a freqüência dos efeito sonoros pouco usuais, a conversa poderá constituir um *ruído*.

7. O Papel da Ação na Auto-Organização Secundária

Assim, encontramo-nos uma vez mais diante do problema da aquisição de critérios de relevância mais amplos, seja para a seleção de padrões informacionais, seja para a própria distinção entre ruído e outras formas de variedade. Ainda que estejamos longe de visualizar uma solução satisfatória para este problema, parece aceitável a consideração de que a aquisição de tais critérios exige algo além do mero hábito, estabelecido pela freqüente exposição de sistemas com memória e sensibilidade a determinadas regularidades nômicas. A aquisição de tais critérios aparentemente não se reduz à simples repetição de *relações causais*, mas decorre de uma *habilidade inteligente* que se encontra entrelaçada às *ações*. Isto porque processos naturais de seleção podem ser estabelecidos através da locomoção de um agente num meio ambiente repleto de variedades e dificuldades, instaurando-se um princípio regulador básico, segundo o qual são eliminados, ao longo do tempo, aqueles sistemas com capacidade defeituosa de adaptação.

As ações bem sucedidas, eventualmente decorrentes de mecanismos elementares de tentativa e erro, de "feedback" e, naturalmente, de um princípio de harmonia ou coerência, poderiam ser vistas como estabelecendo um primeiro filtro para o fluxo da diversidade de estados do mundo. As marcas deixadas na memória por estas ações atuariam como

"variáveis essenciais" no estabelecimento de critérios de relevância para os sistemas bem adaptados, os quais por sua vez constituiriam os tijolos das habilidades inteligentes.

Se as habilidades inteligentes estão, em grande parte, entrelaçadas às ações bem sucedidas, que por sua vez forneceriam subsídios para o estabelecimento de critérios de relevância na seleção da diversidade informacional, então as redes neurais artificiais, como se encontram atualmente estruturadas, não podem possuir, em princípio, tais critérios. À medida que elas não dispõem, até o momento, de habilidades motoras, não se deve esperar encontrar nelas propriedades auto-organizadoras além daquelas envolvidas na formação de representações para o reconhecimento de padrões informacionais simples. Uma conseqüência de tal limitação pode ser sentida tanto no modelo de Smolensky como nas redes neurais tradicionais, que possuem pouca plasticidade: na presença do fluxo da diversidade ambiental elas não conseguem evitar, a não ser localmente, o aumento da desordem dos seus padrões de conectividades.

Uma característica destas habilidades inteligentes, em contraste com os meros hábitos, é que elas envolvem *regulação* e *controle*, em geral decorrentes de um treinamento *dinâmico* e *crítico*. Este treinamento exige uma constante avaliação acerca da sua adequação em relação aos fins estabelecidos, bem como a capacidade de alteração dos mesmos quando necessário. Seguindo as trilhas de RYLE (1949), poderíamos dizer que tais habilidades constituem *estados disposicionais complexos*, adquiridos através de treinamento, os quais, sob certas condições, dão lugar à ação inteligente.

O estudo das habilidades inteligentes tem resistido ao tratamento oferecido pelos materialistas reducionistas e mecanicistas. Ele tem resistido igualmente às investigações dos dualistas cartesianos, uma vez que, ao atribuir à alma, imortal e exclusivamente humana, esta propriedade, eles tornam problemático o seu entendimento a partir de uma perspectiva científica. Em particular, as habilidades inteligentes, entendidas como manifestações das propriedades da alma humana, não poderiam, em princípio, ser encontradas em redes neurais artificiais, nem seriam objetos de estudos da Física, Biologia ou qualquer ciência natural.

unesp

Em contraste com esta visão dualista, os materialistas reducionistas procuram fornecer explicações causais para as habilidades inteligentes identificando-as a determinados estados cerebrais. Contudo, ainda não se dispõe de uma teoria materialista que distinga satisfatoriamente, por exemplo, hábitos e habilidades inteligentes.

Conforme ressalta RYLE (1949), ambas as tentativas parecem sugerir soluções para um pseudoproblema, que consiste em caracterizar a mente como uma substância (material para os reducionistas e imaterial para os dualistas), e os estados mentais, inteligentes ou não, como instanciações desta substância. Se, diferentemente, entendemos a mente não como uma substância, mas como um *processo*, então concluiremos que sem um estudo da ação realizada por seres inteligentes não teremos como compreender os critérios de relevância que regem o seu processo de auto-organização.

No caso específico das redes neurais artificiais, já se pode notar, por exemplo nos trabalhos de VERSCHURE (1992), entre outros, uma insatisfação crescente no que diz respeito à caracterização dos processos de auto-organização nos modelos conexionistas "estáticos". Dando continuidade às idéias de Ashby e Wiener, Verschure procura introduzir elementos da *Teoria da Ação* em um robô, de forma a caracterizá-lo como um modelo que se auto-organiza secundariamente.

A Figura 2 ilustra duas etapas do processo de auto-organização do sistema elaborado por VERSCHURE *et al.* (1992) na forma de um robô, no qual ocorre a formação de representações de um ambiente onde obstáculos são colocados no seu caminho. A tarefa a ser realizada pelo robô, na Figura 2a, consiste em aprender a desviar-se dos obstáculos utilizando apenas a informação disponível no meio ambiente e leis de aprendizado do tipo "regra de Hebb". Conforme aparece na Figura 2c, após um certo período de treinamento o robô aprende a desviar dos obstáculos, ilustrando, na opinião do autor, um genuíno processo de auto-organização.

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

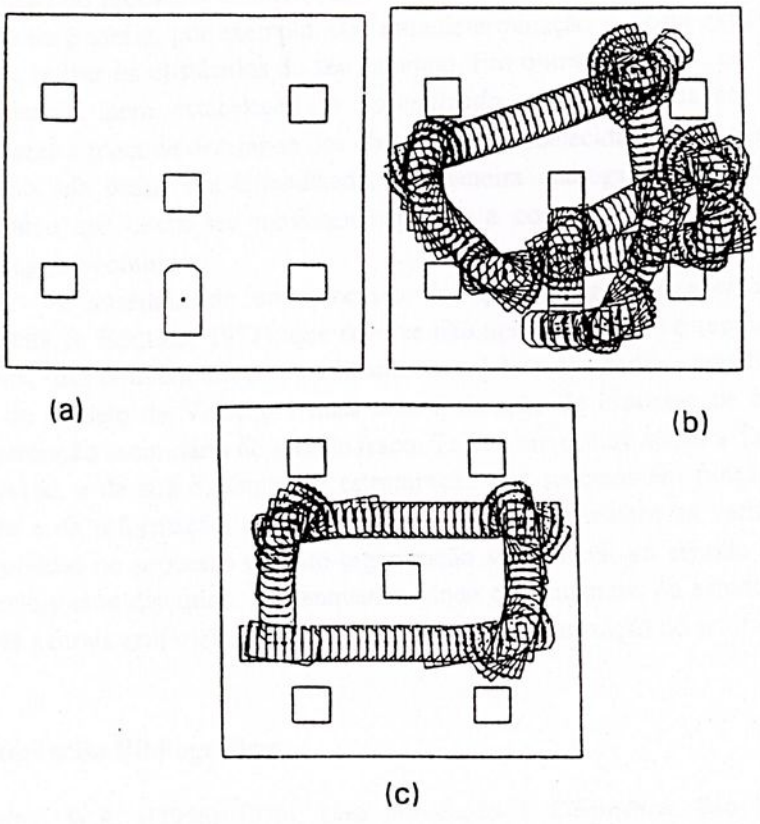


*Figura 2. Auto-organização no robô construído por Verschure para locomoção no meio ambiente. (a) Registro da situação inicial em que se encontra o robô, e dos cinco obstáculos fixos (b) Movimentos iniciais do robô, nos quais ele se choca com os obstáculos e com as paredes. (c) Ilustração do estado final alcançado pelo robô, no qual ele aprendeu a evitar os obstáculos (*VERSCHURE et al., *1992, p. 188).*

Ainda que o presente modelo – para a simulação de estados representacionais envolvidos no processo de auto-organização da percepção/ação –, seja um dos mais sofisticados existentes na Ciência Cognitiva, ele não constitui, para nós, um bom modelo ilustrativo de auto-organização secundária. Isso porque o robô de Verschure *se locomove mas não age* no meio ambiente. Ele mantém uma posição essencialmente *passiva* na

resolução do problema de interação com os objetos do meio; em momento algum ele poderia, por exemplo, sem uma determinação imposta exteriormente, retirar os obstáculos do seu caminho. Em outras palavras, ele não experiencia, nem estabelece um *aprendizado ativo* da situação. Ao satisfazer a meta de desviar-se dos obstáculos, estabelecida por um agente externo, ele cessa seu aprendizado, de maneira análoga a um relógio mecânico que cessa seu movimento quando a corda acaba, sem poder prosseguir evoluindo.

A ausência de um processo de *aprendizagem participativa* (POPPER & ECCLES, 1977), que envolve não apenas controle e regulação crítica, mas também um compartilhamento coletivo dos dados aprendidos, faz do modelo de Verschure mais uma ilustração do processo de auto-organização secundária no sentido fraco. Talvez um estudo sobre a Teoria da Ação, e de sua dinâmica de estruturação dos sistemas em função do ruído e da informação, possa nos ajudar a entender melhor as variáveis envolvidas no processo de auto-organização secundária, no sentido forte anteriormente discutido. Por enquanto, ainda continuamos, no estudo das redes neurais artificiais, com o processo de auto-organização do relojoeiro.

Referências Bibliográficas

ASHBY, W.R. ([1956] 1970), *Uma Introdução à Cibernética*. São Paulo: Perspectiva. (Tradução de *An Introduction to Cybernetics*. Londres: Chapman & Hall.)

—— (1962), "Principles of the Self-Organizing System", in VON FOERSTER, H. & ZOPF, JR., G.W. (org.), *Principles of Self-Organization*. Oxford: Pergamon, pp. 255-278. (Reimpresso in ASHBY, W.R. (1981), *Mechanisms of Intelligence*. Seaside (EUA): Intersystems, pp. 65-89.)

ATLAN, H. ([1979] 1992), *Entre o Cristal e a Fumaça*. Rio de Janeiro: Zahar. (Tradução de *Entre le Cristal et la Fumée*. Paris: Seuil.)

CAUDILL, M. & BUTLER, C. (1992), *Understanding Neural Networks*, vols. 1 e 2. Cambridge (EUA): MIT Press.

DEBRUN, M. (1996), "A Dinâmica da Auto-Organização Primária", neste volume, pp. 23-57.

DESCARTES, R. ([1641] 1979), *As Meditações*, in *Os Pensadores*. São Paulo: Abril Cultural, pp. 73-142. (Original em latim: *Meditationes de Prima Philosophia*. Paris: M. Soly.)

HEBB, D.O. (1949), *Organization of Behavior*. Nova Iorque: Wiley. (Reimpressão das pp. xi-xix, 60-78, in ANDERSON, J.A. & ROSENFELD, E. (orgs.) (1988), *Neurocomputing*. Cambridge (EUA): MIT Press, pp. 45-56.)

HINTON, G.E. & SEJNOWSKI, T.J. (1986), "Learning and Relearning in Boltzmann Machines", in RUMELHART, D. & MCCLELLAND, J. (orgs.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1: *Foundations*. Cambridge (EUA): MIT Press, pp. 282-317.

KOHONEN, T. (1984), *Self-Organization and Associative Memory*. Berlim: Springer.

LA METTRIE, J.O. DE ([1747] 1982), *Homem-Máquina*. Lisboa: Estampa. (Tradução de *L'Homme-Machine*. Leiden: É. de Luzac.)

MCCULLOCH, W. & PITTS, W. (1943), "A Logical Calculus of the Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics 5*: 115-133. (Reimpresso in ANDERSON, J.A. & ROSENFELD, E. (orgs.) (1988), *Neurocomputing*. Cambridge (EUA): MIT Press, pp. 18-27.)

NEWELL, A. & SIMON, H. (1972), *Human Problem Solving*. Englewood Cliffs (EUA): Prentice-Hall.

POPPER, K.R. & ECCLES, J.C. ([1977] 1992), *O Cérebro e o Pensamento*. Campinas: Papirus; Brasília: UnB. (Tradução de *The Self and its Brain*. Berlim: Springer.)

RUMELHART, D. & MCCLELLAND, J. (orgs.) (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1: *Foundations*. Cambridge (EUA): MIT Press.

RYLE, G. (1949), *The Concept of Mind*. Londres: Hutchinson.

SMOLENSKY, P. (1986), "Information Processing in Dynamical Systems: Foundations of Harmony Theory", in RUMELHART, D. & MCCLELLAND, J. (orgs.), *Parallel Distributed Processing: Explorations in the*



Microstructure of Cognition, vol. 1: *Foundations*. Cambridge (EUA): MIT Press, pp. 194-281.

TURING, A. M. ([1950] 1973), "Computadores e Inteligência", in EPSTEIN, I. (org.), *Cibernética e Comunicação*. São Paulo: Cultrix/Edusp, pp. 45-82. (Tradução de "Computing Machinery and Intelligence", *Mind 59*: 433-460. Republicado in FEIGENBAUM, E. A. & FELDMAN, J. (orgs.) (1963), *Computers and Thought*. Nova Iorque: McGraw-Hill, pp. 11-35.)

VERSCHURE, P.F.M.J. (1992), "Taking Connectionism Seriously: The Vague Promise of Subsymbolism and an Alternative", in *Proceedings of the 14th Annual Conference of the Cognitive Science Society*. Hillsdale (EUA): Erlbaum, pp. 653-658.

VERSHURE, P.F.M.J.; KRÖSE, B.J.A. & PFEIFER, R. (1992), "Distributed Adaptive Control: The Self-Organization of Structured Behavior", *Robotics and Autonomous Systems 9*: 181-196.

VON FOERSTER, H. (1960), "On Self-Organizing Systems and their Environments", in YOVITS, M.C. & CAMERON, S. (org.), *Self-Organizing Systems*. Oxford: Pergamon, pp. 31-50. (Reimpresso in VON FOERSTER, H. (1984), *Observing Systems*. 2ª ed. Seaside (EUA): Intersystems, pp. 2-22.)

WIENER, N. (1961), *Cybernetics*. 2ª ed. Cambridge (EUA): MIT Press.

(Recebido em dezembro de 1994.)

Informação e cognição: uma proposta de (dis)solução do problema mente-corpo.

Maria Eunice Quilici Gonzales
Depto de Filosofia – UNESP - Campus de Marilia
gonzalez@marilia.unesp.br

Abstract

The concept of **information** seems to play a fundamental role in the models and theories developed by cognitive scientists about the nature and structure of mental states, but , paradoxically, very few analysis of this concept are found in cognitive science. Based upon Dretske's ideas, we propose here an analysis of the concept of information and indicate the relevance of such a concept to cognitive studies of the mind/brain dilemma. A way out of this dilemma is sketched from this informational perspective.

Resumo:

O conceito de informação parece desempenhar um papel fundamental nos modelos e teorias desenvolvidas pelos cientistas cognitivos, mas, paradoxalmente, pouquíssimas análises deste conceito são encontradas na ciência cognitiva. Propomos aqui uma análise do conceito de informação, fundamentada nas idéias de Dretske, e indicamos a relevância de tal conceito para os estudos do dilema mente/cérebro. Uma saída para este dilema é esboçada a partir da perspectiva informacional.

Introdução

Uma das suposições básicas dos cientistas cognitivos, nos seus trabalhos de elaboração de modelos explicativos da natureza e estrutura dos estados mentais, é que estes comportam uma dimensão representacional, passível de simulação através

de computadores. Mesmo admitindo a existência de diferenças óbvias entre computadores e seres vivos, possuidores de estados mentais inteligentes, os cognitivistas ressaltam a existência de um elemento comum fundamental entre ambos, que consiste na capacidade de *processar informação*.

A noção de informação está sempre presente nos discursos dos funcionalistas computacionais e dos cognitivistas em geral, ainda que poucos - como Dretske, Atlan e Smolensky - se preocupem atualmente com a análise do conceito de informação. Se esta parece ser a situação atual dos cientistas cognitivos, o mesmo não ocorreu com os seus progenitores, cuja memória é, segundo Dupuy (1994), ignorada pelos seus descendentes de maneira problemática e sintomática. Assim é que encontramos nas décadas de 50 e 60 uma intensa discussão sobre os conceitos, intrinsecamente correlacionados, de *informação e auto-organização*, principalmente nos trabalhos de Wiener, von Foester, Mackay, Shannon e Weaver, que compartilharam do movimento cibernético originador da atual ciência cognitiva. Sem entrar em detalhes sobre os caminhos e descaminhos por que têm trilhado as tentativas de explicação da natureza dos processos mentais na ciência cognitiva, seria interessante investigar algumas das razões deste aparente desinteresse para com o estudo do conceito de informação. Afinal, se é a capacidade de processar informação que marca o elemento comum entre homens e máquinas, tornando, supostamente, plausível a simulação de estados mentais dos primeiros através dos segundos, então, não mereceria o conceito de informação uma análise cuidadosa, pelo menos por parte dos filósofos que pretendem, além de unificar o projeto interdisciplinar desta ciência, atuar como árbitro nas análises dos métodos e pressupostos adotados nessa área?

Uma razão possível para o estado atual de desinformação sobre a informação na ciência cognitiva pode ser concebida a partir da tendência adotada pelos cognitivistas na década de 70. Estes elaboravam, na Inteligência Artificial, modelos de explicação da natureza dos estados mentais, modelos estes puramente sintáticos, estruturados a partir de regras fixas, dissociadas dos seus componentes físico-ambientais. A noção de informação adotada nesses modelos era restrita àquela proposta por Shannon, na sua teoria matemática de informação, que é estritamente

unesp

quantitativa e não leva em consideração os seus aspectos semânticos (este ponto será retomado na próxima seção) .

Uma outra razão possível para a atual ausência de análises do conceito de informação pode ser inferida a partir de uma sugestão de Küppers (1990), na sua entusiasmante obra ***Information and the origin of life***. Nesta obra, ele sustenta que o entendimento da noção de informação talvez tenha que ser deixado para o futuro: assim como o homem da idade do bronze e do ferro, diz ele, lidavam com estes elementos, mas não dispunham de estruturas conceituais apropriadas para defini-los, também nós, habitantes da era da informação, teremos que aguardar o desenvolvimento das ciências para podermos ir além das metáforas na descrição deste conceito.

Enquanto esse dia não chega, vamos tatear no campo das controvérsias para ver se encontramos algum elemento iluminador, relevante para o nosso estudo dos estados mentais. No que se segue, introduziremos, na seção I, algumas reflexões acerca da noção de informação, apoiadas nas idéias de Dretske em *Knowledge and the flow of information*, para então utiliza-las no estudo do problema da relação mente/cérebro.

O problema da relação mente/cérebro **(MC)** será discutido nas seções II e III como um caso particular do problema da relação mente/corpo, estabelecido no interior do paradigma cartesiano. Seguindo as trilhas de Ryle (1949), adotaremos como hipótese de trabalho a suposição de que este problema pode ser dissolvido através de um exame cuidadoso dos seus pressupostos. Após realizar tal exame, esboçamos uma proposta de (dis)solução do problema **MC** a partir da perspectiva informacional.

I- Notas sobre o conceito de informação

O termo informação, tal como empregado no senso comum, envolve as noções de novidade, conhecimento, dados acerca de alguém ou algo, comunicação ou informe, instrução, direção, etc. Todas estas noções contém, implícita ou explicitamente, um **significado** que é articulado por agentes conscientes.

Além destas conotações, o termo informação é empregado, por vezes entre biólogos e filósofos, para caracterizar aquele **elemento ativo que *dá forma*** à matéria e molda a conduta, através do estabelecimento de **regularidades** no fluxo das variações de sistemas complexos. Supõe-se que tais *regularidades*, quando aplicadas no plano microscópico dos processos biológicos, dariam lugar à formação do código e programa genéticos. No plano macroscópico, do comportamento individual e coletivo dos seres vivos, a informação, entendida como força ativa, direcionaria, instruiria, treinaria e moldaria condutas. Este segundo sentido do termo informação (dar forma a) constituirá o elemento chave para a nossa tentativa de (dis)solução do dilema mente/cérebro.

Uma terceira concepção da informação é desenvolvida por Dretske, na obra *Knowledge and the flow of information.* Inspirado na abordagem matemática da informação desenvolvida por Shannon e Weaver (1949), ele propõe uma teoria da *informação* entendida como uma *"objective commodity"* cuja existência independe da atividade de agentes conscientes. Contudo, diferentemente de Shannon, Dretske está interessado em explicar, a partir de uma concepção objetiva, os aspectos semânticos da informação.

Nas suas investigações sobre a natureza da informação, Dretske se propõe a explicar a noção de *significado* nos sistemas de processamento de *informação* do tipo humano. Para tal, ele desenvolve uma teoria semântica da informação a partir das seguintes hipóteses:

1. Quando certas condições, *entre outras possíveis*, ocorrem com certa probabilidade em uma fonte **F**, então **F** pode ser considerada um elemento gerador de informação.
2. Sempre que mudanças ocorrerem em um estado ou situação **S1** de uma fonte geradora de informação, produzindo, *nomologicamente*, uma mudança em un outro estado ou situação **S2**, e tal mudança for registrada , uniformemente, em **S2** através de um sinal **R**, dizemos que **R** contém informação sobre **S1**.

Enfatizando a idéia de *correlação uniforme* entre a ocorrência de certas condições em uma fonte de informação e um sinal que lhe corresponde, Dretske argumenta que o ***conteúdo informacional*** de um sinal, evento ou estrutura pode ser expresso por uma sentença descrevendo as condições das quais dependem, de *modo legiforme*, as propriedades de um sinal na fonte.

A noção de dependência ou **correlação nômica** é fundamental tanto para Shannon (e os cibernéticos de sua época) quanto para Dretske. Os primeiros a utilizaram para medir a quantidade média de informação gerada em uma fonte, em função das freqüências de ocorrência de eventos que guardam entre si uma relação de dependência, manifesta através de algum tipo de sinal. Dretske, por sua vez, utiliza esta noção para definir o ***conteúdo***

unesp

informacional de um sinal: um sinal **R**, diz ele, carrega a informação específica que **S** é **F** sempre que a ocorrência de **S** implicar **necessariamente** a ocorrência de **F** e tal correlação for registrada em **R**. Assim, por exemplo, o tocar do telefone (expresso através do sinal triim! trim!..) carrega a informação (S) que alguém discou o seu número (F), em todas as circunstâncias em que o telefone estiver funcionando normalmente.

Ainda que a correlação entre a discagem de um certo número e a produção de um som característico seja inicialmente contingente, uma vez estabelecida (entre inúmeras outras possíveis), tal correlação torna-se necessária na trama informacional. através da sua recorrência. Nesta perspectiva, eventos correlacionados ao acaso não constituem informação. Mesmo que o telefone tenha tocado no exato momento em que a camélia caiu do galho no jardim, tal correlação, que se estabeleceu ao acaso, não transmite informação sobre quedas de camélias ou o tocar de telefones.

Conforme explicitado em Pereira Jr. e Gonzales (1995), existe uma diferença, relevante para as nossas considerações, entre uma relação causal e uma relação informacional. A primeira pode ser entendida como uma manifestação de uma sucessão entre eventos **C** e **E** tal que, sempre que **C** ocorre, então **E** ocorre. Isto é, **C** é condição suficiente de **E**. Uma *relação informacional,* por sua vez, expressa uma relação de causalidade de segunda ordem, uma vez que ela requer correlações entre padrões ou grupos de correlações. Além disso, a ocorrência do mesmo tipo de evento **C**, nas mesmas circunstâncias, poderia conduzir a efeitos diversos **E**, **F**, **J**, etc. ou a nebhum deles. Assim, por exemplo, posso estar diante de uma situação ou evento e não estabelecer com eles uma relação informacional - a chave que procuro por horas pode estar diante dos meus olhos. Alternativamente, posso **não** estar diretamente em contato com uma fonte de informação e, mesmo assim, ter acesso indireto à informação por ela emitida através, por exemplo, de TV, rádio, etc. Assim, na relação informacional, **C** não é conndição necessária nem suficiente de **E**. Contudo, se (e quando) uma relação informacional se estabelece, a dependência entre um sinal e as condições na fonte que o geraram deve ser inequívoca. De outra forma, estaríamos diante de ruído ou ambiguidade, mas não de informação Em outras palavras, **uma vez estabelecida** uma relação informacional, a transmissão de informação requer, não apenas um conjunto de correlações quaisquer, mas uma rede de dependências nômicas entre as condições na fonte e as propriedades do sinal que expressam tais condições. Esta hipótese acerca da natureza das relações informacionais é fundamental para entender a caracterização dretskeana de **conteúdo informacional**, segundo a qual:

unesp

Um sinal **R** carrega a informação que "**S é F**" se e somente se a probabilidade condicional de **S** ser **F**, dado **R** (e K= informação acumulada) é 1 (mas, dado apenas K, é menos do 1). (Dretske, 1981, p. 65)

A partir desta definição de conteúdo informacional de um sinal, juntamente com as mencionadas hipóteses (1) - que caracteriza uma fonte geradora de informação, e (2) - que admite a possibilidade de um sinal expressar as condições de ocorrência de certos estados de uma fonte, Dretske caracteriza a noção de **conhecimento perceptual.** Este seria um estado do organismo, possivelmente um estado cerebral, uma crença descrita por um sinal, cujo conteúdo informacional corresponde, fielmente, àquelas condições na fonte que o gerou.

A definição dretskeana do conhecimento (perceptual) é oferecida como uma alternativa para a definição tradicional do conhecimento (em termos de crenças verdadeiras justificadas), que reconhecidamente envolve circularidade. Tal definição informacional do conhecimento constitui um dos alicerces fundamentais sobre a qual se assenta a nossa proposta de (dis)solução do problema mente/cérebro que analisamos a seguir.

II - Breve histórico do problema da relação mente/corpo e o caso particular do problema mente/cérebro

Ainda que os antigos já tivessem questionado a natureza do pensamento, foi principalmente a partir de Descartes que os filósofos passaram a investigar, sistematicamente, a natureza dos estados mentais e sua relação com o corpo, desenvolvendo-se um debate acerca da *distinção* entre mente e corpo e, mais recentemente, entre mente e cérebro. Antes de introduzir alguns elementos sobre este debate, vamos contrastar, brevemente, as visões de mundo aristotélica e cartesiana, de forma a fundamentar a nossa proposta de (dis)solução do problema mente/cérebro nas seções subseqüentes.

Estamos tão habituados a refletir cartesianamente nas nossas considerações acerca do problema da relação mente/corpo que torna-se difícil conceber uma caracterização alternativa para o mesmo. Contudo, julgamos que, de modo a entender

unesp

a própria formulação cartesiana do problema, com suas vantagens e dificuldades, seria interessante situá-la no pano de fundo do paradigma aristotélico.

O primeiro ponto de interesse para nós deste paradigma é que a abordagem aristotélica, ao atribuir grande importância ao *movimento*, caracteriza e classifica os objetos em termos do que eles fazem, e não em termos de *suas propriedades intrínsecas e imutáveis,* como o faz Descartes. Conforme ressalta Betchel (1988), esta diferença sutil e fundamental na maneira de "ver" os objetos leva a linhas de investigação radicalmente diferentes. Assim, Aristóteles concebe os corpos como *matéria organizada* de acordo com uma *forma* particular. É, principalmente, através da *forma*, que os objetos, animados ou inanimados, são classificados. No caso particular dos seres vivos, a *forma* é caracterizada através do conceito de *psyche* ou "*alma*". Esta não deve ser entendida como uma *substância acoplada ao* corpo, mas, sim, em termos das próprias *capacidades* e *características definidoras* do organismo, que lhe conferem uma identidade.

Descartes, rejeitando a concepção aristotélica de matéria e forma, compartilha a concepção da matéria que marcou a revolução científica do século XVII. A partir dessa época a matéria passa a ser entendida como *passiva e inerte*: os *corpos materiais*, cuja propriedade intrínseca é a *extensão*, ocupam um lugar no espaço e são nele movidos graças à ação de forças externas, sendo *mecanicamente* afetados por estas de acordo com *leis* físicas. O problema começa a se esboçar quando esta caracterização dos corpos é estendida aos seres humanos. Como explicar a ação humana a partir da perspectiva do movimento dos corpos regidos por leis mecânicas?

O ideal mecanicista da física newtoniana, que traz no seu bojo a expectativa do poder de explicação, previsão e controle da natureza, apesar de agradar imensamente o lado científico cartesiano, não satisfaz as aspirações do Descartes filósofo e religioso, que visa fundamentar o conhecimento e a ação humana livre. A teoria cartesiana dos animais máquinas, segundo a qual os animais são concebidos como meros sistemas mecânicos, cujas partes são arranjadas de forma a realizar determinadas funções, não resolve o problema, que agora se coloca, o qual consiste em explicar como é possível a ação humana livre e responsável. Para tal, será preciso acoplar ao corpo humano uma substância pensante imaterial, a *alma*, que,

unesp

diferentemente dos corpos físicos, não ocupa lugar no espaço, não é perecível, nem decomponível e, principalmente, não está sujeita às leis mecânicas da física. Esta substância, caracterizada negativamente, em oposição ao corpo, terá o poder de movê-lo segundo critérios de relevância condizentes com a razão humana livre.

Temos instaurado então, no paradigma cartesiano, dois tipos de substância ou propriedades básicas. Por um lado, a substância *material*, cuja propriedade intrínseca é a *extensão* e, por outro lado, a substância *espiritual*, cuja propriedade intrínseca é o *pensamento*. A primeira estaria sujeita às leis da física e da geometria, e a segunda à um tipo de causalidade não mecânica, que pemitiria desenvolver e fundamentar o conhecimento humano, condição de possibilidade do livre arbítrio. O ser humano expressaria, com exclusividade, a junção destas duas substâncias através do corpo e da alma, reunidos pela da obra divina.

Assim caracterizada a junção entre a alma e o corpo humanos, resta a Descartes explicar aquilo que será conhecido, nos séculos posteriores, como *o problema da relação entre a mente e o corpo: como é possível a interação de substâncias de natureza tão distintas?*

Nas *Paixões da alma*, Descartes analisa vários casos de interação entre estados do corpo (por exemplo, estado do fígado) e estados mentais (como a inveja, que estaria presente nos indivíduos com excesso de bílis), ressaltando que não temos idéias claras e distintas sobre a interação entre a alma e o corpo, enquadrando o nosso conhecimento à seu respeito no domínio das especulações. Dentro deste domínio, ele sugere a conhecida solução da glândula pineal, que, situando-se no cérebro, funcionaria como sede da alma. Esta glândula, disposta na forma de uma bexiga, afetaria o movimento dos pequeníssimos corpúsculos, denominados *espíritos animais*. Estes, ao serem deslocados através da corrente sangüínea, chegariam até a suposta glândula, por meio dos canais nervosos, influenciando assim o movimento do corpo.

A solução da *glândula pineal*, sede da alma, sugerida como uma tentativa de explicar a união entre a mente e o corpo não é, obviamente, uma boa solução. Na condição de substância física, esta glândula não traz qualquer novidade a respeito da possibilidade de interação entre ela e a substância espiritual, imaterial. Ciente desta

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

dificuldade, Descartes irá deixar nas mãos de Deus a tarefa de estabelecer, com seu poder infinito, esta interação. Ele deixará também a sinalização, bastante polemica, da necessidade de uma ciência especial, *distinta da física*, para o estudo do homem.

Mesmo ineficiente, a sugestão cartesiana da glândula pineal, à medida em que esta se situa no cérebro, sugerirá, aos pesquisadores do futuro, uma forma de delimitar o problema da relação mente/corpo, que passará então a ser conhecido como o problema mente/cérebro. Ao invés de investigar como o corpo se relaciona com a mente, a versão moderna do problema será: como o cérebro se relaciona com a mente?

Conforme veremos, o problema MC permanece não resolvido. Além disso, o amplo debate entre monistas e dualistas, que se estabeleceu principalmente entre filósofos, cognitivistas e psicólogos, parece deixar intocada a raiz do problema - qual seja, a busca de uma explicação, não mecanicista, da ação humana racional e livre.

III- Propostas de solução do problema mente/cérebro e algumas de suas limitações

O dualismo entre matéria e mente pode ser dividido em *dualismo de substância e dualismo de propriedade*. O primeiro, mais conhecido dos dois, é representado pelo *dualismo cartesiano*, introduzido na seção II, segundo o qual o ser humano é composto por duas substâncias distintas - espiritual, cuja propriedade central é o pensamento e a material, cuja propriedade central é a extensão.

O segundo, *dualismo de propriedade*, algumas vezes conhecido como *epifenomenalismo*, admite, sem postular a existência de duas substâncias distintas, a possibilidade da mente ser material, ou até mesmo puramente cerebral, não existindo independente da matéria cinzenta. Contudo, de acordo com esta concepção, existe um *domínio próprio do mental* que se caracteriza através das propriedades e relações que se estabelecem entre os estados cerebrais, mas que não se reduzem à matéria, nem podem ser explicados unicamente pelas leis da física, biologia ou da neurologia. Neste

unesp

domínio se encontram os desejos, pensamentos, crenças, tristezas, alegrias, etc., que caracterizam os fenômenos mentais, emergentes da (mas não redutíveis à) atividade cerebral. Esta vertente do dualismo, não muito conhecida, tem sido recentemente objeto de investigação dentro dos estudos de auto-organização, aos quais faremos referência na parte final deste ensaio.

O *dualismo de substância*, por sua vez, tem sido objeto de admiração e crítica por três séculos. Do lado dos admiradores, a adesão dos filósofos, psicólogos e pessoas do senso comum foi tão grande que as idéias de Descartes passaram a ser consideradas como constituindo o que Ryle (1949) chama de *a doutrina oficial*.

Do lado dos críticos, encontramos a divisão entre monistas materialistas e idealistas absolutos. A nossa análise estará centrada nas idéias dos primeiros, que constituem atualmente fonte de interesse para a filosofia da mente e ciência cognitiva. Estes tiveram entre os seus representantes, por exemplo, no século XVIII, o mecanicista francês de La Mettrie, que na obra *O homem máquina*, defende a hipótese segundo a qual a mente humana nada mais é do que uma forma complexa de organização mecânica da matéria. "Adaptando" algumas das idéias de Descartes sobre o animal máquina, de La Mettrie estende a hipótese do mecanicismo aos seres vivos em geral (incluindo o homem), admitindo-os como máquinas vivas dotadas da capacidade de auto-locomoção. Para ele, o homem é uma máquina complexa, cuja dinâmica de organização deu lugar, por excelência, à capacidade de pensar e aprender, segundo determinados princípios mecânicos que atuam sobre a matéria:

> Considero o pensamento tão pouco incompatível com a matéria organizada , que parece ser uma propriedade desta, tal como a eletricidade, a faculdade motriz, a impenetrabilidade, etc." (de La Mettrie, p. 3).

Através do cérebro, que La Mettrie julgava constituir a peça central da nossa maquinaria, o homem pode estocar e modificar informações, ajustando e aprimorando a sua própria organização. Com seu trabalho, chocante para a época, de La Mettrie parece ter sido um dos precursores dos modelos de redes neurais artificiais auto-organizadas, que constitui um dos principais focos de estudo da moderna ciência cognitiva.

unesp

Os *monistas materialistas* atuais dividem-se principalmente em *Materialistas Reducionistas e Eliminativistas*. Entre os primeiros encontramos os nomes de Place, Armstrong e Smart, como os principais defensores da *teoria da identidade cérebro/mente*. Entre os segundos estão Churchland, Feyerabend e Feigl, e outros.

A teoria da identidade, divide-se, por sua vez, em *identidade de tipo* (mind-brain type identity theory) e identidade de *membros específicos* (token identity theory). A primeira sustenta que todos os casos de um tipo particular de estado mental são idênticos, e não simplesmente correlacionados, aos casos de um certo *tipo* de evento cerebral. Por exemplo, a alegria que sinto agora é idêntica à alguma instância de um tipo de evento cerebral, digamos correspondente à ativação de um grupo de neurônios de um certo tipo.

A segunda vertente da teoria da identidade, a *token identity theory,* constitui uma versão mais fraca do que a primeira, no sentido que ela considera que a existência de um estado mental pressupõe a existência de um estado cerebral, mas não exige que determinado tipo de estado mental corresponda a um tipo específico de estado cerebral. Assim, como coloca Betchel, esta teoria sustenta que:

(a) toda vez que me encontro em um particular estado mental, este estado mental é idêntico a um estado cerebral, mas, (b) em outras ocasiões, quando me encontro no mesmo estado mental, posso ter um estado cerebral diferente. (Betchel, 1988, p.107).

Uma conseqüência óbvia desta posição é que não se pode inferir os estados mentais a partir de suas descrições físicas equivalentes. Conforme veremos na terceira parte deste ensaio, os cognitivistas funcionalistas da linha lógico-computacional, como Fodor e Pylysyn, entre outros, compartilham desta vertente da teoria da identidade.

A teoria da identidade, nas suas duas versões, tem sofrido várias críticas por parte dos filósofos, sendo que as mais interessantes se apoiam, em geral na *lei de Leibniz* da identidade dos indiscerníveis, segundo a qual, se dois termos se referem ao mesmo objeto, então qualquer propriedade que se atribua ao objeto referido pelo primeiro termo deve poder ser igualmente atribuída a esse objeto pelo segundo termo e vice-versa (Betchel, 1989, p. 96). Para refutar a teoria da identidade, os críticos enumeram propriedades que podem ser atribuídas aos estados mentais mas não aos

unesp

estados cerebrais e vice-versa. Por exemplo, acredita-se que temos acesso direto, por introspeção, aos nossos estados mentais mas não aos nossos estados cerebrais; os estados cerebrais ocupam lugar no espaço mas não podemos dizer que os estados mentais ocupem um lugar no espaço, etc. Réplicas a tais críticas consistem em argumentar que a lei de Leibniz não se aplica a este tipo de questão. Assim, pode-se dizer de Antônio, que nunca estudou astronomia, que ele conhece a estrela da manhã sem conhecer Vênus, ainda que elas sejam idênticas. Da mesma forma, posso conhecer os meus estados de dor, sem conhecer neurociências, ignorando que tais estados são estados cerebrais de um certo tipo.

Dificuldades com a teoria da identidade levaram alguns de seus adeptos iniciais, a abandoná-la, adotando um materialismo mais radical, conhecido como *materialismo eliminativista.* Deixando de lado a tentativa de estabelecer um elo entre as neurociências e a psicologia, os materialistas eliminativistas acreditam que, quando as neurociências estiverem plenamente desenvolvidas, o jargão da psicologia será completamente abandonado. Cognitivistas como Churchland, acreditam que, da mesma forma que abandonamos a linguagem mítica, da ira dos deuses, ao nos referirmos, por exemplo aos trovões, para empregar, com o avanço da física, uma linguagem cientificamente aceita, que explica o mesmo fenômeno em termos de cargas elétricas, assim também, quando tivermos uma teoria apropriada dos nossos estados mentais, deixaremos de falar em crenças, desejos, etc. que expressam noções obscuras da psicologia popular.

As objeções ao materialismo eliminativista são muitas, sendo que as mais interessantes apontam para as dificuldades de se conceber, num futuro próximo, qualquer progresso nas neurociências que venham a estabelecer a revolução *desejada* pelos adeptos desta vertente do materialismo. Além disso, eles também não oferecem qualquer proposta para lidar com aquilo que na seção II deste ensaio consideramos a raiz do polêmico dualismo cartesiano, a saber - a necessidade de se buscar um paradigma apropriado para explicar a ação humana, em princípio livre.

Para concluir, apresentamos um esboço de (dis)solução do problema MC, fundado nas teorias de auto-organização e informação, que parece dar lugar ao tratamento do que consideramos a raiz da problemática do dualismo cartesiano.

unesp

IV- Sugestões para um esboço de (dis)solução do problema MC

Sugerimos na seção III que a posição cognitivista do funcionalismo computacional, ainda em voga na moderna ciência cognitiva, é compatível com a vertente da teoria da identidade conhecida como *token identity theory*. Conforme mencionamos, os adeptos desta teoria sustentam que cada estado mental particular é idêntico a um estado cerebral, mas que o mesmo estado mental pode ter, em situações e ocasiões diversas, um estado cerebral diferente. Para eles, um mesmo estado mental pode ser instanciado em diferentes sistemas físicos, e, consequentemente não se pode inferir os estados mentais a partir de suas descrições físicas equivalentes. No jargão da Inteligência artificial se diz que um mesmo programa, simulando um estado mental particular, pode ser instanciado em diferentes hardwares. Contudo, o estudo dos hardwares não permite inferir as regras que estruturam o programa.

Mesmo admitindo a relevância das neurociências para o estudo da cognição, estes pesquisadores chegam à conclusão, paradoxal, que a análise apropriada do sistema cognitivo não deve ser realizada no plano físico, uma vez que não se pode inferir os estados mentais a partir de suas descrições físicas equivalentes. A análise dos estados mentais deve se situar no plano abstrato, tendo como foco a racionalidade e as regras que caracterizam o pensamento racional. O estudo dos aspectos físicos correspondentes a tais estados mentais deve ser deixado para os neurologistas.

A posição paradoxal a que parecem chegar estes cognitivistas e adeptos da teoria da identidade tem despertado inúmeras críticas entre filósofos, psicólogos e cognitivistas da vertente conexionista, atualmente em fase de desenvolvimento na ciência cognitiva. Sem entrar em detalhes a respeito destas críticas vamos esboçar uma solução para esse debate, inspirada nas idéias de Gilbert Ryle, na tentativa de esboçar uma solução para este debate, que resulta na dissolução do problema da relação mente/corpo.

Na sua obra *The concept of mind*, Ryle denomina, ironicamente, a teoria cartesiana de "a teoria do fantasma na máquina". Para Ryle, o conflito de Descartes, que consistia em conciliar sua postura de homem da ciência com aquela do homem



religioso, leva-o a cometer um erro categorial ao tentar classificar a mente como uma **substância** imaterial. Este erro categorial consiste em tentar classificar ou enquadrar elementos pertencentes à uma determinada categoria lógica em outra categoria à qual eles não pertencem. Ryle exemplifica este tipo de erro com a ilustração de um estudante a quem se procura ilustrar o conceito de Universidade. Após ter-lhe sido mostradas as salas de aula, biblioteca, salas dos professores, etc., o estudante pergunta : e onde está a Universidade? Outro exemplo engraçado é o do menino que ao assistir um batalhão em um desfile e, após ter visto passar os seus membros ordenadamente, pergunta à mãe: e onde está o batalhão? Ou, ao se observar uma partida de futebol: onde está o espírito de time? Com estes e outros exemplos, Ryle procura ilustrar o tipo de erro categorial decorrente da tentativa de enquadrar os conceitos relacionais do tipo, Universidade, batalhão, espírito de time etc. em conceitos correspondentes à elementos físicos do tipo sala de aula, jogadores de futebol, soldados etc.

Para Ryle, Descartes teria incorrido no mesmo tipo de erro categorial ao tentar enquadrar o conceito de espírito ou mente, " por oposição" ao de corpo, nos moldes da física da época, ao mesmo tempo que buscava explicar suas características intrínsecas fundamentais, que identificam a alma livre e responsável. O resultado deste erro categorial consistiu em classificar a mente como uma **substância** imaterial, um "fantasma" na máquina, que inevitavelmente daria lugar ao problema da relação mente e cérebro, entre outros.

Este problema se dissolve, na opinião de Ryle, ao deixarmos de entender a mente como uma substância e passarmos a entendê-la como um processo dinâmico, de constante atualização e criação de *estados disposicionais* que habilitam o organismo a agir. A destruição do fantasma na máquina destruiria também as vertentes *materialistas e dualistas* da mente, as quais constituem, para Ryle, versões opostas de esforços equivocados para solucionar um pseudo-problema. O estudo da mente estaria intrinsecamente ligado ao estudo dos estados disposicionais do organismo e de suas capacidades, adquiridas através do hábito.

Na mesma linha de argumentação, poderíamos caracterizar agora os estados mentais a partir da perspectiva informacional. Estes seriam estados disposicionais do

unesp

organismo que, quando instanciados possivelmente no cérebro, poderiam ser expressos através de sinais com conteúdos informacionais específicos (no sentido proposto por Dretske). Sempre que tais conteúdos correspondessem, fielmente, às condições especificadas pelas fontes informacionais que os geraram, eles representariam crenças verdadeiras, desempenhando um papel informativo na história do organismo. Quando tal correspondência não se verificasse, teríamos crenças falsas, que necessitariam correção, através de mecanismos de aprendizagem .

Neste sentido, os estados mentais seriam estados informacionais, entendidos como forças ativas que dariam forma à matéria (no plano microscópico) e à conduta (no plano macroscópico). O estudo das relações informacionais no plano macroscópico deixa em aberto a possibilidade de uma análise do que na seção II consideramos a raíz do dualismo cartesiano, qual seja, a necessidade de se explicar a ação livre e responsável.

Ao tentar estabelecer uma distinção entre os nossos atos, realizados na condição de agentes portadores de uma vontade controladora de tendências muitas vezes desenfreadas, e aquilo que nos acontece na condição de pacientes sujeitos às leis da natureza e do acaso, deparamo-nos com duas tendências geraia, bem delineadas por Atlan (1992). A primeira, do determinismo, reduz a consciência e a vontade livres a meras ilusões. As ações são reduzidas à seqüências de eventos físicos causalmente relacionados e possivelmente previsíveis por um "deus laplaciano". Em oposição, a segunda tendência, característica do absolutismo espiritualista, distingue a mente da matéria e instaura uma consciência imaterial, propriedade exclusiva da alma humana.

A perspectiva informacional que acabamos de esboçar, sem concordar seja com o determinismo ou com o absolutismo espiritual, deixa espaço para a pergunta: Haveria um correspondente informacional da consciência e da vontade nos organismos ou sistemas que se auto-organizam?

A esperança é que um agente, de posse da *informação* acerca das várias implicações do seu agir, seja tanto mais livre quanto mais esta informação puder ajudá-lo a discriminar as consequências adequadas e inadequadas de sua ação, em função de sua sobrevivência.

Liberdade, inteligência e noções afins, estariam, assim, intrinsecamente ligadas à capacidade informacional dos organismos na medida em que ela contribuísse para a sua sobrevivência. Conforme indicado por Debrun (1996b), a atividade de um tal agente

unesp

"...quando é dona de si, dominando seus objetos, capaz de reflexividade, de memorização explícita e de projetos - deveria ser recolocada, de um lado, no seu contexto (ou base) orgânico e ambiental para que se possa melhor entender seu funcionamento. E, de outro lado, cotejada com outras atividades cognitivas, em particular a do mesmo sujeito quando, visando o seu corpo, fica "perto" dele em vez de sobrevoa-lo (ou tentar sobrevoar) como um objeto externo" (p. 38).

Sem a pretensão de se conhecer por inteiro ou de encontrar verdades absolutas, o nosso sujeito" informacional" navegaria no fluxo das informações disponíveis no seu meio e estabeleceria os rumos da sua ação a partir da dinâmica resultante das várias forças circundantes. Sem um centro organizador ou alma, ele buscaria a sua identidade no processo de auto-organização que molda o seu viver.

Bibliografia

Abrantes, P . (org) (1994). *Epistemologia e cognição*; Brasília: UNB

Ashby, W. R. (1962). Principles of the self organizing systems. In von Foerster & Zopf (orgs). *Principles of self-organization*: Pergamon Press.

Atlan, H. (1992). *Entre o cristal e a fumaça*: ensaio sobre a organização do ser vivo. Trad. Vera Ribeiro; Rio de Janeiro : Ed. Zahar.

Betchel, W. (1988). *Philosophy of Mind:* an overview for Cognitive Science. Hillsdale: Erlbaum.

Blakemore, C. & Greenfield, S. (Edts) (1989). *Mindwaves*. Oxford: Blacwell Ltd.

Caudill, M. & Butler, C. (1993). *Naturally intelligent systems*. Cambridge: MIT Press.

Churchland, P. S. (1986). *Neurophilosophy*: towards an unified science of mind and brain; Cambridge: The MIT Press.

Debrun, M. (1996a). *A dinâmica da auto-organização Primária*; in Debrun Debrun, M.; Gonzales, M.E. & Pessoa, Jr. (Orgs). Auto-organização. Coleção CLE. Campinas: Ed. Unicamp, pp. 23-57.

Debrun, M. (1996b). *Auto-organização e ciências cognitivas*. In: Gonzales, M.E. & Lungarzo, C; Milidoni, C. & Pereira Jr, A. & Wrigley (orgs), pp 29-38.

Debrun, M. ; Gonzales, M.E. & Pessoa, Jr. (Orgs) (1996). *Auto-organização:* Estudos interdisciplinares. Coleção CLE, Campinas: Unicamp.

Descartes, R. (1979). *As Meditações*. Trad. de Guinsburg, J. & Prado Junior, B. São Paulo, Edit. Abril Cultural.

_____. (1979). *As paixões da alma.* São Paulo, Edit. Abril Cultural.

unesp

Dorffner. G. (Edit) (19997). *Neural networks and a new artificial intelligence*. London: Thomson Computer Press.

Dretske, F.(1981). *Knowledge and the flow of information*. Cambridge: Bradford Books

Dupuy , J. (1994). *Nas origens das ciências cognitivas*; Trad. de Roberto Leal Ferreira. São Paulo: Edit. UNESP.

Fodor, J. A. (1988). *Psychosemantics*. New York: Thomas Crowel.

Gardner, H. (1987). The mind's new science: a history of cognitive revolution; New York: Basic Books.

Gibson, J.J. (1979). *The ecological approach to visual perception.* Boston: Houghton-Mifflin.

Gilholy, K., Keane, M. & Erdos, G. (Eds.) (1990). *Lines of thinking*: Reflections on the psychology of thought. England: John Willey & Sons.

Gonzales, M. E.; French, T. & Treffner, P. (1990). *A Naturalistic Approach to Mental representation*. In: Gilhooly, pp. 57-69.

Gonzales, M. E. (1991). Redes Neurais e representação mental: um ensaio sobre harmonia e racionalidade. *TRANS/FORM/AÇÃO*. São Paulo, v. 14, pp. 93-108.

Gonzales, M. E. (1994). *Um estudo cognitivo-informacional das representações mentais*. In: Abrantes, P. (org), pp. 127-146.

Gonzales, M. E. (1996). Ação, causalidade e acaso nas redes neurais auto-organizadas. In Debrun, M. ; Gonzales, M.E. & Pessoa, Jr. (Orgs), pp. 273-298.

Gonzales, M.E. & Lungarzo, C; Milidoni, C. & Pereira Jr, A. & Wrigley (orgs) (1996). *Encontro com as Ciências Cognitivas*. Ed. da Faculdade de Filosofia e Ciências, campus de Marília, Unesp.

Hebb, D. (1949). *The organization of behaviou.*; New York: John Wiley & Sons.

Hinton, G. & Sejnowski, T. (1986). *Parallel models of associative memory.* New Jersey: Erlbaum.

Küppers, B. (1990). *Information and the origin of life.* Cambridge: MA. The MIT

Kohonen, T.(1984). *Self-organization and associative memory*. Berlin: Spring-Verlag.

La Mettrie, J. O. De (1747/1982). *Homem-máquina*. Lisboa: Estampa.

Mackay, D. M. (1969). *Information, mechanism and meaning.* Cambridge, MIT Press.

Pereira Jr., A & Gonzales, M. E. (1995). Informação, organização e linguagem. In: Évora, F. (org.). *Espaço e Tempo.* Coleção CLE, Campinas: Unicamp.

Rumelhart , D. & MacClelland, J. (1986). Parallel distributed processing: Explorations. In: *The microstructure of cognition*. vol.1; Cambridge, MIT. Press.

Ryle, G. (1949) . *The concept of mind.* Londres, Hutchinson.

Shannon, C. & Weaver, W. (1949). *The mathematical theory of communication.* Illinois: The University of Illinois Press.

Smolensky, P. (1986). *Information processing in dynamical systems:* Foundations of harmony theory. In: Rumelhart & MacClelland (1986, 390-431).

Smolensky, P. (1988). On the proper treatment of connectionism. In: *The Brain and Behavioral Sciences*, 11, 1-74.

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

Stonier, T. (1990). *Information and the internal structure of the universe.* London: Springer-Verlag.

Searle, J. (1984). *Mente, cérebro e ciência.* Tradução de Artur Morão; Biblioteca de Filosofia Contemporanea; Edições 70.

Verschure, P. (1997*).* *Connectionist explanation: taking positions in the mind-brain dilemma*. In: Dorffner. G. (19997), pp. 133-137.

Von Foerster & Zopf (orgs) (1962). *Principles of self-organization*: Pergamon Press.

Wiener, N. (1961). *Cybernetics or control and communication in the animal and the machine*. Cambridge: MIT Press.

unesp

PARTE III

Auto-Organização, Identidade Pessoal e Conhecimento Comum

Auto-Organização, org. por I.M.L. D'Ottaviano & M.E.Q. Gonzales.
Coleção CLE 30, Campinas, 2000. p. 69-81.

3

Auto-organização, Autonomia e Identidade Pessoal[1]

MARIA EUNICE Q. GONZALEZ

Departamento de Filosofia, FFC, UNESP – Campus de Marília, SP
gonzales@marilia.unesp.br

MARIANA C. BROENS

Departamento de Filosofia, FFC, UNESP – Campus de Marília, SP
mbroens@marilia.unesp.br

JULIANA SERZEDELLO

Departamento de Educação Física, UNESP – Campus de Rio Claro, SP
jugui@uol.com.br

1. Introdução

Uma das grandes novidades que surge com o movimento cibernético na década de 50 é a noção de **causalidade circular**. Difícil de definir mas fácil de exemplificar, a noção de causalidade circular remete-nos a situações em que o efeito de uma causa volta-se para a própria causa, alterando-a e, simultaneamente, alterando o efeito que, de novo, afetará a causa e assim sucessivamente. Um exemplo desse tipo de causalidade pode ser verificado nos indivíduos que, ao se tornarem pais, sabem bem o quanto os filhos alteram a sua identidade e a sua forma de ser, mudança esta que, por sua vez, afeta diretamente o filho, que dinamicamente conti-

[1] Trabalho apresentado no *Tenth International Conference on Perception and Action*, Agosto, 1999. Universidade de Edimburgo, Escócia. Apoio: FAPESP.

nuará afetando a identidade dos pais e assim indefinidamente. De modo semelhante, indivíduos criam vários tipos de organizações sociais que passam a afetar sua identidade pessoal. As mudanças individuais, por sua vez, refletem-se na organização que os indivíduos criam, passando a afetar cada um deles e assim por diante de modo dinâmico e contínuo.

O aspecto revolucionário desta noção de causalidade circular, que marcou o início da nova cibernética, reside na possibilidade de construir máquinas que, incorporando um tal mecanismo, teriam a habilidade de gerar e alterar hábitos em sua interação com o meio ambiente. Até bem recentemente, as máquinas, por mais sofisticadas que fossem, não passavam de meros artefatos passivos, como os patos mecânicos, bonecas falantes, relógios ou computadores digitais manipuladores de símbolos. Foi, justamente, com o surgimento de noções relacionadas à causalidade circular, autonomia, auto-organização, etc. que robôs móveis começaram a ser desenvolvidos (VERSHURE, 1992), dotados da habilidade de interagir com o meio ambiente e de criar hábitos através de processos de **auto-organização**. Consideramos que essas máquinas constituem instrumentos úteis para ilustrar a capacidade da matéria de criar **disposições ou tendências dinâmicas**, constitutivas dos hábitos. Estes por sua vez são fundamentais para entender o processo de formação de identidade em sistemas processadores de informação.

Apresentaremos, na seqüência, as noções de **estado disposicional** e de **auto-organização** com o propósito de analisar o processo de formação de identidade pessoal em termos da capacidade para adquirir e mudar hábitos.

2. Estados disposicionais e o processo de formação de identidade

O conceito de *disposição* foi objeto de análise filosófica durante séculos sem muito sucesso. Desde Aristóteles foram concebidas disposições ou formas como forças potenciais que poderiam dar lugar a diferentes atualidades. A principal dificuldade associada a esta noção é que as disposições

não podem ser observadas e, em muitos casos, suas manifestações somente podem ser preditas com um grau muito vago de precisão. De modo caricatural, pode-se dizer que até recentemente as disposições eram vistas como fantasmas, cuja ação poderia ser sentida por aqueles que acreditassem em sua existência, não sendo observável ou passível de previsão.

Atualmente, com os avanços da Química e da Física, as disposições podem ser entendidas em termos de organizações moleculares, invariantes, de elementos físicos e químicos que, em circunstâncias adequadas, causam realidades específicas e observáveis. Estas podem produzir efeitos que afetam suas causas produzindo eventos ainda mais complexos. Desse modo, por exemplo, dada a organização molecular de certos tipos de sais, estes tornam-se solúveis em água. A organização molecular do hidrogênio e do oxigênio na água, por sua vez, permite que ocorra a solubilidade dos sais. Neste exemplo, as disposições podem ser explicadas e sua ação calculada com precisão.

Como observa TURVEY (1988), as disposições são propriedades complementares de **sistemas** e, assim, sempre apresentam-se em pares. Elas indicam aquelas possibilidades no mundo real que darão lugar a atualidades quando (e se) encontrarem condições adequadas.

No caso dos organismos, as disposições podem ser ilustradas através da concepção Gibsoniana de *affordance* (disponibilidade) que pode ser definida como "... a combinação invariante de superfície, substância e propriedades tomada com referência a um animal" (GIBSON, 1979). Essas disponibilidades constituem estados disposicionais; elas caracterizam propriedades complementares do sistema organismo/meio ambiente, permitindo aos animais movimentarem-se e agirem de acordo com as leis que estruturam tanto os organismos quanto o meio ambiente. Assim, por exemplo, as superfícies planas tornam possível o caminhar, regulando-o, para organismos que como nós desenvolveram pernas conforme a nossa história bioquímica e evolutiva.

Gibson salienta que as *affordances* têm existência independente dos estados subjetivos dos organismos; elas existem como conseqüências



naturais (governadas por leis) e reguladoras de um conjunto de estados disposicionais, decorrentes de possíveis desenvolvimentos na história do sistema organismo-meio ambiente. Porém, dado que pode haver muitas disponibilidades na relação organismo-meio, sua "individuação" pode indicar marcas importantes na formação da identidade dos sistemas. Assim, por exemplo, quer estejamos tristes ou alegres, a água nos possibilita (*affords*) o ato de beber; o fogo, o ato de queimar; certos tipos de alimento, o ato de comer, etc. Tais *affordances* existem graças as leis físico-biológicas e sociais que moldam a nossa ação, limitando o domínio dos eventos que nos cercam. A probabilidade de uma disposição ser estabelecida de um modo específico revela informações sobre a natureza do sistema no qual ela se efetiva e, por conseguinte, sobre a natureza do organismo que atua em tal processo de seleção.

Se aceitarmos a suposição da inseparabilidade entre o organismo e o meio ambiente (entendida aqui como um sistema unificado de disposições no mundo que indicam as ações possíveis dos organismos), então podemos dispensar o vocabulário representacionalista, tão comum nas teorias cognitivas e filosóficas. Pois, nesta perspectiva, as representações mentais seriam não apenas desnecessárias mas antieconômicas, uma vez que os estados disposicionais indicariam, por si só, diretamente, as possibilidades de ação aos organismos.

Em vez de procurar entender como as representações mentais se ajustam aos objetos no mundo, nos preocupamos em compreender **como** e **por que** os organismos escolhem *affordances* específicas presentes no sistema ao qual eles pertencem. Acreditamos que ao identificar tais escolhas, e o seu modo de realização, estaremos reunindo as marcas que caracterizam o processo de formação de diferentes identidades.

De um modo muito simplificado, podemos abordar esta questão investigando sistemas do tipo robô-meio ambiente como o *Roboser* (VERSHURE, 1992) que incorpora mecanismos de causalidade circular e de auto-organização. Este sistema consiste de um robô foto-sensível que se move em uma superfície plana impulsionado por fontes externas de luz. Tal como foi



projetado, *Roboser* tem um mecanismo de aversão a qualquer tipo de contato físico, de forma que, enquanto é atraído por uma fonte de luz, afasta-se dos objetos, sempre que existir qualquer possibilidade de colisão.

Especificadas as condições acima, quais seriam as possibilidades de "ação" para tal robô?

Ao analisarmos o comportamento do *Roboser*, podemos identificá-lo como um mecanismo de processamento de informação, que se locomove em superfícies planas, dirigindo-se aos focos luminosos e desviando-se dos obstáculos. Se uma fonte luminosa, anexada a um anteparo de vidro transparente, for colocada na frente do robô, qual será o seu comportamento? As leis de estruturação do sistema robô-meio ambiente permitem prever que ele se moverá em círculo até que uma das fontes do conflito seja eliminada. Tal mapeamento nos dá elementos para inferir a identidade de *Roboser* caracterizado como um mecanismo de processamento de informação que possui limites claramente estabelecidos.

Em função de tais dados, seria correto supor que o *Roboser* tem uma **identidade pessoal**? À primeira vista, a resposta é não, pois, de acordo com a noção comum de identidade pessoal não faz sentido qualquer discurso sobre identidade pessoal em robôs; eles são sistemas pré-programados, sem liberdade para escolher qualquer *affordance* disponível no mundo. Porém, se incluímos neles, como é o caso com o *Roboser*, mecanismos de causalidade circular e de auto-organização, então fica difícil dar uma resposta definitivamente negativa a esta pergunta. A dificuldade acontece, em parte, porque não temos uma idéia clara sobre as noções de auto-organização, autonomia e identidade pessoal e, por conseguinte, sobre sua aplicação em robôs ou mesmo nas ações dos organismos. No que se segue, faremos algumas especulações sobre esses conceitos.

a) Identidade pessoal, auto-organização e autonomia

O problema da identidade pessoal, tal como descrito atualmente, assume uma forma aparentemente paradoxal. Em estudo dedicado à aná-

lise da concepção humeana da identidade pessoal, GREEN (1999) observa que o problema consiste em explicar como é possível que indivíduos diferentes, vivendo em épocas distintas, possam ser idênticos a um outro indivíduo que também vive em época distinta (p. 103).

Como é possível, por exemplo, que o senhor X, que atualmente cuida de flores e ajuda carinhosamente crianças e idosos abandonados, seja o mesmo indivíduo que teve uma infância bem comportada e que hoje é acusado, com provas contundentes, de torturar e matar dezenas de pessoas durante um regime militar? Que identidade é esta que permite tamanha diversidade de caráter? Que indivíduos compõem o indivíduo X?

Uma resposta para esse problema é oferecida por alguns filósofos clássicos que, como DESCARTES (1978), postulam a existência de uma alma imortal, que dotaria o seu possuidor – exclusivamente humano – de uma consciência racional. Esta, por sua vez, permitiria o emprego de representações mentais, através das quais o sujeito representaria o mundo de acordo com as diferentes fases de sua existência. Tais representações manipuladas por um sujeito racional podem conter erros e contradições que serão dissolvidas e unificadas pela sua consciência. Esta realizaria, assim, o papel unificador das diferentes fases da vida do indivíduo, garantindo a sua identidade pessoal. As representações do sujeito, mas não ele próprio, comportam divergências, contradições, etc.

A dificuldade em se aceitar atualmente esta concepção metafísica do sujeito é que ela não pode ser cientificamente analisada: a existência de almas, exclusivamente humanas, e da razão como condição necessária para a existência da consciência parece estar longe do alcance da ciência.

Essa postura metafísica clássica foi amplamente investigada e criticada por vários filósofos. HUME (1962), por exemplo, critica a concepção cartesiana de alma imortal, cuja existência independe da experiência dos indivíduos, que trazem consigo, no nascimento, a marca do seu criador divino. Sem abandonar, entretanto, a noção de representação mental, Hume argumenta que a **idéia** de identidade pessoal surge dos hábitos que adquirimos na nossa prática de associação de idéias simples. Estas, sendo

originadas das sensações produzidas pelos estímulos que chegam aos nossos órgãos sensoriais, povoam o "palco vazio" da nossa existência, gerando a idéia, provisória, de uma identidade pessoal. Desta forma, a identidade pessoal parece ser, para Hume, apenas uma idéia complexa, entre outras.

No presente trabalho esboçamos uma hipótese alternativa tanto à concepção cartesiana quanto a humeana, no que se refere à identidade pessoal. Abandonando provisoriamente a noção de representação mental, tão cara aos filósofos cognitivistas, procuramos caracterizar a identidade pessoal como um processo de auto-organização que adquire relativa autonomia. Tal processo se estabeleceria, inicialmente, através de encontros aleatórios, resultantes do cruzamento de cadeias causais independentes, entre um organismo recém formado e o seu meio ambiente. Teríamos, nesta etapa, o início do processo que DEBRUN (1996) denominou de *auto-organização primária*. Nele, as tendências, disposições ou organizações se configuram e adquirem identidade sem um centro organizador absoluto ou qualquer programa pré-estabelecido. A medida em que tais disposições vão se cristalizando, e adquirindo certa autonomia, elas poderão dar lugar à geração e alteração de hábitos que, por sua vez, iniciarão o processo de identidade pessoal.

A habilidade de criar e mudar hábitos permite aos organismos agir de modo a favorecer a própria sobrevivência e a ajustar seu comportamento conforme as características ambientais; mudando o ambiente e sendo afetado pelas mudanças de acordo com a dinâmica da causalidade circular. Exemplos deste tipo de auto-organização podem ser vistos em organismos, como nós, capazes de aprender com os erros e absorver ruídos para aumentar o seu grau de complexidade. Na terminologia de DEBRUN (1996), tal habilidade caracterizaria os processos de *auto-organização secundária*.

Tanto na auto-organização primária quanto na secundária estabelecem-se, entre os elementos do processo, tendências, formas ou estados disposicionais sem a influência de um controle central. Contudo, algumas formas estáveis de organização, estabelecidas nos processos de auto-

unesp

organização secundária, adquirem a habilidade de criar e abandonar hábitos desenvolvendo uma relativa **autonomia** e comportando-se como um núcleo independente em relação ao meio circundante. Este estágio do processo de auto-organização secundária é descrito por DEBRUN (1996) como um processo de *cristalização* ou de fechamento. O sistema, assim estabelecido, necessitará de pouca informação externa para sobreviver; os seus elementos constituintes se bastam para a manutenção da ordem estabelecida. Neste estágio, várias possibilidades de desenvolvimento estariam disponíveis ao sistema, dependendo do seu grau de complexidade. Contudo, certos hábitos se instalam formando uma matriz relativamente isolada do seu meio. Tudo se passa como se o meio existisse apenas como um alimentador básico de uma estrutura que possui existência própria e autônoma. É justamente nesta fase que o sistema estará criando condições para o estabelecimento de uma identidade própria.

Em resumo, a hipótese central deste trabalho consiste em sugerir que, dadas as condições adequadas do meio ambiente, um processo de formação de identidade pode ocorrer em sistemas (não necessariamente humanos) que, adquirindo a habilidade de criar e mudar hábitos auto-organizadamente, passam a agir de maneira relativamente autônoma. No caso dos seres vivos, essa identidade não é puramente biológica ou fisiológica, mas é um produto de disposições específicas que são criadas pelas ações dos organismos no meio ambiente.

O resultado dessa interação do organismo com o meio não poderia ser previamente determinado, uma vez que ele resulta de uma dinâmica de cooperação entre elementos complexos que encontram-se em constante mutação, e cujo encontro traz sempre alguma forma de novidade. No caso dos seres humanos, essa interação, acontecendo sistematicamente entre processos ambientais, neurológicos, sociais e cognitivos, produz uma diversidade extremamente rica e complexa. Além disso, ela fornece a matriz na qual são estabelecidos hábitos no processo de formação de uma identidade flexível. Tal matriz pode ser vista como um processo constituído durante a vida de um indivíduo, não sendo definida apenas

unesp

geneticamente ou pelas condições do meio, uma vez que é produto do encontro de elementos heterogêneos. Apesar de ser relativamente fechada, esta matriz incorpora a possibilidade de mudanças de acordo com as condições estabelecidas em torno dos hábitos que a identificam.

De acordo com as hipóteses que desenvolvemos aqui, a identidade pessoal não se resume a uma idéia complexa (como sugere Hume) nem decorre de uma alma imortal (como sugere Descartes). Ela constitui um complexo processo de auto-organização que, com o tempo, adquire relativa estabilidade e autonomia na geração de hábitos, formas ou tendências. Ela *não pertence* aos indivíduos, como poderiam lhes pertencer as representações mentais. Os hábitos são elementos constitutivos dos indivíduos. Através dos hábitos podemos entender os conflitos do Senhor X que, possuidor de uma memória, registra dados sobre o seu passado que não se harmonizam com os seus estados disposicionais atuais.

b) O papel dos hábitos no processo de formação de identidade

Mencionamos na seção anterior que alguns filósofos clássicos abordam o tema da identidade em termos da noção de sujeito metafísico ou transcendental, dotado de uma razão única, que distinguiria o gênero humano dos demais organismos. Até o momento rejeitamos essa concepção ao considerar a formação da identidade como um processo cuja principal característica é a capacidade de auto-organização – não exclusiva dos seres humanos – para criar e alterar hábitos, seguida do estabelecimento de certa autonomia. Admitimos, inclusive, a possibilidade de o *Roboser* poder adquirir uma identidade, contanto que as condições ambientais adequadas pudessem ser obtidas e que ele conseguisse operar corretamente com processos de auto-organização e de causalidade circular. A pergunta natural que surge aqui é: tal concepção, aplicada ao contexto humano, não reduziria os indivíduos ao estado de autômato, sem liberdade e, por conseguinte, sem nenhuma identidade real?

Na história da filosofia, numerosos debates têm sido registrados sobre esta pergunta. Assim, por exemplo, Pascal (1963) argumenta que



se, por um lado, a ênfase em hábitos parece impedir o exercício da liberdade individual, por outro lado, a concepção de Descartes – que relaciona a identidade a uma alma essencial e imortal, fonte de liberdade humana – deveria ser abandonada. Pascal, crítico veemente do pensamento cartesiano, procura mostrar a importância da criação e manutenção de hábitos no comportamento individual e coletivo. Ele sustenta que o principal componente do agir humano é a memória corporal de um hábito bem sucedido para a manutenção da vida individual e a garantia da paz na ordenação social.

Se é possível o exercício da liberdade, argumenta Pascal, esse exercício deve, por princípio, ser coletivo, na medida em que os costumes sociais somente adquirem esse estatuto quando pertencem a toda a sociedade ou, pelo menos, à maioria de seus membros. No entanto, o costume coletivo surge enquanto tal pela repetição constante de formas particulares de agir e pelo reconhecimento implícito das vantagens trazidas por essas práticas individuais.

A liberdade consiste, para Pascal, em poder criar novos hábitos que, gradativamente, vão sendo incorporados pelos demais indivíduos; hábitos que, pela filtragem temporal, vão mostrando sua excelência em relação aos hábitos anteriores, até serem incorporados em nosso modo de agir espontâneo ou serem abandonados pela assimilação de um novo costume. Para ele, há tão somente costumes legitimados por uma somatória de experiências bem sucedidas.

Com efeito, Pascal aponta no fragmento 125 da Edição Lafuma dos *Pensamentos*:

> Que são nossos princípios naturais, senão princípios de hábitos? E nas crianças, os que receberam com os hábitos dos pais, como a caça entre os animais? Hábitos diferentes dão-nos princípios naturais diversos, é o que nos prova a experiência e, se existem princípios que o hábito não pode fazer desaparecer, há-os também do costume contra a natureza, inapagáveis por esta, ou por um segundo costume. Tudo depende da disposição.

unesp

Também no fragmento 126 da mesma obra Pascal aponta a força dos costumes ao afirmar:

> Temem os pais que o amor natural de seus filhos se extinga. Que espécie de natureza será essa, então, suscetível de extinção? O hábito é uma segunda natureza que destrói a primeira. Mas que é a natureza? Por que não é o hábito natural? Receio muito que essa natureza não seja ela própria senão um primeiro hábito, assim como o hábito uma segunda natureza".

Assim é inadmissível para Pascal a concepção de que há uma "essência humana" anterior a qualquer interação com o mundo e que, supostamente, regularia a ação individual. Tudo o que os indivíduos recebem é um conjunto de disposições historicamente geradas. As escolhas particulares estabelecidas pelos agentes marcariam sua individuação.

A lição deixada por Pascal é que se o ser humano parece apenas repetir mecanicamente hábitos biológica e historicamente predeterminados, tal repetição mecânica, entendida aqui no sentido literal, perde a conotação pejorativa. Pois, a atividade humana consiste na repetição das formas de agir que se mostraram mais adequadas para a preservação da vida e o abandono das que se mostraram inadequadas para o mesmo fim, com sua posterior substituição por novos hábitos. Afinal, é através da repetição que são preservadas certas formas de agir que permitem a sobrevivência individual e a preservação da melhor forma de organização social.

Mas, neste caso, como podemos assegurar minimamente a liberdade como possível? A liberdade, entendemos, pode ser concebida como a habilidade para abandonar aqueles hábitos que se mostraram inadequados, seguida pela habilidade de criar hábitos novos e mais satisfatórios, através dos processos de auto-organização. Indivíduos criativos cumprem bem esta tarefa, enquanto os menos criativos seriam resistentes a mudanças, necessitando sempre de receitas pré-estabelecidas para a sua ação. Além disso, a criação de um novo hábito na interação do indivíduo com o meio só é possível desde que haja a percepção dos **limites**, não apenas dos hábitos abandonados, como também daqueles hábitos, por definição

temporários, que substituirão os antigos. Essa percepção de limites na interação dos sistemas (processadores de informação) com o meio parece ser possível apenas aos indivíduos possuidores de uma história evolutiva, isto é, apenas aos seres vivos. Nesse sentido, o *Roboser* não é livre e não possui uma identidade própria.

Referências Bibliográficas

ASBHY, W. (1962) Principles of the self-organising system. In VON FOESTER, H., ZOPF, G.W. (eds.). *Principles of self-organisation.* London : Pergamon, p.225-278.

DEBRUN, M. (1996) A idéia de auto-organização. In: DEBRUN, M., GONZALES, M.E.Q., PESSOA JR. O. (orgs.). *Auto-organização*: estudos interdisciplinares. Campinas : UNICAMP/CLE, p. 03-23. (Coleção CLE, v. 18.)

DESCARTES, R. (1978) *Méditations méthaphysiques.* Paris : J. Vrin.

GIBSON, J.J. (1979) *The ecological approach to visual perception.* Boston : Houghton Mifflin.

GONZALES, M.E.Q., ALVES, E.H., SETTE, A.M. (1996) Self-organisation in the development of representation action couplings. *Studies in Perception and Action*, v. III. (Eightieth International Conference on Perception and Action.)

GREEN, M. (1999) The idea of a momentary self and Hume's Theory of Personal Identity. *British Journal for the History of Philosophy,* v. 7, n.1, p. 103-122.

HUME, D. (1962) *On human nature and the understanding.* London : Collier Macmillan Publishers.

PASCAL, B. (1963) *Oeuvres complètes.* Paris : Éditions du Seuil.

TURVEY, M.T. (1988) Notas de aulas ministradas na University of Connecticut.

VERSHURE, P. (1992) Taking connectionism seriously: The vague promise of subsymbolism and an alternative. *Proceedings of the 14th Annual Conference of the Cognitive Science Society*. Hillsdale (EUA) : Erlbaum, p. 653-658.

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

Conhecimento comum e auto-organização[1]

Willem Ferdinand Gerardus Haselager
w.haselager@nici.kun.nl
www.nici.kun.nl/~haselag
Dpt. of Artificial Intelligence/Cognitive Science,
Nijmegen Institute for Cognition and Information, University of Nijmegen, The Netherlands
Post-Graduate Program in Cognitive science and Philosophy of Mind, Philosophy Department, UNESP, Brazil
&
Maria Eunice Quilici Gonzalez
gonzalez@marilia.unesp.br
Pós-graduação na area de Ciência Cognitiva e Filosofia da Mente,
Departamento de Filosofia, UNESP,
Av. Hygino Muzzi Filho 737, Marília, SP, 17525-900, Brasil

Abstract
According to traditional cognitive science, our common sense behavior is the consequence of reasoning and planning processes on the basis of common sense knowledge. Specifically, common sense knowledge (our beliefs and desires) is taken to consist of propositional attitudes. We will argue that the traditional view of everyday behavior as caused by propositional attitudes provides an artificial impression of the nature of our everyday behavior. The root of this problem lies within classical epistemology, with its attempt to analyze knowledge in terms of justified true belief.
We propose an alternative, systemic, view of common sense knowledge that is based on the theory of self-organization and embodied embedded cognition. Common sense knowledge should not to be conceived as beliefs *about* the environment, but rather as dispositions to interact *with* the environment. These dispositions are the result of self-organizing dynamic couplings between body and world and are incorporated instead of represented. Combined with a pragmatist approach to epistemology, this systemic conception of common sense knowledge may provide a proper perspective on the nature of our everyday behavior.

Resumo
De acordo com Ciência Cognitiva tradicional, na esteira das concepções da psicologia popular, nosso comportamento é uma conseqüência de processos racionais e planejados com base no conhecimento comum. Especificamente, considera-se que o conhecimento comum (que reúne crenças e desejos) pode ser explicado a partir da relação causal que se estabelece entre atitudes proposicionais. Procuraremos mostrar que a perspectiva tradicional sobre o comportamento cotidiano, como sendo causado por atitudes proposicionais, gera uma concepção artificial sobre a natureza do comportamento comum. A raiz deste problema se encontra na epistemologia clássica, uma vez que ela concebe o conhecimento em termos de crença verdadeira justificada.
Neste trabalho propomos uma perspectiva alternativa, sistêmica, do conhecimento do senso comum apoiada na Teoria da Auto-Organização e na Teoria da Cognição Incorporada e Situada. Argumentamos que o conhecimento do senso comum não pode ser propriamente concebido como crença verdadeira justificada *a respeito* do meio ambiente, mas, sim, como disposições para

---

[1] Este artigo foi aceito para publicação em Coletânea de Filosofia da Mente organizada por Eduardo Rabossi em *La mente y sus problemas: temas actuales de Filosofia de la Psicologia.* Buenos Aires: Catálogos Editora. Uma investigação detalhada do problema do Teeteto, enunciado neste artigo, é apresentada no primeiro capítulo desta Coletânea.



interagir no meio ambiente. Tais disposições resultam de acoplamentos dinâmicos e auto-organizados entre o corpo e o mundo, sendo, antes de tudo *incorporados,* o invés de *representados*. Combinada com uma epistemologia próxima ao Pragmatismo, esta concepção sistêmica do conhecimento do senso comum pode fornecer uma perspectiva apropriada sobre a natureza de nosso comportamento cotidiano.

Introdução

Herdeira da tradição Platônica, a ciência cognitiva tradicional procura caracterizar o conhecimento através de estruturas simbólicas abstratas que são manipuladas por mecanismos de recepção passiva de informação. Como no mito platônico da caverna - que caracteriza os indivíduos como seres acorrentados, receptores passivos de sombras que lhes chegam através de imagens luminosas projetadas na parede da caverna – os cognitivistas tradicionais elaboram modelos computacionais do processo cognitivo. Tais modelos simulam (ou mesmo reproduzem) o comportamento, supostamente inteligente, de sistemas de processamento de informação que, operando com representações simbólicas, respondem à estímulos formalmente estruturados.

Nesse cenário, conforme procuraremos mostrar, a ciência cognitiva tradicional, influenciada pela Epistemologia clássica, caracteriza o conhecimento comum em termos de atitudes proposicionais resultantes de planejamentos racionais estruturados a partir de regras abstratas aplicadas a manipulação de símbolos. De acordo com essa concepção, nosso comportamento é freqüentemente entendido como uma conseqüência de nossas crenças e desejos. Assim, por exemplo, se acreditamos que a Argentina é um país muito bonito, e se desejamos estar em um pais bonito, é muito provável que procuraremos viajar para a Argentina, se as circunstancias assim permitirem. Nossas crenças e desejos causam nosso comportamento. Podemos explicar nosso comportamento indicando nossos desejos e crenças relevantes, e outras pessoas podem prever, até certo ponto, nosso comportamento se elas conhecerem nossos desejos e crenças. No que se segue argumentaremos que tal caracterização do conhecimento comum é inapropriada e que ela gera mais problemas do que soluções

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

A 'Psicologia popular'

Na filosofia e na ciência cognitiva, a concepção acima esboçada é conhecida como psicologia popular ('folk psychology'). A idéia central aqui é que conduzimos nossa vida cotidiana, da forma que o fazemos, através de crenças e desejos (esperanças, medos, etc.) que moldam nosso comportamento. Gastamos um tempo enorme falando com os outros sobre nossos (e de outros) estados mentais.

De acordo com alguns teóricos (principalmente, Fodor, Stich e Churchland), as pessoas aprendem a psicologia popular muito cedo (a partir dos quatro anos, Perner, 1991) na forma de uma teoria. Isto é, elas aprendem que, quando alguém quer X (digamos um brinquedo) então aquela pessoa geralmente procurará obter X. As pessoas usam esse corpo de conhecimento da psicologia popular para prever e explicar o comportamento dos outros. Esta concepção é algumas vezes conhecida como 'a teoria da teoria': trata-se da teoria segundo a qual a psicologia comum é uma teoria que é aprendida em torno de quatro anos de idade.

Uma concepção alternativa é dada pela teoria da simulação; a idéia básica aqui é que as pessoas se imaginam no lugar de outro (imaginam que eles possuem as crenças e desejos de outros)/ imaginam o que elas fariam em tais circunstancias, e utilizam suas próprias decisões para prever o comportamento da outra pessoa. Por exemplo, imaginamos que estamos na situação de Antônio, e que, como Antônio, pensamos que se P então Q; se, novamente como Antônio, desejássemos Q, faríamos P. Em tais circunstâncias, podemos prever que Antônio também fará P. Apesar das diferenças entre a *teoria da teoria* e a teoria da simulação, investigações empíricas sobre a natureza da psicologia popular têm nos deixado com mais perguntas do que respostas. Resultados recentes de investigações empíricas indicam que algumas vezes empregamos *a teoria da teoria* e outras vezes empregamos a *teoria da simulação* para prever e explicar o comportamento dos outros (Stone & Davies, 1995; Perner & Kühberger, no prelo, p. 5).

Conhecimento comum: atitudes proposicionais

Qual é a característica comum das duas teorias? Ambas caracterizam os estados mentais (como crenças e desejos) como atitudes proposicionais. Se acreditamos que a Argentina é um país bonito, nos encontramos, supostamente, em uma relação computacional com uma proposição que

unesp

instancia o conteúdo: 'Argentina é um país bonito'. Podemos ter diferentes atitudes em relação a tal proposição: posso acreditar, duvidar, ter esperanças, e assim por diante.

Queremos focalizar aqui uma hipótese muito popular na ciência cognitiva tradicional, segundo a qual as nossas crenças e desejos devem ser caracterizados em termos de proposições em relação as quais assumimos certas atitudes. Esta é uma visão do senso comum que está muito próxima da concepção segundo a qual o conhecimento é essencialmente proposicional. Conforme detalharemos adiante, a Epistemologia clássica estabelece, ou procura estabelecer, uma nítida distinção entre conhecimento e crença. De acordo com a tradição filosófica, o conhecimento possui algo mais que as meras crenças. Mas em nenhum momento o caráter proposicional de ambas é colocado em dúvida. Esta pressuposição nos parece problemática porque ela nos induz a uma concepção do senso comum entendido como um conjunto de descrições sobre o mundo. Além disso, ela conduz a uma concepção do comportamento comum como sendo guiado por conhecimento descritivo sobre o mundo, a partir do qual raciocinamos.

Segundo essa concepção, mesmo nas situações mais mundanas, estamos constantemente engajados na atividade de planejar, raciocinar e tomar decisões. Como argumentaremos abaixo, esta concepção é inadequada como uma caracterização das causas de nosso comportamento comum. Ela nos conduz a uma concepção equivocada a respeito de nossas atividades cotidianas.

Senso comum ou senso artificial?

A concepção vigente na ciência cognitiva tradicional, segundo a qual nosso comportamento comum é uma conseqüência das inferências formuladas com base em atitudes proposicionais, tem conduzido a sérios problemas. Para mencionar um exemplo:

> "Consideremos um exemplo bem conhecido de problema cotidiano: que conhecimento necessitamos para formular um simples plano para quebrar um ovo nas bordas de uma vasilha, com a intenção de colocar o conteúdo do ovo na vasilha". (Elio, 2002, p.11).

A Inteligência Artificial (IA) tradicional tem tentado resolver problemas como este há muito tempo. Desde o momento em que McCarthy & Hayes (1969, p.463) tentaram criar um programa computacional capaz de 'agir inteligentemente no mundo', desde então a IA tem se deparado com a dificuldade de modelar o conhecimento comum e o raciocínio cotidiano. Até o momento, a IA

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

não foi bem sucedida na produção de qualquer máquina que se aproxime de algo que atue inteligentemente no mundo.

Esta falha da IA tradicional em obter tal meta tem sido discutida intensivamente em relação ao conhecido problema dos *frames* ('the Frame Problem'). Uma parte importante deste problema é que as pessoas possuem uma quantidade surpreendente de conhecimento comum. Não apenas é extremamente difícil de formular *o que* sabemos (veja, por exemplo, o projeto de Cyc: Guha & Lenat, 1990; 1993), como também é ainda mais problemático modelar nosso uso eficiente e flexível das partes relevantes do nosso conhecimento (Haselager, 1997; 1999; Haselager & van Rappard, 1998).

Entendemos que a raiz da dificuldade reside na própria formulação do problema da natureza do conhecimento cotidiano. É a própria suposição que o conhecimento comum deve ser caracterizado em termos de atitudes proposicionais, e que nosso comportamento é uma conseqüência de raciocínio sobre proposições, que conduz a dificuldades. Para caracterizar nossa posição em relação ao exemplo acima mencionado, da quebra do ovo, julgamos que em muitas situações do dia a dia não formulamos planos e o conhecimento que desempenha um papel relevante na nossa conduta não é de um tipo proposicional. Conforme procuraremos argumentar, as suposições da ciência cognitiva possuem raízes na epistemologia clássica que é essencialmente platônica.

Conhecimento e opinião verdadeira: a Epistemologia clássica

Um dos problemas mais discutidos na Epistemologia clássica diz respeito á distinção entre conhecimento e opinião verdadeira. Tal problema foi originalmente formulado por Platão que, no diálogo *Teeteto*, considera várias situações em que um indivíduo pode possuir crenças verdadeiras acerca de eventos no mundo, mas não possuir conhecimento sobre tais eventos. Assim, por exemplo, Wilma pode acreditar que choverá, em uma data e hora específicas, sem possuir qualquer conhecimento sobre meteorologia. Ainda que sua crença se mostre verdadeira e que, de fato, chova na data e hora previstas, ainda assim, dada sua ignorância sobre meteorologia, não se considera que ela possua conhecimento sobre a ocorrência da chuva naquele momento. O problema do Teeteto se coloca quando perguntamos : o que é que um

unesp

indivíduo, dotado de crenças verdadeiras sobre um evento, deve possuir além dessas crenças para que possamos considerá-lo um conhecedor da ocorrência de tal evento?

Várias propostas de solução para o problema do Teeteto vem sendo discutidas entre os filósofos por muitos séculos sem uma solução que seja, consensualmente, considerada satisfatória. Contudo, a grande maioria dos filósofos admite que a solução para o problema do Teeteto deve partir da suposição que o conhecimento envolve um procedimento de justificação racional de crenças verdadeiras. Nesse contexto, o problema do conhecimento se desloca para o problema da *justificação* racional: em que condições um sujeito, possuidor .de crenças verdadeiras, estaria apto para justificá-las?

Uma dificuldade central do problema do conhecimento, tal como caracterizado na tradição platônica, diz respeito à circularidade que parece estar envolvida na escolha dos candidatos que poderiam fundamentar o processo de justificação racional de crenças verdadeiras. Assim, por exemplo, processos de justificação de crenças que se sustentam em verificações empíricas envolvem problemas de circularidade, uma vez que a escolha dos dados observacionais relevantes, que supostamente fundamentariam uma verificação empírica, requerem conhecimento; precisamente o que estamos procurando definir. O mesmo pode ser dito de outros processos de justificação que envolvem, por exemplo, elementos estatísticos, éticos, etc. Uma vez que não é qualquer dado estatístico ou elemento ético que pode ser qualificado como um candidato apropriado para justificar uma crença verdadeira, eles exigem um critério apropriado de seleção, cuja escolha parece pressupor o conhecimento que procuramos definir.

Uma tentativa recente de resolução do problema do Teeteto foi proposta por Dretske no seu livro *Knowledge and the flow of information* (1981). Nessa obra, Dretske investiga o conceito de *informação* argumentando que um estudo cuidadoso de tal conceito nos indica uma saída para a circularidade presente nos processos de justificação racional do conhecimento. Grosso modo, a proposta de Dretske envolve duas etapas. Na primeira, apoiando-se na teoria matemática da comunicação desenvolvida por Shannon e Weaver (1948), ele fornece uma caracterização naturalista da informação: trata-se de um elemento objetivo, cuja existência independe de mentes conscientes ou de qualquer fator subjetivo. O mundo físico é considerado

unesp

por Dretske como estando repleto de relações legiformes que se estabelecem entre conjuntos, padrões ou estruturas de naturezas diversas. A informação é justamente aquele indicador de relações que pode ser objetivamente registrado por um receptor (humano ou não) que esteja devidamente equipado para tal.

O caráter legiforme e objetivo da informação, tal como caracterizada acima, indica , no entender de Dretske, um forte candidato para resolver o problema do Teeteto. Tendo caracterizado a informação de maneira naturalista, Dretske propõe, na segunda etapa de sua argumentação, uma definição do conhecimento entendido como " ...crença verdadeira sustentada em *informação"* Uma vez que sua concepção de informação não pressupõe a existência de conhecimento (e nem mesmo de uma consciência), e dado o caráter legiforme da informação que assegura a sua objetividade, ele argumenta que temos, finalmente, uma definição do conhecimento que não envolve circularidade.

Não é nossa intenção debater aqui a validade e o alcance da proposta de Dretske.

Nosso principal objetivo consiste em indicar a natureza da tradição platônica na epistemologia e suas conseqüências problemáticas para a pesquisa na I.A. no que diz respeito ao comportamento e ao conhecimento comum. No que se segue, examinaremos uma visão alternativa do senso comum, que se sustenta na Teoria da Cognição Incorporada e Situada e na Teoria da Auto-Organização.

Cognição situada e incorporada e a Teoria da Auto- Organização

No mito platônico da caverna, os observadores são forçados a permanecer passivos. Suas crenças sobre o mundo não podem ser colocadas em prática. Contudo, o senso comum está diretamente ligado à prática e à ação. Supostamente, o senso comum é o que nos habilita a atuar como atuamos à medida que interagimos cotidianamente com aquilo que nos circunda. Nessa perspectiva, rejeitamos a idéia segundo a qual o conhecimento do senso comum (crenças) deve ser interpretado como descrições (proposicionais) do mundo, formuladas a partir de uma perspectiva de observadores passivos. Em contraste, propomos um esboço de uma interpretação do senso comum fundada em uma perspectiva de indivíduos ativos. Nossa sugestão básica é que

unesp

o conhecimento do senso comum deve ser concebido como inteiramente incorporado e situado na natureza, consistindo de padrões potenciais de interação dos indivíduos com o mundo. Esta concepção ajusta-se aos desenvolvimentos recentes na área de ciência cognitiva, algumas vezes indicada pelo nome de 'Cognição Enativa', 'Cognição incorporada' ou 'Cognição incorporada e situada' (CIS) e também a Teoria da Auto-Organização (TAO). A TAO e a CIS ressaltam que a cognição desempenha um papel adaptativo e flexível em nossas atividades no mundo. Algumas noções da CIS e TAO, centrais para nosso objetivo, dizem respeito ao mundo, ao corpo, ao cérebro, à interação entre organismo e ambiente, à temporalidade e aos processos de auto-organização. Estas noções são brevemente descritas a seguir:

(1) O mundo: sistemas cognitivos estão situados no meio ambiente e seus comportamentos estão, essencialmente, imersos no mundo. Ao invés de conceber o mundo como um componente separado e distinto dos processos cognitivos, o mundo é considerado como parte deles.

(2) O corpo: sistemas cognitivos são essencialmente sistemas incorporados, com dinâmicas corporais intrínsecas (isto é, tendências de coordenação espontâneas, Kelso, 1995) contribuindo para o comportamento do sistema e, de acordo com Damásio (1994), para a racionalidade. A percepção e a ação são consideradas como uma unidade funcional ao invés de componentes separados,

Uma boa ilustração de como a cognição e o comportamento dependem de especificidades do corpo e do meio ambiente é fornecida por Thelen & Smith (1994). Bebês que já aprenderam a engatinhar (aos oito meses e meio de idade) não hesitam ao descer uma rampa acentuada de 10 a 40 graus. Após cair algumas vezes, eles param de engatinhar na rampa: eles aprendem os riscos presentes nessa atividade. No momento em que começam a andar (de 12 a 14 meses), eles têm que aprender tudo novamente. Na mesma situação, 10 de 15 crianças descem rampa abaixo sem hesitação. Algumas crianças quando colocadas em frente da rampa, na posição de engatinhamento, não descem por ela, mas quando colocadas em pé, descem "como se tivessem se tornado instantaneamente ignorantes da existência prévia da rampa" (Thelen & Smith,1994, p. 220).

unesp

(3) O cérebro: o cérebro não é concebido como 'espelho' do mundo, como um sistema primariamente movido por representações e processamento de informação, mas sim como um dispositivo de controle da ação (ver, por exemplo, em VARELA et al., 1991, a noção de 'cognição enativa').

(4) Interação e pareamento: em todos os planos de descrição o comportamento pode ser entendido como emergindo de interações contínuas e não lineares entre organismo e ambiente.

(5) O tempo: os processos cognitivos se desenvolvem (*unfold*) em tempo real; a necessidade de agir e reagir no tempo funciona como pressão da seleção operando nos organismos bem como em seus padrões comportamentais.

(6) Auto-Organização: sistemas complexos reúnem um grande número de componentes interagindo (por exemplo, células, tecidos, sistemas de órgãos, subsistemas funcionalmente definidos para o movimento, a percepção, etc) entre si de tal forma que todos contribuem para o seu comportamento global. Devido à interação não linear dos componentes de tais sistemas, padrões auto-organizados de comportamento podem emergir.

A perspectiva básica que a TAO e a CIS fornecem é que muito de nosso comportamento cotidiano emerge em tempo real a partir da junção dinâmica entre corpo, cérebro e meio ambiente. A CIS e a TAO enfatizam as possibilidades de interação do organismo com o meio ambiente. Cada situação fornece certas possibilidades de ação e torna outras mais difíceis. Cabe ressaltar que existe uma conexão muito próxima entre esta concepção e a noção gibsoniana de *affordance:* "behavioral opportunities in a niche for an organism, relative to coordinated action and perceived by the organism-in-action." (Clancey, 1997, p.251-258).

A partir dessa perspectiva, entendemos que é mais frutífero interpretar um organismo como um 'repertório móvel de comportamentos" do que como um 'observador passivo acorrentado'. Nosso conhecimento do senso comum é incorporado e situado, ele está em nossas ações e não em nossas descrições do que conhecemos sobre o mundo. Se olharmos para o exemplo da 'quebra do ovo', acima mencionado, o problema de como quebrar um ovo não se resume a uma consulta na base de dados de nosso conhecimento proposicional para formular um planejamento simples. Ao invés disso, o meio ambiente e suas propriedades possibilitam, selecionam ou explicitam certos

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

comportamentos. A borda da vasilha permite a quebra da casca do ovo: o movimento que fazemos com nossos braços e mãos não é planejado e calculado, mas emerge da dinâmica intrínseca de nosso corpo, do peso do ovo, da influencia da gravidade e de muitos outros fatores. Nesse sentido, os aspectos gerais do organismo na situação do senso comum se auto-organizam em uma seqüência apropriada de comportamentos emergentes. Crenças e desejos, estão, então, intrinsecamente, e não inferencialmente, relacionados ao comportamento. Não existe, nessa perspectiva, a necessidade de uma exploração exaustiva de uma imensa base de dados do senso comum para escolher as partes relevantes de modo similar a 'um planejamento comportamental simples'. O conhecimento do senso comum não é descritivo e nem proposicional; tampouco se encontra estocado em alguma base de dados. Ao contrário, o conhecimento está nos padrões comportamentais que um organismo tem a sua disposição.

Conclusão

Como apontamos, uma das principais questões da epistemologia consiste em estabelecer a diferença entre conhecimento e crença. A interpretação clássica considera que o processo de obtenção do conhecimento começa com crenças, a partir das quais se inicia a busca pelas condições que devem ser adicionadas à crença (notadamente justificação racional) de modo a qualificá-la como conhecimento. A natureza da crença é tomada como um ponto de partida relativamente não problemático para análise posterior do conceito de conhecimento. Contudo, sugerimos que graças à importância que a epistemologia atribuiu às noções de justificação e de verdade, a interpretação da crença foi construída de uma maneira particular. E é a própria interpretação da noção de crença - entendida como uma atitude proposicional - que tem como conseqüência o interesse da Epistemologia nos processos de justificação e de legitimação. Mais especificamente, é a ênfase que se coloca na noção de verdade em relação ao conhecimento que conduz à interpretação da crença como sendo proposicional por natureza. Assim, a concepção das crenças como atitudes proposicionais não é considerada um ponto de partida passível de discussão para o desenvolvimento da pesquisa sobre a natureza do conhecimento, mas, antes, uma conseqüência de uma pesquisa epistemológica.

A partir da idéia de que crenças são atitudes proposicionais, necessitamos um pequeno passo para sugerir as operações inferenciais como meios de se combinar representações portadoras de

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

conteúdo proposicional das crenças. Nesse cenário, o senso comum se torna 'intelectualizado, 'racionalizado' e 'computacionalizado'. Com base nessa concepção, a IA tentou modelar nossa capacidade cotidiana de agir inteligentemente no mundo, sustentada em planejamento e raciocínio. O fracasso da IA para atingir suas metas pode servir como uma indicação da inadequação da suposição de que as crenças possuem um caráter intrinsecamente proposicional.

Neste trabalho procuramos esboçar os contornos de uma concepção alternativa do senso comum baseada nas idéias de cognição situada e incorporada e nos pressupostos da teoria da auto-organização. De acordo com esta perspectiva, os organismos são concebidos como agentes situados no mundo ao invés de sujeitos passivamente observando e pensando sobre o mundo. Em relação ao conhecimento do senso comum, sugerimos, então, que este se caracteriza em termos de relações que se estabelecem entre o agente e o mundo que torna possível (affords) suas ações.

Referências

Elio, R. (2002). *Common sense, reasoning & rationality.* Oxford: Oxford University Press.

Clancey, W.J. (1997). *Situated cognition: On human knowledge and computer representation.* Cambridge: Cambridge University Press.

Damasio, A.R. (1994). *Descartes´ error: Emotion, reason, and the human brain.* New York: Avon Books.

Davies, M. & Stone, T. (1995). (Eds.). *Mental simulation.* Oxford: Blackwell.

Dretske, F. (1981). *Knowledge and the flow of information.* Cambridge, MA: MIT Press.

Guha, R.V. & Lenat, D.B. (1990). Cyc: a midterm report. *AI Magazine,* 32-59.

Guha, R.V. & Lenat, D.B. (1993). Re: Cycling paper reviews. *Artificial Intelligence (61),* 149-174.

Haselager, W.F.G. (1997). *Cognitive science and folk psychology: The right frame of mind.* London: Sage.

Haselager W.F.G. (1999). On the potential of non-classical constituency. *Acta Analytica, 22,* (pp. 23-42).

Haselager W.F.G. & van Rappard J.F.H. (1998). Connectionism, systematicity, and the frame problem. *Minds and Machines, 8*(2), (pp. 161-179).

Kelso, J.A.S. (1995). *Dynamic patterns: The self-organization of brain and behavior.* Cambridge, Massachusetts: MIT-Press.

McCarthy, J. & Hayes, P.J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In D. Michie & B. Meltzer (Eds.). *Machine Intelligence 4*. 463-502.

Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: Bradford Books/ MIT-Press.

Perner, J. & Kühberger, A. (forthcoming). Framing and the theory-simulation controversy: Predicting people's decisions. *Mind & Society* (Special Issue of contributions to ESF workshop "Natural Bases of Human Inference").

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18

Shannon, C.E. & Weaver W. (1949). *The mathematical theory of communication.* Urbana: University of Illinois Press.

Thelen, E., & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, Massachusetts: MIT-Press.

Varela, F.J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT press.

Agradecimentos

Agradecemos aos nossos colegas do Grupo de Estudos Cognitivos da UNESP (especificamente a Maria Cândida Del-Masso, Carmem Beatriz Milidoni e Mariana Claudia Broens), bem como as agências de fomento FUNDUNESP, FAPESP e NICI, Holanda, por financiarem esta pesquisa.

cm 1 2 3 4 5 6 7 unesp 1 12 13 14 15 16 17 18