

REDES NEURAIS E REPRESENTAÇÃO MENTAL: UM ENSAIO SOBRE HARMONIA E RACIONALIDADE

Maria Eunice Quilici GONZALES*

RESUMO: Investigaremos, a partir da perspectiva da Ciência Cognitiva, a noção de representação mental, no domínio da percepção visual humana. Ênfase é dada ao paradigma Conexionista, ou de Redes Neurais, de acordo com o qual tais representações mentais são descritas como estruturas emergentes da interação entre sistemas de processamento de informação que se auto-organizam – tais como o cérebro – e a luz estruturada no meio ambiente. Sugerimos que essa noção de representação mental indica uma solução para uma antiga polêmica, entre Representacionistas e Eliminativistas, acerca da existência de representações mentais no sistema perceptual humano.

UNITERMOS: Percepção visual; conexionismo; redes neurais; ciência cognitiva; padrões de informação; representação mental; auto-organização.

*“Só a natureza é divina, e ela não é divina...
Se falo dela como de um ente
É que para falar dela preciso usar da língua dos homens
Que dá personalidade às cousas,
E impõe nome as cousas.
Mas as cousas não têm nome nem personalidade:
Existem, e o céu é grande e a terra larga,
E o nosso coração do tamanho de um punho fechado...”*
(Alberto Caeiro)

INTRODUÇÃO

O que significa *ver*? Com esta questão David Marr introduziu seu artigo *Vision: Philosophy and the Approach* e, adotando a concepção do senso comum, sugeriu que *ver* é saber o que é algo, e onde este algo se encontra, através do olhar. A visão, diz ele, é o processo de descobrir, por meio de *imagens*, aquilo que está presente no mundo, e onde esse elemento se encontra (15, p. 103).

Embora discordemos em vários aspectos do trabalho de Marr, concordamos com seu ponto de partida, segundo o qual a visão é, entre outras coisas, um processo de

descoberta. Esse processo de descoberta envolve uma atividade de “representação” interna do fluxo de informação que caracteriza objetos e eventos no meio ambiente. Além disso, o processo de descoberta envolvido na visão possui *intencionalidade*, isto é, ele é sempre direcionado em relação a um objeto, evento ou cena particular. Essa intencionalidade caracteriza a percepção visual humana como um processo cognitivo que envolve expectativas, crenças e outras atitudes representacionais meta-direcionadas.

Até agora, nada de novo foi dito aqui pois a hipótese segundo a qual a percepção é um processo cognitivo, que envolve representação mental, vem sendo, há séculos, objetos de discussão entre filósofos e psicólogos (Cf. 7, 10, 11, 12, 15, 19, etc.). A contribuição que acreditamos trazer com a nossa análise diz respeito a natureza das representações mentais, que se supõem presentes na percepção visual.

Embora a noção de representação mental também seja objeto de investigação de filósofos, psicólogos, e outros, há muito tempo, a tentativa de tratar esse assunto a partir da perspectiva da Ciência Cognitiva, como faremos aqui, é relativamente nova.

Dado o caráter interdisciplinar da Ciência Cognitiva, problemas de interesse comum a diferentes áreas podem ser abordados a partir de perspectivas diversas. Esse caráter interdisciplinar da Ciência Cognitiva proporciona um horizonte amplo e rico de informações para o estudo de problemas, tais como o problema da natureza das representações mentais, que muitas vezes se mostram sem perspectiva de solução nos seus tradicionais âmbitos de abordagem.

De acordo com essa perspectiva, investigaremos a natureza das representações mentais em quatro etapas. Na primeira, *Secção 1*, faremos uma breve introdução à Ciência Cognitiva, de modo a situar no seu interior a nossa investigação. Na *Secção 2*, apresentaremos a polêmica entre os chamados *Representacionistas*, que postulam a existência de estruturas representacionais *abstratas* no nosso sistema cognitivo-perceptual (8, p. 7), e os *Eliminativistas*, que negam a existência de tais representações. Uma solução para essa polêmica é sugerida, na *Secção 3*, onde a natureza das representações é analisada a partir de um ramo da Ciência Cognitiva denominado *Conexionismo* (Connectionism), ou *Redes Neurais* (Neural Networks).

Finalmente, concluiremos com algumas especulações acerca da relação existente entre as noções de *representação mental*, *harmonia* e *racionalidade*, no paradigma de redes neurais.

1. CIÊNCIA COGNITIVA E O PROBLEMA DAS REPRESENTAÇÕES MENTAIS

A Ciência Cognitiva é uma área de investigação interdisciplinar que existe há mais ou menos trinta anos. Embora o termo *Ciência Cognitiva* tenha sido oficializado em 1973 (21, p. 8), suas raízes filosóficas podem ser encontradas nos antigos debates gregos sobre a natureza do conhecimento humano. Uma diferença básica entre o trabalho dos gregos e dos cognitivistas modernos a esse respeito reside no uso recente de métodos computacionais para testar as teorias propostas.

Uma caracterização geral de *Ciência Cognitiva* é fornecida por Gardner que a define como um esforço contemporâneo, empiricamente baseado, para responder antigas questões epistemológicas – particularmente aquelas concernentes à natureza do conhecimento humano, seus componentes, fontes, desenvolvimento, etc. (9, p. 6).

A maioria dos trabalhos em *Ciência Cognitiva* inclui, com algumas pequenas diferenças, as seguintes suposições (Cf. 11, p. 8-11):

- 1) O estudo do conhecimento humano requer investigação das capacidades representacional e computacional da mente.
- 2) Há muito o que se ganhar com os estudos interdisciplinares.
- 3) Os computadores constituem instrumentos úteis para entender a mente humana.

A suposição (1) tem as suas raízes no século XVII, quando os debates acerca da natureza do conhecimento humano tomou um novo rumo com os trabalhos de Descartes (5) e outros. Estes introduziram novas perspectivas aos problemas concernentes à relação entre mente e corpo, linguagem e pensamento, percepção e sensação dos objetos. Em particular, Descartes reivindicou que o nosso conhecimento dos objetos externos, existentes no meio ambiente, é sempre mediado por *representações mentais*.

Assim, desde Descartes, questões concernentes à natureza das representações internas ocupam a atenção de filósofos e psicólogos. Atualmente, os aspectos *computacionais* das representações são analisados, constituindo um dos focos de atenção do nosso estudo sobre percepção visual.

A suposição (2), sobre a natureza interdisciplinar da *Ciência Cognitiva*, é baseada na crença de que pesquisadores com formações diferentes podem contribuir para um entendimento da mente humana ao compartilharem o conhecimento já desenvolvido em suas áreas de estudo. Essa crença tem como meta a obtenção de um estado de interação entre cientistas e filósofos de forma a criar um campo unificado de pesquisa, com um vocabulário comum, para descrever e explicar a atividade cognitiva.

A suposição (3), sobre a relevância dos computadores para o entendimento da mente humana, é baseada na crença de que:

(3a) Computadores podem ser usados como instrumentos para *testar* a precisão das hipóteses e teorias acerca do funcionamento da mente humana.

(3b) Os computadores podem servir como bons *modelos* para ilustrar a forma de funcionamento da mente humana.

A suposição (3a) reflete uma crença comum em *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 (9, p. 18). O sucesso de um programa computacional, que simula o comportamento de um organismo, é considerado como boa evidência para a precisão e plausibilidade das hipóteses e teorias que esse programa descreve. Por outro lado, se uma teoria, ou conjunto de hipóteses, for incompleta ou vagamente enunciada, a chance de ela vir a ser implementada em termos computacionais é muito baixa.

A idéia de igualar completude, ou ausência de vagueza, com computabilidade encontra grande oposição entre filósofos e psicólogos (Cf. 2, 6), que argumentam ser errônea a suposição de uma estreita relação entre modelos computacionais e teorias incompletas ou vagas. Embora essa polêmica possa ter interesse filosófico, o que mais interessa aos cognitivistas é que a prática de elaboração dos modelos computacionais tem mostrado a convivência e o valor prático da suposição 3(a).

A suposição 3(a) é conveniente porque ela sugere um critério científico e acessível para testar as teorias psicológicas. Ela possui um valor prático porque pode se fazer previsões acerca do comportamento do sistema através de sua simulação computacional.

A suposição (3b) – que os computadores servem como bons modelos do funcionamento da mente humana – é igualmente polêmica, dada as enormes diferenças existentes entre organismos e computadores. Contudo, a idéia subjacente a esta hipótese é que os computadores fornecem grande apoio para a concepção *funcionalista* da atividade cognitiva.

Uma versão informal do funcionalismo nos diz que, apesar de as máquinas e os seres humanos serem fisicamente diferentes, ambos possuem sistemas de processamento de informações, cujas operações mentais/computacionais podem ser entendidas em termos de suas *organizações funcionais*. Atualmente existem dois ramos principais do funcionalismo na Ciência Cognitiva, os quais denominaremos: (i) *funcionalismo lógico-computacional* e, (ii) *funcionalismo neuro-computacional* (11, p. 11-20). A diferença central entre (i) e (ii) é que, no primeiro caso, os processos mentais são estudados como se fossem apenas computações abstratas – independentemente de suas realizações físicas e do meio ambiente – que desempenham papéis funcionais no sistema cognitivo (Cf. 17, 18, etc.). Em contraposição, os funcionalistas neuro-computacionais enfatizam a relevância de considerações acerca dos aspectos *físicos*, *neurofisiológicos* e *ambientais* para o estudo e compreensão dos processos mentais.

A diferença de enfoque no estudo dos processos mentais existente entre as abordagens (i) e (ii) é particularmente relevante para a nossa presente análise das representações mentais. Como ficará claro a seguir, os funcionalistas lógico-computacionais descrevem as representações mentais como *abstrações* que funcionam como *mediadores* entre o organismo e o seu meio ambiente.

Para os funcionalistas neuro-computacionais, entre os quais esta autora se inclui, as representações mentais são estruturas emergentes da ativação de *unidades neurônio-símile* (*neuron like units*) que se *auto-organizam em função* da informação disponível no meio ambiente.

As discussões acerca da existência e da natureza das representações mentais assumem um caráter polêmico entre os estudiosos da percepção. Dada a relevância dessa polêmica para o entendimento do que se segue, vamos nos deter, brevemente, na análise de alguns dos seus pontos de discordância.

2. REPRESENTACIONALISMO VERSUS ELIMINATIVISMO

Foi Descartes quem iniciou, no século XVII, o estudo sistemático das “representações mentais”. Além de estabelecer a famosa *distinção mente/corpo*, ele argumentou que o nosso conhecimento do mundo externo é sempre mediado por objetos mentais abstratos, conhecidos como “idéias”, que de alguma forma representam o ambiente fora do sujeito. Descartes (5) insistiu que, embora as idéias, em geral, representem o mundo externo, a existência dessas idéias é independente das coisas que elas representam. Elas poderiam existir ainda que o mundo externo não existisse. Uma consequência dessa visão cartesiana é que podemos, acredita-se, dispensar considerações sobre o mundo externo ao sujeito quando investigamos a natureza da mente humana.

Seguindo a tradição cartesiana, estudos dos estados mentais são desenvolvidos entre filósofos, e funcionalistas lógico-computacionais (Secção 1), sem que informações sobre o meio ambiente e o físico dos organismos sejam levadas em consideração (Cf. 5, 8, 17, 18). Assim, por exemplo, a seguinte afirmação de Putnam, sobre os estados psicológicos ilustra a situação:

“... *no psychological state, properly so-called, presupposes the existence of any individual other than the subject to whom that state is ascribed*” (17, p. 220).

Tendo em vista essa tradição, podemos situar melhor os estudos recentes das representações mentais. Estes enquadram-se em dois grupos principais, conhecidos como *Representacionalismo* e *Eliminativismo* (8, p. 7). O primeiro tem como pressuposto a hipótese que existem estados da mente – as “representações mentais” – que codificam estados do mundo (8, p. 7). No caso da percepção visual, essas representações são entidades abstratas que funcionam como mediadores entre o organismo e o seu meio ambiente.

Os pesquisadores da *Inteligência Artificial* (AI, do inglês, *Artificial Intelligence*) (Cf. 16) e os funcionalistas lógico-computacionais são, em geral, Representacionalistas. Como foi mencionado na Secção 1, eles julgam que as representações mentais são entidades abstratas que não apenas podem, mas devem ser estudadas independentemente dos elementos físicos que as instanciam. Na sua forma mais radical, o Representacionalismo pressupõe que a capacidade de manipular e codificar símbolos constitui a essência da atividade representacional da mente (3, p. 12). As codificações caracterizam as representações mentais como sendo *abstrações* para agentes que funcionam como intérpretes. Para manipulá-las os agentes se utilizam de algum tipo de “linguagem do pensamento” que possua estrutura sintática e semântica. As dificuldades inerentes a esse processo de “interpretação” são familiares aos cognitivistas. Em particular, a exigência de um intérprete, que desempenha o papel de um *homunculus* no trabalho de decodificar e atribuir sentido às representações é problemática, posto que ela, obviamente, leva a um regresso infinito.

Em oposição ao Representacionalismo, os *Eliminativistas* (Cf. 3, 10) julgam ser desnecessário, e até desorientador (para os mais radicais) a pressuposição da existência de representações mentais, manipulação de símbolos, etc., nos estudos dos estados mentais. Fodor e Pylyshyn (8), nas suas críticas aos Eliminativistas ressaltam:

“Eliminativists ... think that psychological theories can dispense with such semantic notions as representations. According to Eliminativists the appropriate vocabulary for psychological theorizing is neurological, or perhaps behavioral...” (8, p. 7).

No domínio da percepção, os Eliminativistas negam a existência de qualquer estado representacional que estabeleça uma mediação entre o organismo e o meio ambiente. Ao contrário, eles desenvolvem uma abordagem que *ênfatiza* o estudo da interação entre o organismo e o seu meio ambiente. Tal abordagem foi desenvolvida nos trabalhos de Gibson (10), sendo que sua obra mais conhecida na década de 70 – *The Ecological Approach to Visual Perception* – iniciou uma nova linha de estudos da ação/percepção.

Como veremos a seguir, os Conexionistas se posicionam no meio do caminho entre os Eliminativistas e Representacionalistas. Do lado dos primeiros, os Conexionistas consideram ser de extrema importância para o estudo do sistema cognitivo as informações acerca da interação entre o organismo e o seu meio ambiente. Entretanto, do lado dos Representacionalistas, os Conexionistas explicitamente admitem a existência, e o papel fundamental, das representações mentais nos atos cognitivos/perceptuais. A pergunta imediata que se coloca é: Como é possível a existência de uma posição coerente entre duas posições aparentemente antagônicas? A resposta para essa questão envolve a reelaboração do conceito de “representação mental”; e essa é a principal inovação desenvolvida no interior do paradigma Conexionista que introduziremos a seguir.

3. UMA ABORDAGEM CONEXIONISTA DAS REPRESENTAÇÕES MENTAIS

O *Conexionismo*, também conhecido como *PDP (Parallel Distributed Processing)* Processamento em Paralelo de Informação Distribuída, ou *Redes Neurais (Neural Networks)*, é um ramo da Ciência Cognitiva (representado pelo funcionalismo neurocomputacional) que estuda as propriedades de sistemas dinâmicos, com características semelhantes àquelas do cérebro humano. No caso da percepção visual, os conexionistas elaboram modelos computacionais para simular mecanismos que, supostamente, atuam no cérebro e são responsáveis pela organização do sistema visual.

O Conexionismo foi reconhecido oficialmente como uma área específica de estudos no final da década de 70, estabelecendo-se então uma nova linha de pesquisa do sistema cognitivo, que se tornou conhecida na Europa e USA pela sua contribuição à tradicional *Inteligência Artificial (AI)*. Contudo, ao invés de focalizar a pesquisa nos métodos de resolução de problemas e regras abstratas de manipulação de símbolos (como o fazem os pesquisadores da AI), os conexionistas sugerem um nível menos abstrato de descrição do sistema cognitivo. Tal nível é conhecido como nível *sub-simbólico* de descrição (22, p. 261), porque ele descreve estados *físicos* do sistema cognitivo, os quais constituem, por hipótese, o substrato dos níveis de abstração hierarquicamente superiores deste sistema.

Os elementos primitivos de análise dos modelos conexionistas são as unidades físicas, denominadas *neurônio-símile* (*neuron like units*). Estas unidades possuem, como os neurônios reais, um nível de ativação que varia em função de influências externas, isto é, em função das suas conexões sinápticas com outras unidades neurônio-símile, que respondem aos estímulos providos do meio ambiente. A Figura 1 abaixo ilustra esta situação.

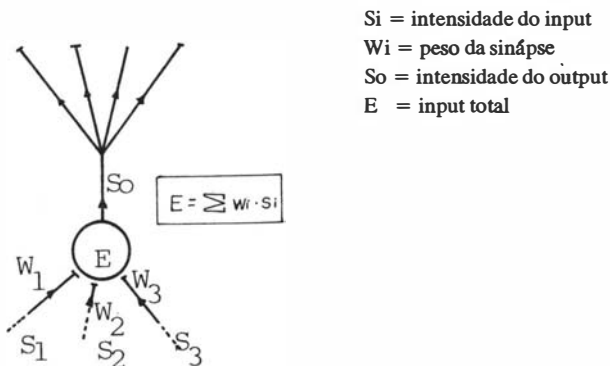


FIG. 1 – Esquema simplificado de um neurônio artificial

Estas unidades, ao serem ativadas emitem sinais a outras unidades que, por sua vez, dependendo da intensidade do sinal recebido, transmitem, em cadeia, um outro sinal. Populações de unidades neurônio-símile formam redes de camadas estruturadas: uma camada de unidades de entrada ou *input units*; uma camada de unidades de respostas, ou *output units* e, finalmente, zero ou mais camadas intermediárias entre *input* e *output*, ou *hidden units*. A Figura 2 ilustra um segmento de rede neural.

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 que existem entre essas unidades.

No caso da visão, a informação disponível no meio ambiente é estruturada na forma de arranjos ou padrões de luz (*light array*) que são refletidos dos objetos devido à sua localização espacial e à sua constituição físico-química. Esses padrões, que chamaremos de *padrões informacionais*, são apresentados às redes neurais na forma de vetores cujos elementos codificam dados estruturais que caracterizam objetos específicos. Eles incluem, por exemplo, as frequências e intensidades específicas dos raios luminosos que são refletidos dos objetos no campo visual do sistema. Esses elementos ao entrarem em contato com as unidades de entrada do sistema geram níveis de ativação, os quais, ao atingirem um valor limite (*threshold*), se propagam através da rede. Essa ativação será amplificada ou inibida em função dos pesos das conexões das unidades da rede, sendo a sua propagação estendida até que níveis específicos de

ativação sejam produzidos nas suas unidades de saída. Os valores destes níveis de ativação são associados a vetores que, idealmente, caracterizam os dados estruturais dos padrões informacionais que se quer identificar.

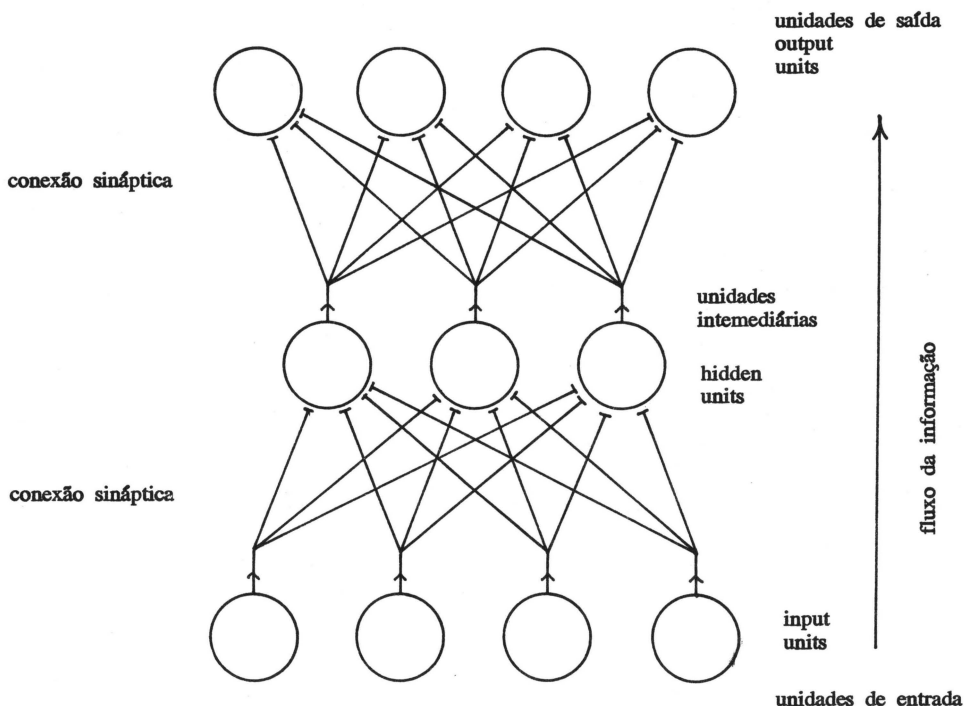


FIG. 2 – Esquema de uma rede neural artificial

Uma vez que, por hipótese, o meio ambiente é constituído 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. Para um sistema que ainda não recebeu treinamento, o peso destas conexões poderá ter um valor inicial arbitrário, estabelecendo-se, conseqüentemente, correlações arbitrárias entre este último e as regularidades no meio ambiente. Nesta situação, dado que não existe uma correspondência precisa entre as regularidades e os pesos das conexões, o sistema não terá como evitar erros na identificação de padrões informacionais. Contudo, é possível, através de treino onde se empregam “regras” específicas para o ajuste de pesos (Cf. 14, 20, 22), ajustar gradativamente os pesos das conexões até que eles converjam à um padrão estável que representará, as regularidades no meio ambiente, idealmente sem ambigüidades.

Tais *estruturas estáveis*, que *emergem* da interação entre os padrões de atividades das unidades neurônio-símile e os padrões de informação no meio ambiente, consti-

tuem as *representações mentais* do sistema. Elas se comportam como se estivessem *minimizando* uma quantidade que desempenha o papel de energia física da rede (esse ponto será retomado na Seção 4). A existência destas representações permite à rede identificar os padrões informacionais que lhes deram origem.

Uma vez que as representações são formadas, o sistema 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 o sistema foi treinado. Ilustrações desse processo de emergência de representações em sistemas conexionistas são amplamente conhecidas entre os cognitivistas. Um exemplo simples desse processo de formação de padrões estáveis é fornecido por Churchland (4, p. 205-208). Este considera um sistema sonar que é treinado para distinguir, em um ambiente aquático, ecos provenientes de rochas naturais e de minas explosivas. A Figura 3, extraída de Churchland (4, p. 206), ilustra um tal sistema.

A rede em questão possui 13 unidades na camada de entrada, sete unidades intermediárias e duas unidades de saída. Inicialmente, um selecionador de frequências seleciona 13 frequências que atuam, correspondentemente, nas unidades de entrada como níveis de ativação, cujos valores são, por convenção, alguma fração do valor 1 (por exemplo 0.3, 0.6, etc.). Tais valores são propagados através da rede, sendo alterados à medida que o sistema interage com os ecos, até produzir um par de níveis de ativação nas unidades de saída. Como Churchland ressalta (4, p. 205), apenas duas unidades de saída são necessárias, pois espera-se que a rede produza como *output* um vetor de ativação com valores próximos a $(0,1)$, quando um eco das *minas* entrar como *input*, e um vetor de ativação com valores próximos a $(1,0)$, quando um eco das rochas entrar como *input*.

Como inicialmente os valores dos pesos das conexões entre as unidades são estabelecidos ao acaso, o sistema apresentará respostas aleatórias às informações apresentadas às suas unidades de entrada. Neste caso, não se pode dizer que exista uma representação interna, propriamente dita, no sistema sobre quem é quem no ambiente. A novidade aqui consiste em que o sistema pode *aprender* a ajustar os pesos das suas conexões, em geral, com o auxílio de um computador que emprega uma regra para tal finalidade. Esse processo de aprendizagem é descrito por Churchland, da seguinte forma:

“We procure a large set of recorded samples of various (genuine) mine echoes..., and a comparable set of genuine set of rock echoes, keeping careful track of which is which. We then feed these echoes into the network, one by one, and observe the output vector produced in each case. What interests us in each case is the amount by the actual output vector differs from what would have been the ‘correct’ vector...” (4, p. 205-206)

Os erros detectados em cada apresentação são corrigidos, à medida que o sistema interage com as frequências registradas, através de graduais alterações nos pesos das conexões sinápticas reduzindo-se, a cada etapa, a discrepância entre os valores “cor-

retos” e os valores fornecidos pelo vetor de *output*. Se, após um grande número de apresentações dos ecos ao sistema, os pesos de suas conexões sinápticas tiverem sido ajustados de modo a eliminar totalmente a referida discrepância, então ele passará a distinguir os ecos das rochas dos ecos das minas. Neste caso, dizemos que a rede possui uma representação adequada dessa distinção entre ecos de *rochas* e ecos de *minas*.

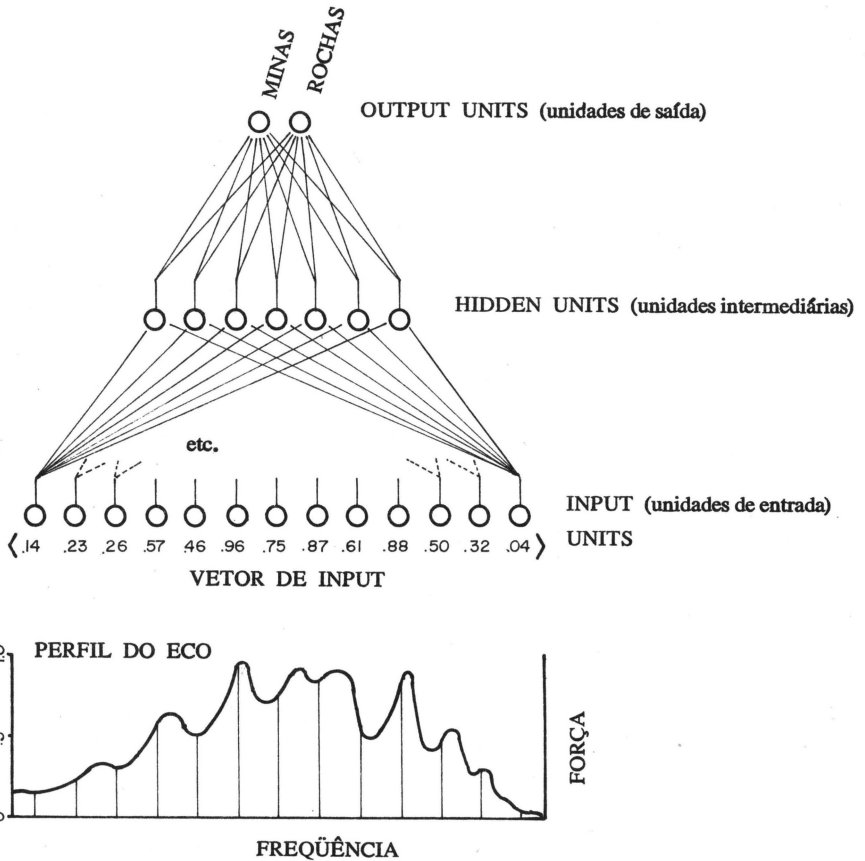


FIG. 3 – Uma rede neural artificial para o reconhecimento de padrões sonoros

Exemplos do tipo que acabamos de mencionar tem sido desenvolvidos com sucesso em casos de reconhecimento de padrões sonoros, visuais e lingüísticos, em vários modelos conexionistas (Cf. 4, 20, 22). Apesar das pesquisas nesta área estarem apenas iniciando, o sucesso obtido até agora com os modelos que simulam aspectos da atividade representacional humana indica que um novo paradigma está se formando para investigar questões acerca do nosso sistema cognitivo.

Um ponto importante a ser ressaltado é que o vocabulário usado neste paradigma conexionista, o qual inclui termos tais como “unidades neurônio-símile”, “nível de ativação das unidades de *input*”, etc., distingue as teorias conexionistas das demais teorias tradicionais do sistema cognitivo. Em particular, o uso destes termos emprestados da psicologia, biologia, física e neurofisiologia, permite uma nova interpretação naturalista do conceito de representação mental, cujas principais características acabamos de esboçar (para uma versão detalhada dessa concepção naturalista ver (10) e (11)). Essa nova interpretação parece indicar uma solução para a disputa entre Representacionista e Eliminativistas, mencionada na Seção 2. Pois, com os primeiros, pode-se afirmar não apenas a existência, mas o papel fundamental que as representações mentais, assim caracterizadas, desempenham no sistema perceptual/cognitivo humano. Ao mesmo tempo, com os Eliminativistas, essa noção de representação mental enfatiza a relevância da interação entre o sistema e o seu meio ambiente nos estudos do sistema cognitivo, sem cuja interação tais representações não existiriam.

Para concluir, vamos esboçar algumas idéias acerca da relação entre as noções de representação mental, harmonia e racionalidade, dentro da concepção conexionista apresentada.

4. REPRESENTAÇÃO MENTAL, HARMONIA E RACIONALIDADE

Na introdução deste trabalho sugerimos que a *intencionalidade* presente na percepção visual caracteriza-a como um processo cognitivo que envolve crenças, expectativas e outras atitudes representacionais meta-direcionadas. A existência desse tipo de atitude intencional por sua vez parece depender, em grande parte, de fatores sociais, culturais e contextuais em geral. Sabidamente, a observação de um cientista, de que há, por exemplo, um aparelho de raio X na sala, depende, de maneira determinante, de fatores contextuais, teóricos (13) e culturais que não se resumem aos dados sensoriais que chegam à sua retina.

Diante dessa constatação, a pergunta imediata que surge é: como pode uma teoria naturalista, do tipo conexionista, explicar o caráter intencional da visão? Para responder a essa pergunta, dois pontos precisam ser ressaltados. O primeiro tem por objetivo deixar claro que as hipóteses esboçadas aqui sobre a natureza das representações mentais não fornecem, nem pretendem fornecer, subsídios para uma teoria eliminativista, ou puramente reducionista, dos estados mentais. As representações mentais não se reduzem, na nossa perspectiva, a entidades materiais. Como insistimos

anteriormente, elas são *estruturas informacionais emergentes* da interação entre populações de neurônios e padrões informacionais no meio ambiente. A natureza dessa interação não é necessariamente causal mas, sim, *informacional* (Cf. 7, 10 e 11 para detalhes sobre esse tópico). Nesta relação informacional existe um elemento de *covariância* entre os padrões de atividades neural e os padrões informacionais no meio ambiente. É essa covariância que caracteriza um padrão emergente da atividade neural como uma representação mental – do tipo detector de sons, por exemplo. Dada a suposição da existência de covariância entre os padrões neurais e informacionais, a ativação de um padrão neural pode indicar, e em circunstâncias adequadas representar, a ocorrência de informação de um tipo específico (de som, por exemplo) que está sendo processada pelo sistema.

É claro que a covariância a que nos referimos deve ser regida por regularidades *funcionais* adaptativas que permitam ao sistema se *auto-organizar* em função de mudanças no meio ambiente. Em alguns sistemas conexionistas, a auto-organização ocorre com a auxílio de mecanismos que atuam no ajuste de pesos das conexões entre as unidades neurônio-símile. Tais mecanismos, conhecidos, como *learning rules* (regras de aprendizagem), permitem ao sistema corrigir erros na detecção de padrões e ajustar o seu comportamento em função de sua interação com a informação recorrente no meio ambiente. Esse elemento adaptativo-funcional diferencia as representações internas de sistemas complexos, que se auto-organizam, dos simples registros, realizados por sistemas tais como um termômetro ou termostato, de variações que ocorrem no meio ambiente.

Assim, para retomar a questão da intencionalidade, mencionada no início desta Seção, não propomos aqui hipóteses puramente reducionistas da intencionalidade dos atos perceptuais, mas sim, uma investigação acerca de sua base informacional. Essa base, como indicamos anteriormente, não se reduz a entidades materiais, ainda que as entidades materiais sejam absolutamente relevantes para a sua existência. Como enfaticamente observou Wiener:

“Information is information, not matter or energy. No materialism which does admit this can survive at the present day.” (23, p. 132)

Em segundo lugar, cabe lembrar que há vários níveis de investigação do sistema perceptual/cognitivo. O paradigma conexionista sugere um estudo dos níveis de organização desse sistema a partir da análise de suas bases neurofisiológica e informacional. Contudo, o conexionismo está engatinhando nos seus esforços para a realização desse projeto. Não existe, ainda, uma teoria conexionista, no sentido próprio do termo, que explique de forma satisfatória o elemento intencional da percepção. O que dispomos no momento são esboços rudimentares do que poderá vir a constituir uma teoria cognitiva da percepção, que explique o seu caráter intencional.

Tendo em vista os dois pontos acima, podemos agora investigar algumas idéias acerca da relação entre representação mental, racionalidade e harmonia.

No que diz respeito à relação entre representação mental e racionalidade, é importante observar que a natureza em si não parece ser racional, intencional ou não inten-

cional. Não faz sentido para nós filósofos e cognitivistas a idéia de que rios, montanhas, pedras, etc., se comportam racional ou irracionalmente. Isso porque, os objetos naturais apenas *são* – existem. As ações, crenças, e atitudes intencionais em geral, passam a ser consideradas racionais ou irracionais no interior do nosso *universo representacional*. Por isso é comum o pressuposto de que o estudo da racionalidade deve ter como base de referência o conceito de representação mental.

Conforme foi ressaltado nas palestras anteriores, a noção de *racionalidade* não pode ser caracterizada por um atributo único. Ela inclui vários elementos, entre outros as noções de eficiência, coerência, capacidade de fazer inferências corretas, etc. No domínio da percepção/ação a racionalidade requer a capacidade, por parte do agente, de selecionar e representar *coerentemente* a informação disponível no meio ambiente, e agir em concordância com essa informação.

No caso específico da percepção visual, a racionalidade está principalmente relacionada às noções de representação coerente e representação eficiente. A importância destas noções se torna evidente se considerarmos, como Marr (15) que ver é saber o que é algo e onde esse algo se encontra no ambiente. Pois, a identificação de algo requer, de acordo com a concepção discutida aqui, que o sistema represente a informação visual, recorrente no meio ambiente, de maneira inequívoca. Deve haver uma correspondência única entre um arranjo luminoso e sua representação no interior de um sistema treinado para reconhecer padrões visuais. Se assim não fosse, o sistema poderia identificar, no momento t_1 , por exemplo, um arranjo luminoso refletido de flores como “flores”, sendo que no momento t_2 o mesmo arranjo luminoso poderia ser identificado como “jacarés”. Essas representações, além de não serem eficientes para o sistema (que poderia eventualmente desaparecer ao confundir flores com jacarés), não estariam auxiliando na tarefa de “ver”, tal como caracterizada anteriormente. Pois, não haveria aqui uma identificação, propriamente dita, de quem é quem no meio ambiente.

Em termos conexionistas se considera que a presença de tais *inconsistências* aumenta o “grau de entropia” (quantidade de desordem) no sistema, e diminui a sua eficiência no reconhecimento de padrões. Em tal circunstância, onde uma mesma representação possui conteúdos diversos e incompatíveis, o sistema no qual ela ocorre poderia ser considerado como apresentando um certo grau de “irracionalidade”.

A função dos treinos nas redes neurais, através dos quais se empregam “regras de aprendizagem” (*learning rules*) para ajustar os pesos das conexões entre as unidades neurônio-símile, é justamente reduzir o valor correspondente ao seu grau de entropia. Quando esse valor atinge um mínimo, estruturas estáveis são formadas, às quais nos referimos na Seção 3. Como foi indicado, estas estruturas estáveis – candidatas a representações internas – se comportam como se estivessem minimizando uma certa quantidade que desempenha o papel de energia física da rede.

Smolensky (22) esboçou uma teoria conexionista das representações mentais, conhecida como *Harmony Theory* (Teoria Harmônica), a qual sugere uma ponte entre a teoria matemática da informação, por um lado, e a termodinâmica e mecânica

estatística, por outro. Como nos demais modelos conexionistas, o peso das conexões entre as unidades neurônio-símile codifica a frequência com que os padrões informacionais invariantes ocorrem no meio ambiente. Quanto mais frequentemente uma regularidade no meio ambiente for detectada mais alto será o valor do peso destas conexões.

Na Teoria Harmônica, as conexões entre unidades neurônio-símile que registram regularidades no meio ambiente são denominadas “átomos de conhecimento”. Conjuntos de átomos de conhecimento, que possuem um certo grau de coerência dão lugar a estruturas estáveis conhecidas como *esquemas*. Uma vez que essas estruturas são formadas, o sistema pode realizar “abstrações”, no sentido que, ainda que mudanças ocorram na direção, posição, tamanho, etc., de um padrão a ser reconhecido, mesmo assim tais mudanças não afetam o processo de reconhecimento. Além disso, certos elementos que não são percebidos num momento específico podem ser antecipados. O sistema é capaz de “inferir” aspectos de um padrão visual que, devido a certas condições de visibilidades, não são registrados. O reconhecimento é feito através de mecanismos que completam os padrões informacionais cujas estruturas básicas estão presentes nos esquemas, dos quais o caso em questão é apenas uma instância.

Uma das novidades do modelo de Smolensky é que, uma vez que os esquemas são formados, a rede “aprende” a se movimentar de um esquema para o outro em função de mudanças no meio ambiente e do grau de harmonia, ou coerência, nas suas representações. Quando novos estímulos chegam ao sistema ele tende a detectar aqueles conjuntos invariantes de caracteres que *concordam* ou *se harmonizam* com os seus esquemas. Estímulos não compatíveis com tais esquemas introduzem caos no sistema, aumentam 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 *input* ambientais). Nestas circunstâncias, a Teoria Harmônica determina que novos conjuntos de estímulos sejam examinados pela rede e que esta selecione os elementos que são compatíveis com os esquemas já adquiridos.

O modelo proposto por Smolensky sugere uma descrição da dinâmica do processo perceptual, em particular no que se refere à detecção de eventos e mudanças que ocorrem no meio ambiente. De acordo com essa perspectiva, os atos perceptuais podem ser caracterizados em termos da aquisição de representações perceptuais *consistentes*. Estas representações perceptuais são concebidas como estruturas que codificam regularidades no mundo e preparam o organismo para agir. Uma hipótese similar é sugerida por Armstrong (1), que analisa as nossas “crenças perceptuais” – entendidas como instanciações de atitudes intencionais – caracterizando-as como estados do organismo que o habilitam a agir.

A idéia central subjacente a essa concepção da dinâmica do sistema perceptual pode ser resumida através de um princípio que denominaremos “*princípio de relevância*”. De acordo com este princípio, sistemas de processamento de informação

cérebro-símile *maximizam* a eficiência de seu comportamento através da seleção *coerente* de informação disponível no meio ambiente, sendo que tal seleção deve envolver um mínimo possível de energia 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, preparando-o para “agir”.

Como os modelos de redes neurais não são determinísticos, o critério de relevância se aplica apenas de forma probabilística. Algumas vezes, padrões informacionais, ou aspectos destes padrões, são representados tornando inconsistentes as representações pré-existentes no sistema. Nestas circunstâncias, a tendência será uma reorganização dos átomos de conhecimento existentes ou a eliminação da nova representação. Quando esta eliminação não ocorrer, o sistema tenderá a reorganizar suas representações de modo a obter novas estruturas estáveis que direcionarão o seu comportamento futuro, de forma racional. Assim, a existência temporária de inconsistências no sistema pode ter um papel criativo na formação de novas estruturas representacionais que, eventualmente, serão estabilizadas.

No caso das representações mentais humanas, este modelo se desenvolve em um meio informacional extremamente complexo, o qual inclui, entre outras coisas, informações a respeito de outras mentes. É claro que as representações dos modelos conexionistas não possuem, nem de longe, a riqueza, flexibilidade e complexidade das representações mentais humanas. Contudo, os modelos conexionistas sugerem alguns princípios, tais como aqueles discutidos acima, que parecem estar presentes no processo de representação mental dos seres racionais.

A expectativa é de que a fusão de esforços, que se combinam na Ciência Cognitiva, para o estudo desses princípios possibilite o estabelecimento de alguns alicerces para o desenvolvimento de uma ciência do sistema cognitivo humano.

GONZALES, M. E. Q. Neural networks and mental representation: an essay on harmony and rationality. *Trans/Form/Ação*, São Paulo, v. 14, p. 93-108, 1991.

ABSTRACT: We develop an analysis of the notion of mental representation, in the domain of visual perception, from a Cognitive Science perspective. Emphasis is given on a Connectionist view, according to which mental representations are emergent properties of the interaction between brain-like systems and structured light in the environment. We suggest that such a notion of mental representation indicates a way out of an ancient dispute between Representationalism and Eliminativism regarding the existence of mental representation in the human perceptual system.

KEYWORDS: Visual perception; connectionism; neural networks; cognitive science; patterns of information; mental representation; self-organization.

REFERÊNCIAS BIBLIOGRÁFICAS

1. ARMSTRONG, D. M. *A materialist theory of the mind*. New York: Humanities Press, 1968.
2. BAKER, L. R. *Saving beliefs: a critique of physicalism*. New Jersey: Princeton University Press, 1987.
3. BICKHARD, M., RICHIE, D. *On the nature of mental representation: a case study of James Gibson's theory of perception*. London: Praeger, 1983.
4. CHURCHLAND, P. M. Cognitive activity in artificial neural networks. In: OSHERSON, D. N., SMITH, E. E., ed. – *Thinking*. Cambridge: MIT Press, 1990. v. 3, p. 199-227.
5. DESCARTES, R. *As meditações*. Trad. de J. Guinsburg e Bento Prado Júnior. São Paulo: Abril Cultural, 1979.
6. DREYFUS, H. *What computers can't do*. New York: Harper and Row, 1972.
7. DRETSKE, F. *Knowledge and the flow of information*. Cambridge: Bradford Books, 1981.
8. FODOR, J. A., PYLYSHYN, Z. W. Connectionism and cognitive architecture: a critical analysis. *Cognition*, Lausanne, v. 28, p. 3-71, 1988.
9. GARDNER, H. *The mind's new science: a history of cognitive revolution*. New York: Basic Books, 1987.
10. GIBSON, J. J. *The ecological approach to visual perception*. Boston: Houghton-Mifflin, 1979.
11. GONZALES, M. E. Q. *A cognitive approach to visual perception*. Colchester: Essex University, 1989. Tese (Doutorado).
12. GONZALES, M. E. Q., FRENCH, T., TREFFNER, P. A naturalistic approach to mental representation. In: GILHOOLY, K., KEANE, M., ERDOS, G., ed. – *Lines of thinking: reflections on the psychology of thought*. London: John Wiley & Sons, 1990, p. 57-69.
13. HANSON, N. R. *Patterns of discovery*. Cambridge University Press, 1958.
14. HEBB, D. O. *Organization of behavior*. New York: John Wiley, 1949.
15. MARR, D. *Vision*. San Francisco: Freeman, 1980.
16. NEWELL, A., SHAW, J., SIMON, H. *Human problem solving*. New Jersey: Englewood Cliffs, 1972.
17. PUTNAM, H. The meaning. In: GUNDERSON, K. *Mentality and machines*. New York: Crowell, 1975.
18. PYLYSHYN, Z. W. *Computation and cognition*. Cambridge, MA: MIT Press, 1984.
19. REID, T. (1785) *Essays on the intellectual powers of man*. London: Sheldon Press, 1986.
20. RUMELHART, D., MACCLELLAND, J. *Explorations in parallel distributed processing*. Cambridge, MA: MIT Press, 1988, v. 1.
21. SCHEERER, E. *Towards a history of cognitive science*. *Cognitive Science*, Norwood, NJ, v. 115, p. 7-9, 1988.
22. SMOLENSKY, P. Information processing in dynamical systems: foundations of harmony theory. In: RUMELHART, D., MACCLELLAND, J., *Explorations in parallel distributed processing*. Cambridge: MIT Press, 1986. v. 1. p. 390-431.
23. WIENER, N. *Cybernetics, or control and communication in the animal and the machine*. 2. ed. Cambridge, MA: MIT Press, 1961.