

Maurizio de Queiroz Oliveira

Ciência Cognitiva e Auto-Organização: Um
Encontro no Modelo Conexionalista da Mente

Marília-SP
1999



1410001026



MAURIZIO DE QUEIROZ OLIVEIRA

**CIÊNCIA COGNITIVA E AUTO-ORGANIZAÇÃO: UM ENCONTRO
NO MODELO CONEXIONISTA DA MENTE**

Dissertação apresentada à Faculdade de Filosofia e Ciências da Universidade Estadual Paulista "Júlio de Mesquita Filho", Campus de Marília, para obtenção do título de Mestre em Filosofia (Área de Concentração: Ciência Cognitiva e Filosofia da Mente).

Orientadora: Profª Drª Maria Eunice Quilici Gonzales



**MARÍLIA
1999**

MAURIZIO DE QUEIROZ OLIVEIRA

**CIÊNCIA COGNITIVA E AUTO-ORGANIZAÇÃO: UM
ENCONTRO NO MODELO CONEXIONISTA DA MENTE**

COMISSÃO JULGADORA

DISSERTAÇÃO PARA OBTENÇÃO DO TÍTULO DE MESTRE

Presidente e Orientador : Maria Eunice Quilici Gonzales
2º Examinador : José Roberto Castilho Piqueira
3º Examinador : Carmen Beatriz Milidoni

Marília, 16 de Novembro de 1999



AGRADECIMENTOS

Somos gratos a todos os pessoas que colaboraram para tornar este trabalho possível. Em particular:

DADOS CURRICULARES

A Coordenação do curso de pós-graduação em ciência cognitiva e filosofia da mente do Câmpus de Marília - Unesp;

MAURIZIO DE QUEIROZ OLIVEIRA

Aos professores e colegas do curso de Mestrado em Ciência Cognitiva e Filosofia da Mente;

Aos funcionários da Biblioteca do Câmpus de Marília - Unesp;

NATURALIDADE Orindiúva-SP

A direção da Faculdade de Odontologia do Câmpus de Araçatuba - Unesp Prof. Dr. João César Bedran de Castro;

FILIAÇÃO Nivercio Moisés de Oliveira

Eremita de Queiroz Oliveira

Aos meus funcionários do Polo Computacional da Faculdade de Odontologia do Câmpus de Araçatuba - Unesp;

1980-1984 Curso de Graduação em Engenharia Elétrica
Faculdade de Engenharia de Ilha Solteira – Unesp

A direção e os funcionários do Polo de Odontologia do Câmpus de Araçatuba - Unesp.

1993- Analista de Informática do Polo Computacional do
Câmpus de Araçatuba - Unesp

AGRADECIMENTOS

Somos gratos a todas as pessoas que colaboraram para tornar este trabalho possível, em particular:

À Coordenação do curso de pós-graduação em ciência cognitiva e filosofia da mente do Câmpus de Marília - Unesp;

Aos professores e colegas do curso de Mestrado em Ciência Cognitiva e Filosofia da Mente;

Aos funcionários da Biblioteca do Câmpus de Marília - Unesp;

À direção da Faculdade de Odontologia do Câmpus de Araçatuba - Unesp, Prof. Dr. João César Bedran de Castro;

Aos meus colegas, Katia e Antonio, funcionários do Polo Computacional da Faculdade de Odontologia do Câmpus de Araçatuba - Unesp;

À direção e todos os funcionários da Biblioteca da Faculdade de Odontologia do Câmpus de Araçatuba - Unesp.

DEDICATÓRIA

Ao meu pai (*in memoriam*) e a minha mãe: apoio e vida

A Priscila pelo apoio e compreensão

Prof. Jose Roberto Castilho Piqueira

pelas sugestões que envolveram
este trabalho.

Profª. Carmen Beatriz Miodonni

devo interesse a observações relevantes
sobre este trabalho.

AGRADECIMENTOS ESPECIAIS

Prof^a. Maria Eunice Quilici Gonzales

pela oportunidade, dedicação, apoio e
por apontar caminhos;

Prof. José Roberto Castilho Piqueira

pelas sugestões que enriqueceram
este trabalho;

Prof^a. Carmen Beatriz Milidonni

pelo interesse e observações relevantes
sobre este trabalho.

SUMÁRIO

INTRODUÇÃO GERAL	1
CAPÍTULO I - INTRODUÇÃO À CIÊNCIA COGNITIVA	
Apresentação	5
Seção I - Principais caracterizações da ciência cognitiva	8
Seção II - Breve histórico da ciência cognitiva	13
Seção III - Fundamentos conceituais e principais pressupostos da ciência cognitiva	18
CAPÍTULO II - FUNCIONALISMO : UMA PROPOSTA DE SOLUÇÃO DO PROBLEMA MENTE-CORPO	
Apresentação	29
Seção I - O problema da relação mente-corpo	30
Seção II - Funcionalismo computacional	37
Seção III - Ramificações da ciência cognitiva: inteligência artificial e redes neurais artificiais	45
CAPÍTULO III - O PAPEL DOS MODELOS NA EXPLICAÇÃO DA MENTE	
Apresentação	62
Seção I - O papel dos modelos na ciência cognitiva	64
Seção II - Redes neurais artificiais e os modelos cognitivos	73
Seção III - Limites e alcances dos modelos cognitivos	82
CAPÍTULO IV - AUTO-ORGANIZAÇÃO	
Apresentação	87
Seção I - Resumo histórico das idéias de auto-organização	89
Seção II - Auto-organização : aspectos conceituais	94
Seção III - Auto-organização primária e auto-organização secundária	112
Seção IV - Ciência cognitiva e auto-organização	116
CAPÍTULO V - APLICAÇÕES DA TEORIA DA AUTO-ORGANIZAÇÃO	
Apresentação	120
Seção I - Plausibilidade biológica da rede de Kohonen	121
Seção II - Auto-organização na rede neural de Kohonen	127
Seção III - Ilustrações do modelo computacional da rede de Kohonen	137
CONCLUSÃO	143
REFERÊNCIAS BIBLIOGRÁFICAS	153
RESUMO	164
ABSTRACT	165

LISTA DE FIGURAS

	Página
Figura 1 - Gráfico conceitual da ciência cognitiva com suas disciplinas e diferentes abordagens	25
Figura 2 - Tabela de transição de estados de uma máquina de Turing	39
Figura 3 - <i>Machine table</i> para uma instância de uma máquina de Turing	40
Figura 4 - Estrutura de um computador digital simples	51
Figura 5 - Desenho esquemático de um neurônio.....	55
Figura 6 - Uma rede neural simples.....	56
Figura 7 - Uma rede neural artificial típica	74
Figura 8 - Modelo matemático de uma unidade das RNA	76
Figura 9 - Função sigmóide	77
Figura 10 - Imãs do experimento de Von Foerster: a) antes da agitação b) após a agitação	106
Figura 11 - <i>Tonotopic map</i>	121
Figura 12 - Circuito inibitório lateral pelo qual uma fibra excitada de um agrupamento neuronal pode determinar inibição de fibras adjacentes	123
Figura 13 - Aumento no contraste do padrão de estímulo determinado pelo circuito inibitório	124
Figura 14 - Função Chapéu Mexicano	125
Figura 15 - <i>Kohonen feature map</i>	128
Figura 16 - Dois exemplos de vizinhança topológica: a) retangular, b) hexagonal	130
Figura 17 - Rede de Kohonen	133
Figura 18 - Rede de Kohonen	138

LISTA DE FIGURAS

Página

- Figura 19 - A mudança do valor dos pesos das conexões rotaciona o vetor de pesos para a proximidade do vetor de entrada 139
- Figura 20 - Representação gráfica da evolução dos vetores pesos durante os ciclos de treinamento da rede de Kohonen . 141

LISTA DE ABREVIATURAS

- GPS - *General Problem Solver*
- IA - Inteligência Artificial
- RNA - Redes Neurais Artificiais
- SOFM - *Self-Organizing Features Maps*
- PDP - Processamento Distribuído em Paralelo

INTRODUÇÃO GERAL

Este trabalho tem por objetivo analisar as propriedades de auto-organização que supostamente poderiam ser exibidas em modelos connexionistas utilizados para simular processos cognitivos. Em particular, analisaremos o modelo conhecido como *rede neural de Kohonen*. Este objetivo será realizado em duas etapas, correspondentes as partes I e II da dissertação.

Na parte I, que reúne os capítulos 1, 2 e 3, caracterizamos a área da ciência cognitiva, mostrando suas origens, as principais hipóteses e modelos utilizados para a análise dos processos cognitivos.

Na parte II, que reúne os capítulos 4 e 5, analisamos o tema da auto-organização, nos seus aspectos conceituais, sugerindo uma possível convergência de interesses entre a ciência cognitiva e o paradigma da auto-organização.

O entendimento da natureza da capacidade cognitiva presente na relação homem-meio ambiente faz parte da preocupação de várias áreas do conhecimento. A análise da inteligência, em especial na ciência cognitiva, começou a partir dos anos 50, através de uma cooperação interdisciplinar envolvendo as áreas de psicologia cognitiva, lingüística, cibernética, neurociência, filosofia e antropologia, estabelecendo-se um programa de pesquisas com o objetivo de elaborar uma teoria científica da mente.



Para cumprir com os desafios teóricos e práticos do estudo dos fenômenos mentais, os cientistas cognitivos partem de três suposições fundamentais que caracterizam o campo da ciência cognitiva: a) a existência da capacidade representacional da mente, b) a hipótese segundo a qual o computador constitui um modelo apropriado do pensamento humano e, no plano metodológico, c) a hipótese de que a cooperação interdisciplinar trará resultados férteis para o estudo da mente.

No interior desse programa, o estudo da natureza da cognição é realizado por meio da construção de modelos simulados em computador. O uso de modelos como uma estratégia de análise da atividade cognitiva resultou na formulação de dois paradigmas centrais: a inteligência artificial e o connexionismo ou redes neurais artificiais.

O emprego desses modelos objetiva a construção e teste de hipóteses e teorias a respeito dos fenômenos cognitivos, sem considerar, inicialmente, os aspectos emocionais e histórico-culturais na compreensão destes fenômenos.

Os modelos utilizados no âmbito da ciência cognitiva apoiam-se nos conceitos de informação, *máquina de Turing* e funcionalismo, bem como no funcionamento do sistema nervoso humano. Nos modelos elaborados na inteligência artificial, o sistema cognitivo é caracterizado como um sistema gerador e manipulador de símbolos por meio de regras abstratas. Já nos modelos de redes neurais artificiais, o sistema cognitivo é entendido como um processo emergente de uma ampla interação entre elementos densamente interconectados.



Tais modelos são analisados, ainda, como elementos para formular possíveis respostas a antigos problemas filosóficos; entre eles, o problema da relação mente-corpo. A apresentação deste e de outros problemas, bem como suas propostas de soluções, constituem o núcleo da primeira parte da dissertação.

Na segunda parte, analisamos o tema da auto-organização através do modelo de Rede Neural de Kohonen. Descartada a possibilidade de auto-organização nos modelos de inteligência artificial, passamos a analisar a possibilidade de auto-organização nos modelos conexionistas. Nosso objetivo é verificar até que ponto um modelo, por exemplo do tipo rede de Kohonen, pode exibir propriedades de auto-organização, estabelecendo-se assim uma possível conexão entre a ciência cognitiva e a teoria da auto-organização.

Em suma, nosso objetivo nesta segunda parte é responder a seguinte questão: Pode uma máquina conexionista, particularmente a rede de Kohonen se auto-organizar? Nossa pretensão, ao levantar esta questão, é extremamente moderada diante da dificuldade em fornecer uma resposta precisa. Assim, nos limitaremos a verificar em que sentido o termo auto-organização estaria sendo usado e apontar as principais alegações de ordem teórica, os limites e alcances do modelo de Kohonen diante da possibilidade de auto-organização.



Apresentação

A investigação dos processos cognitivos humanos vem desde a época dos filósofos gregos e ganhou corpo a partir do século XVIII com o advento da psicologia experimental. Desde então, a psicologia tem se desenvolvido como uma ciência autônoma, com métodos próprios e com uma linguagem específica.

Compreender a natureza das capacidades cognitivas humanas é um dos objetivos centrais da psicologia cognitiva. Este livro tem como finalidade apresentar os fundamentos desta ciência, desde suas origens até os avanços mais recentes.

CAPÍTULO I – INTRODUÇÃO À CIÊNCIA COGNITIVA

A psicologia cognitiva surgiu no início do século XX, quando os pesquisadores começaram a estudar os processos mentais de forma sistemática. Esta abordagem se baseia na ideia de que a mente pode ser estudada através de métodos científicos, como a observação, a experimentação e a análise estatística. A psicologia cognitiva tem se desenvolvido rapidamente, tornando-se uma das áreas mais ativas da psicologia contemporânea.

Com relação à definição e ao domínio de atuação da ciência cognitiva, Gardner (1985, p. 19) afirma:

“A ciência cognitiva é o estudo dos processos mentais que permitem ao indivíduo compreender o mundo ao seu redor e agir de acordo com esta compreensão. Ela envolve a interação entre a biologia, a psicologia e a linguística, buscando entender como a informação é processada e utilizada para a tomada de decisões e a resolução de problemas.”

Para cumprir com esta definição, a psicologia cognitiva deve considerar tanto os aspectos biológicos quanto os psicológicos e linguísticos. A abordagem cognitiva busca entender a mente como um sistema integrado, onde diferentes processos trabalham juntos para produzir o comportamento observado.

Apresentação

A investigação dos processos cognitivos humanos vem desde a época dos filósofos gregos e ganhou maior estímulo a partir das reflexões de Descartes sobre a natureza da mente e sua relação com o corpo.

Compreender a natureza das capacidades cognitivas humanas envolvidas na interação homem-meio ambiente – manifestas na memória, aprendizado, atenção, visão, linguagem, etc. - faz parte da preocupação de várias áreas do conhecimento. Na ciência cognitiva, esta preocupação começou a partir dos anos 50 através de uma cooperação interdisciplinar envolvendo as áreas de psicologia cognitiva, lingüística, inteligência artificial, cibernética, neurociência, filosofia e antropologia, formalizando-se um programa de pesquisas com o objetivo de estabelecer uma teoria científica da mente.

Com relação à definição e ao domínio de atuação da ciência cognitiva, Gardner (1995, p. 19) afirma:

Defino a ciência cognitiva como um esforço contemporâneo, com fundamentação empírica, para responder questões epistemológicas de longa data - principalmente aquelas relativas à natureza do conhecimento humano, seus componentes, suas origens, seu desenvolvimento e seu emprego.

Para cumprir com os desafios teóricos e práticos de entender os fenômenos mentais, os cientistas cognitivos conforme

ressaltamos na introdução, partem de três suposições fundamentais que caracterizam o campo da ciência cognitiva: a) a existência da capacidade representacional da mente, b) o computador constitui um modelo apropriado do pensamento humano, c) a cooperação interdisciplinar trará resultados férteis para o estudo da mente. Comentaremos, no decorrer deste capítulo, cada uma destas suposições.

Historicamente, apesar do termo *ciência cognitiva* ser mais intensamente utilizado a partir da década de 70, 1956 é considerado o ano de seu nascimento (Gardner, 1995, p.43). Nessa época, a concepção dominante para avaliar o comportamento inteligente era o behaviorismo, que estudava a relação estímulo-resposta sem enfatizar os aspectos intermediários presentes nesta relação, preocupando-se apenas com as reações observáveis e mensuráveis. Ao contrário desta perspectiva, a vertente cognitivista considera os aspectos internos da interação homem-ambiente, admitindo a existência de representações mentais e tratando a mente como um aparato processador de informação na constituição de tais representações.

Conceitualmente, o programa de estudos da ciência cognitiva incorpora as contribuições das concepções filosóficas Dualista e Monista e suas principais implicações para a análise da mente. No interior deste programa, o estudo da natureza da cognição é realizado por meio da construção de modelos simulados em computador: o modelo de Inteligência Artificial e o Processamento Distribuído em Paralelo ou Redes Neurais Artificiais.

Seção I - Principais Neste capítulo caracterizamos a ciência cognitiva, descrevendo seu perfil histórico, seu domínio de atuação e métodos de análise da mente, por meio de três seções. Na primeira, descrevemos as principais caracterizações da ciência cognitiva e seus objetivos. Na segunda, traçamos um perfil histórico da ciência cognitiva, destacando os principais acontecimentos, trabalhos científicos e pesquisadores que contribuíram para a formação dessa área de estudos. Na última seção, analisamos os principais pressupostos e características metodológicas adotadas pelos cientistas cognitivos para compreender a natureza do conhecimento.

Conforme veremos, vários pesquisadores contribuíram para definir os contornos da ciência cognitiva durante o século XX. Alan M. Turing, por exemplo, desenvolveu, em 1936, a noção de uma máquina abstrata que pudesse simular o raciocínio inteligente, denominada Máquina de Turing. De acordo com Gardner (1995, p.32) o conceito de máquina de Turing resultou na seguinte conclusão: "... na medida em que se possam especificar os passos necessários para a execução de uma tarefa, ela poderá ser programada e executada por uma máquina de Turing".

Segundo o relato histórico realizado por Gardner (1995), Von Neumann desenvolveu a ideia de que um computador eletrônico

Seção I - Principais caracterizações da ciência cognitiva

O interesse em caracterizar a natureza do conhecimento humano é uma antiga inquietação da área de filosofia. Atualmente, além dos filósofos, os pesquisadores das áreas de lingüística, psicologia cognitiva, neurociências e da inteligência artificial procuram interagir no sentido de formar uma área autônoma de estudos sobre a mente. Esse esforço investigativo, para conhecer o quê, como, e por quê os seres humanos aprendem, memorizam, percebem, resolvem problemas e se interagem com o meio ambiente, tem por objetivo levar a definição de uma área de estudos e métodos experimentais necessários à formação de uma concepção científica, interdisciplinar da mente humana. Essa área ficou conhecida como *ciência cognitiva*.

Conforme veremos adiante, vários pesquisadores contribuíram para definir os contornos da ciência cognitiva durante o século XX. Alan M. Turing, por exemplo, desenvolveu, em 1936 a noção de uma máquina abstrata que pudesse simular o comportamento inteligente, denominada *Máquina de Turing*. De acordo com Gardner (1995, p.32) o conceito de máquina de Turing resultou na seguinte conclusão "... na medida em que, se pudesse expressar claramente os passos necessários para a execução de uma tarefa, ela poderia ser programada e executada por uma máquina de Turing".

Segundo o relato histórico realizado por Gardner (1995), Von Newmann desenvolveu a idéia de que um computador eletrônico



poderia ser controlado por um programa armazenado em sua memória interna. Os pesquisadores McCulloch & Pitts em 1943 propõem que uma rede biológica poderia ser modelada em termos de lógica, enfatizando que : "Qualquer coisa que possa ser descrita de forma exaustiva e inequívoca ..é.. realizável por uma rede neural finita apropriada" (Gardner, 1995, p. 33). Durante as décadas de 30 e 40, Wiener apud (Ashby, 1970, p.1) conceituou o termo cibernética definindo-a como " a teoria do controle tanto no homem quanto na máquina" , investigando as características funcionais comuns existentes entre o sistema nervoso e as máquinas.

Outro trabalho importante para o fundamento da ciência cognitiva foi a Teoria Matemática da Comunicação desenvolvida por Shannon em 1949. Esse trabalho descreve a aplicação dos predicados lógicos, do tipo verdadeiro ou falso, ao comportamento dos circuitos elétricos. Essa teoria contribuiu para desenvolver os conceitos iniciais de informação com ênfase em seu caráter estatístico.

A seguir discutiremos algumas caracterizações da ciência cognitiva encontradas na literatura e em alguns centros de pesquisa sobre o assunto. Conforme ficará evidente no que se segue, os cientistas cognitivos seguem várias orientações, porém, guiados pela idéia de procurar uma explicação científica da mente.

Assim, Gardner (1995, p.19), conforme descrito na apresentação deste capítulo, enfatiza que a ciência cognitiva procura esclarecer questões que foram objeto de análise da filosofia por longa data, tais como o problema da relação entre a mente e o corpo, a natureza das

representações mentais, etc. Porém, ao contrário da filosofia, a ciência cognitiva demonstra uma preocupação “empírica” ao abordar estas questões. Modelos cognitivos serão elaborados e testados nesta área.

Segundo Simon & Kaplan (1989, p.1), a ciência cognitiva enfatiza o “estudo da inteligência e dos sistemas inteligentes com particular referência ao comportamento inteligente como a computação”. Esse autor ressalta que apesar de não existir uma caracterização satisfatória do conceito de inteligência, ela se manifesta nos seres humanos quando eles resolvem problemas, escolhem um curso de ação para atingir determinado objetivo, quando respondem questões de maneira apropriada, etc.

Uma terceira caracterização da ciência cognitiva é oferecida por Pylyshyn (1986, p. xi) . Para este autor, a ciência cognitiva “... é uma abordagem para o estudo da mente” com a possibilidade de se tornar um domínio científico tão bem sedimentado quanto a biologia, a química ou a geologia. De maneira análoga a estas áreas, a ciência cognitiva analisaria cientificamente a mente por meio de um conjunto de princípios próprios que a tornaria um campo de pesquisas autônomo. Pylyshyn (1986) destaca, ainda, que o caráter interdisciplinar da ciência cognitiva é fruto da convergência de interesses sobre várias questões filosóficas e da necessidade de compartilhar técnicas de simulação computacional.

Os relatos abaixo, extraídos de
(http://web.psych.ualberta.ca/~mike/Pearl_Street/INTD554/Week1/control.h

tml,1997) refletem a atividade no campo da ciência cognitiva de algumas universidades e centros de pesquisas nesta área.

Na Universidade da Califórnia, em San Diego, a ciência cognitiva vem sendo desenvolvida a partir de uma perspectiva interdisciplinar, envolvendo a neurociência, ciência da computação, antropologia, psicologia, lingüística e a filosofia.

Na Universidade de Princeton, a ciência cognitiva é caracterizada como o estudo interdisciplinar dos processos cognitivos a partir das áreas de antropologia, ciência da computação, lingüística e história da ciência.

No Centro de Ciência Cognitiva de Rutgers os seus pesquisadores entendem que a ciência cognitiva se preocupa em explicar a natureza do comportamento inteligente. O comportamento inteligente é caracterizado como um processo simbólico que pode ser instanciado em sistemas naturais ou artificiais. O desenvolvimento das atividades de pesquisa neste centro é de natureza interdisciplinar envolvendo as áreas de psicologia experimental, ciência da computação, neurociência, filosofia, lingüística, matemática e a engenharia.

As caracterizações descritas acima, evidenciam uma série de características em comum, com diferenças nas ênfases com que são pesquisados diversos temas que compõem a ciência cognitiva. Está presente, na maioria das caracterizações, o caráter interdisciplinar da ciência cognitiva, dela participando, basicamente, as áreas de filosofia, inteligência artificial, lingüística, psicologia e antropologia. Essa cooperação

interdisciplinar visa a dar à ciência cognitiva um vocabulário específico e um conjunto de princípios que a torne uma ciência autônoma. Identifica-se a importância do computador como um elemento experimental no qual os processos cognitivos possam ser instanciados e inclui, entre as suas preocupações, a análise do comportamento inteligente tanto em humanos, como nos animais e nas máquinas.

Em resumo, o cientista cognitivo se preocupa fundamentalmente em investigar a natureza da mente por meio da construção de modelos, assumindo uma abordagem representacional e uma perspectiva interdisciplinar. Na seção seguinte, abordaremos os fatos históricos, os trabalhos científicos e os pesquisadores que marcaram o nascimento da ciência cognitiva, alguns dos quais foram citados nesta seção.



Seção II – Breve histórico da ciência cognitiva

Considera-se 1956 como sendo o ano de nascimento consensual da ciência cognitiva (Gardner 1995, p. 43) e 1973 o ano de utilização formal do termo (Gonzales, 1991, p. 94).

A concepção científica dominante, nesta época, era o behaviorismo, cujo foco de análise da inteligência é o comportamento objetivo e observável. Tal concepção fundamenta-se em alguns pressupostos básicos, entre eles: a) a explicação do comportamento inteligente dos organismos fundamenta-se na relação entre estímulo e resposta; b) a introspecção para o estudo da inteligência não é essencial como parte de seus métodos de análise; c) utiliza-se do método da observação para obter informações a respeito do comportamento.

Ao considerar apenas o efeito do meio ambiente sobre o comportamento do indivíduo, desprezando seus aspectos funcionais e estruturais, o behaviorismo, na visão de alguns psicólogos da vertente comportamentalista, não seria capaz de explicar comportamentos inteligentes específicos.

A reação à análise comportamentalista começou através de alguns membros da própria área behaviorista (Gardner, 1995, p. 27), que começaram a considerar a existência de fenômenos cognitivos que antecedem comportamentos complexos e recorreram a processos introspectivos para explicar a estrutura do pensamento humano. Essa nova forma de estudar o comportamento humano deu origem a uma nova

concepção científica denominada *cognitivismo*, perspectiva teórica que influenciou diretamente o nascimento da ciência cognitiva.

O surgimento da ciência cognitiva foi precedido pela realização de várias conferências, simpósios e publicações das quais participaram cientistas de diversas áreas do conhecimento. Esses acontecimentos foram marcados por sua natureza interdisciplinar, envolvendo, principalmente, antropólogos, matemáticos, filósofos, neurofisiologistas, psicólogos e engenheiros.

O primeiro encontro foi chamado de Conferências Macy e realizado em 1949. Vários encontros foram realizados sob o título de “Conferências sobre Mecanismos Circulares Causais e de *Feedback* em Sistemas Biológicos e Sociais”. O objetivo fundamental que levou vários cientistas a participarem desta série de conferências foi a “ambição de edificar uma ciência geral do funcionamento da mente”. Essas conferências contaram com a participação de Wiener, Von Neumann, McCulloch e Von Foerster, entre outros estudiosos da cibernética (Dupuy, 1996, p. 9).

Derivado da palavra grega *Kybernetes*, que significa piloto, o movimento cibernético foi caracterizado por dois estágios. O primeiro foi marcado por introduzir conceitos como os de informação, comunicação e *feedback*. O segundo estágio, conhecido como *segunda cibernética*, enfatizou a análise dos aspectos autônomos dos sistemas naturais e artificiais preocupando-se com conceitos como os de *causalidade circular*, *informação*, *autonomia*, entre outros, bem como com os estudos sobre *auto-organização*.

Na visão de Dupuy (1996, p. 15), o nome cibernética saiu de “moda” sendo substituído por “ciência da cognição”. Esse autor argumenta que a origem da ciência cognitiva está fortemente ligada à cibernética, que tinha por objetivo o estudo científico da mente :

Daí uma mudança de foco que obscurece o que era o projeto propriamente científico e filosófico da cibernética. Tratava-se, nada mais nada menos, de conduzir a aventura científica à sua apoteose, edificando uma ciência da mente. É realmente essa ambição que faz da cibernética o antepassado das ciências cognitivas.

Outro acontecimento importante no âmbito do movimento cibernético foi o Simpósio Hixon, realizado no MIT - Massachusetts Institute of Technology - em 1948. Seu objetivo principal era abordar questões a respeito de como o sistema nervoso controla o comportamento. Participaram desse simpósio cientistas que iniciaram a abertura do campo da ciência cognitiva, destacando-se entre eles, Von Neumann, que apresentou um paralelo entre o computador eletrônico e o cérebro, e McCulloch, que debateu aspectos do *processamento de informação* no sistema nervoso.

Outro marco importante no surgimento da ciência cognitiva foi o simpósio sobre a teoria da informação realizado no MIT, no período de 10 a 12 de setembro de 1956. Segundo Gardner (1995, p.43), foi neste simpósio que o termo ciência cognitiva foi oficialmente reconhecido. Destacaram-se, pela relevância dos trabalhos apresentados, Newell &

Simon, com o artigo “Máquina de teoria lógica”, cientistas esses que vieram a ser os fundadores do campo da inteligência artificial. Nesse simpósio, Gardner ressalta que ficou mais uma vez evidenciado o caráter interdisciplinar no surgimento do campo da ciência cognitiva, fruto da convergência de interesses entre áreas aparentemente independentes como a psicologia, a lingüística e a ciência da computação, fornecendo o suporte teórico para a simulação computacional dos processos cognitivos.

A noção de informação desempenha um papel importante no campo da ciência cognitiva. Tal importância está relacionada com os modelos da inteligência artificial e das redes neurais artificiais, utilizados para o estudo dos processos cognitivos, concebidos como um sistema processador de *informação*. Nesse sentido, no âmbito da ciência cognitiva, a mente é caracterizada de maneira que suas propriedades “são [...] propriedades de sistemas de processamento de informação” Dupuy (1996, p. 27).

A noção de informação contida na “Teoria Matemática da Comunicação”, de Shannon & Weaver (1949), fornece noções relevantes para o estudo dos processos cognitivos. Essa teoria foi inicialmente elaborada para abordar o problema da transmissão de informação por meio de circuitos de comunicação. Nesse contexto, os autores propõem uma formulação quantitativa da noção de informação associada com a idéia de ocorrência de eventos em um canal de comunicação, idéia esta que será amplamente utilizada na ciência cognitiva.

Seção III - F Em resumo, os trabalhos apresentados nos simpósios e conferências, acima mencionados, tiveram sua importância na estruturação das características fundamentais da atual ciência cognitiva, ou seja, a interação interdisciplinar, a preocupação com o entendimento dos processos cognitivos e a possibilidade da simulação computacional desses processos.

Na seção seguinte examinaremos os aspectos conceituais fundamentais e os principais pressupostos sobre os quais está edificada a ciência cognitiva.



Seção III - Fundamentos conceituais e principais pressupostos da ciência cognitiva

A tentativa de entender os aspectos conceituais que fundamentam a ciência cognitiva remete-nos ao estudo dos problemas filosóficos clássicos. As reflexões filosóficas sobre a natureza do conhecimento funcionam, de acordo com Gardner (1995, p. 58), "... como um ponto de partida lógico para investigações da ciência cognitiva" fornecendo elementos conceituais relevantes para o desenvolvimento teórico dessa ciência. Este autor enfatiza que é evidente a presença dos problemas levantados pelos filósofos, na maioria dos textos que caracterizam o seu domínio. Conforme vimos na introdução deste capítulo, a relevância da filosofia, em seu ramo epistemológico, para a ciência cognitiva, aparece na própria caracterização que Gardner (1995, p.6) dá a esta área ao enfatizar que ela se preocupa em "responder questões epistemológicas de longa data".

Entre as questões amplamente analisadas no âmbito filosófico e que fazem parte da agenda de questões da ciência cognitiva, está o problema da relação mente-corpo. Impulsionada por uma reflexão sistemática do filósofo Descartes, a análise do problema da relação mente-corpo resultou na caracterização do ser humano dividido em duas substâncias: a espiritual e a corporal. O corpo, na concepção cartesiana, é caracterizado como uma substância material extensa, que ocupa lugar no espaço, pode ser decomposta e seu comportamento ser regido por leis

físicas. A alma é caracterizada como uma substância não material, não extensa, que não ocupa lugar no espaço, indivisível, não perecível e não regida por leis físicas.

A caracterização radicalmente oposta entre o corpo e a alma resultou na dificuldade em compreender *como é possível a interação entre duas substâncias de naturezas completamente distintas*. Ou seja, como é possível uma substância extensa (o corpo) interagir com uma substância sem nenhuma extensão (a alma) ? Estabelecido o dualismo cartesiano, várias concepções filosóficas surgiram procurando investigar a natureza dos estados mentais e seu relacionamento com o corpo.

Estas concepções, que atualmente dominam o estudo da mente, podem ser caracterizadas em teorias dualistas e materialistas. A posição materialista enfatiza a existência de uma única substância. A teoria dualista se diversifica em dualismo de substância e dualismo de propriedades. O dualismo de substância ou dualismo cartesiano, tal como caracterizado acima, enfatiza que a propriedade principal da mente é o pensamento; o dualismo de propriedades reivindica que a mente pode ser idêntica a matéria, mas possuindo propriedades não materiais. De acordo com Gonzales (1998, p. 91) no dualismo de propriedades :

..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 por leis da física, biologia ou da neurologia.

A posição materialista rejeita o dualismo mente-corpo e postula uma identidade entre a mente e matéria. De acordo com Churchland (1988, p.2), a teoria materialista da mente reivindica que "os estados e processos mentais são meramente estados e processos sofisticados de um sistema físico complexo : o cérebro ". De acordo com Armstrong & Malcolm (1984, p.191) a teoria materialista sobre a natureza da mente pode fornecer uma explicação completa do homem em termos puramente físico-químicos.

Duas correntes relevantes para a análise dos estados mentais se destacam entre os materialistas: os materialistas reducionistas e os materialistas eliminativistas. A corrente reducionista reivindica que os estados mentais são estados físicos do cérebro, ou seja, a mente se reduz ao físico. Duas versões reducionistas se destacam no estudo da interação entre a mente e o corpo: a teoria da identidade de tipos e a teoria da identidade de particulares.

A primeira sustenta que um tipo específico de evento mental, como a dor, por exemplo, é idêntico a um tipo específico de evento cerebral, como por exemplo a excitação de neurônios específicos.

Na segunda versão reducionista, a teoria da identidade de particulares, a natureza da identidade entre eventos mentais e eventos físicos não é tão forte quanto na identidade de tipos. Esta versão materialista reivindica que a ocorrência de um estado mental é idêntica a um estado cerebral necessário, mas não específico. Portanto, esta visão confirma a identidade entre estados mentais e estados cerebrais, porém



admite a possibilidade de um mesmo estado mental ser identificado a estados cerebrais distintos (Bechtel, 1988, p.107). Neste sentido, a identificação do estado cerebral não garante a identificação do estado mental equivalente.

Outra concepção materialista a respeito dos eventos mentais é o materialismo eliminativo. Tal corrente propõe a eliminação dos fenômenos mentais, tais como entendidos pelo senso comum, sustentando que eles, na realidade, não existem. Os materialistas eliminativos acreditam que a psicologia popular é completamente falsa, argumentando que as crenças, desejos, medos, etc., seriam abandonados e substituídos por uma explicação apoiada no desenvolvimento teórico das neurociências. Assim, os processos cognitivos seriam explicados puramente no plano neuronal. Segundo Churchland (1988) :

...o materialismo eliminativo é uma tese em que nossa concepção de senso comum do fenômeno psicológico constitui uma teoria radicalmente falsa, fundamentalmente defeituosa tanto em princípio e ontologia, uma teoria que eventualmente será transferida, preferivelmente reduzida de forma suave, através de uma neurociência completa.

Por último, entre as teorias sobre os estados mentais acrescentamos a teoria funcionalista, atualmente dominante na ciência cognitiva. A concepção funcionalista sustenta que os estados mentais são estados funcionais caracterizados por um "conjunto de relações causais entre efeitos ambientais no corpo, outros tipos de estados mentais e o

estado ou comportamento corporal” (Churchland, 1988, p.36). Devido a sua importância para a ciência cognitiva, a teoria funcionalista será tratada em maiores detalhes na seção II do capítulo II.

No que segue, iremos analisar os principais pressupostos em que se apoia a ciência cognitiva. Gardner (1995, p.20) aponta cinco pressupostos através dos quais a ciência cognitiva procura compreender a natureza da atividade cognitiva. O primeiro refere-se a necessidade de admitir a existência de entidades abstratas denominadas de *representações mentais*. Estas representações formam o elemento básico da atividade cognitiva e devem ser analisadas, segundo Gardner, separadas dos aspectos biológicos e culturais.

De acordo com Stillings (1987, p.305), a idéia de representação estava presente na obra de Descartes ao argumentar que o nosso conhecimento do mundo é mediado por representações. A hipótese de que os processos cognitivos podem ser analisados a partir de representações mentais é um dos pressupostos metodológicos centrais da ciência cognitiva.

As representações mentais, tais como, símbolos, idéias, regras e imagens são manipuladas, reunidas, recuperadas e computadas formando uma estrutura mental que interage com os estímulos recebidos e os comportamentos emitidos durante a relação homem-meio ambiente. De acordo com Thagard (1998, p.21), “... uma teoria cognitiva postula um conjunto de estruturas representacionais e um conjunto de processos que operam nessas estruturas”.



A existência de representações na mente, como elementos utilizados para se chegar ao conhecimento, também é ressaltado por Dupuy (1996,p.27) ao argumentar que o ato de conhecer é essencialmente um processo de representação.

Para Pylyshyn (1986, p.xii), a idéia de representação mental estabelece um ponto comum entre os problemas filosóficos clássicos e as noções da ciência da computação. A abordagem teórica fornecida pela filosofia, acerca da natureza das representações mentais, aliada a construções de modelos computacionais permite à ciência cognitiva obter resultados empíricos sobre o funcionamento da mente.

Os cientistas cognitivos que admitem a existência de representações mentais trabalham com a hipótese de que estas representações são manipuladas por procedimentos computacionais levando-nos a analisar o segundo pressuposto da ciência cognitiva: o *computador*, entendido como instrumento apropriado para analisar, modelar ou explicar a mente.

De acordo com Gardner (1995, p.20), no âmbito da ciência cognitiva "...existe a crença de que o computador eletrônico é essencial para qualquer compreensão da mente humana". A utilização do computador para estudar os processos cognitivos apoia-se na teoria da máquina de Turing que, em sua essência, sinalizou a possibilidade de simular o comportamento inteligente em um aparato capaz de manipular procedimentos algorítmicos.

Segundo Pylyshyn (1985, p.21), "... a necessidade de admitir as representações para dar explicações do fenômeno cognitivo constitui uma das principais razões do porquê a cognição é fundamentalmente um processo computacional". A característica principal que faz dos computadores digitais um pressuposto central da ciência cognitiva é a sua capacidade de processar informação. O computador como um sistema que recebe, processa e transmite informações permite ao cientista cognitivo elaborar modelos e realizar testes sobre o fenômeno cognitivo apoiado na analogia mente-computador.

A relevância do uso dos computadores na ciência cognitiva não é unânime. Existem aqueles que argumentam ser as abordagens representacional e computacional inadequadas para explicar a natureza da mente em todos os seus aspectos.

O terceiro pressuposto aponta para a necessidade de uma abordagem interdisciplinar para compreender a natureza da mente. Assim, os cientistas cognitivos, por meio da interação entre diferentes áreas, esperam formar uma teoria autônoma com métodos próprios para estudar a mente. Essa natureza interdisciplinar da ciência cognitiva emergiu da necessidade de compartilhar os vários interesses e conceitos comuns obtidos nas áreas da computação, filosofia, antropologia, psicologia, lingüística, neurociências e da inteligência artificial para explicar a atividade cognitiva. De acordo com Stillings (1987, p.10), a neurociência procura entender os aspectos fisiológicos do processamento da informação no sistema nervoso humano; a lingüística pesquisa a estrutura e a natureza da

aquisição da linguagem; os filósofos, realizando um trabalho de base, interessam-se pelos aspectos conceituais da ciência cognitiva; os psicólogos analisam as capacidades mentais humanas, e os cientistas da computação, através da área de inteligência artificial, estão interessados em simular o comportamento inteligente.

Varela et al. (1993, p.7) caracteriza a natureza interdisciplinar da ciência cognitiva de forma gráfica (Figura 1) mostrando a diversidade de orientações teóricas presentes no campo da ciência cognitiva e suas principais disciplinas.

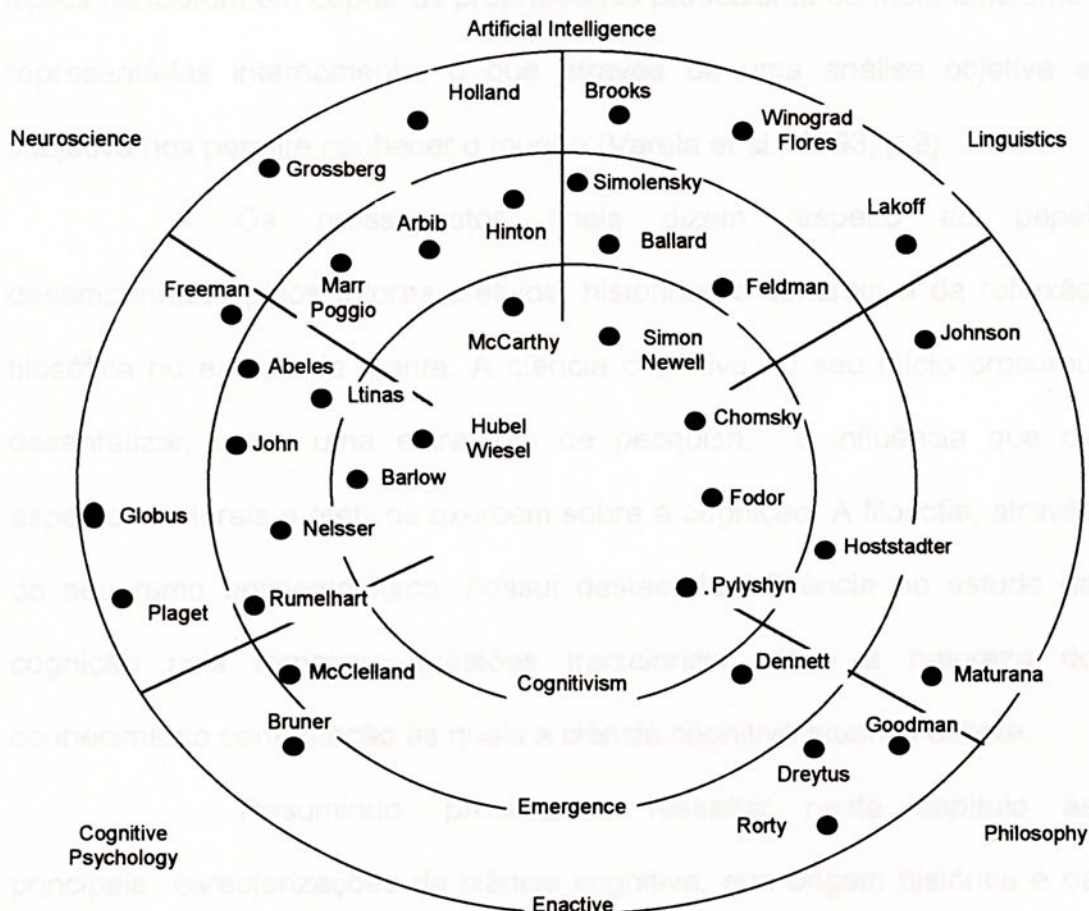


FIGURA 1 - Gráfico conceitual da ciência cognitiva com suas disciplinas e diferentes abordagens. Varela et al. (1993, p. 7)

A figura acima enfatiza que a ciência cognitiva, nos seus primeiros anos, foi dominada pelo paradigma cognitivista, cuja idéia central consiste em estudar a mente como manipulações de símbolos através de regras, sem compromisso com seus aspectos físicos. Ultimamente, surgiram abordagens alternativas como a denominada de connexionismo onde os processos cognitivos são efeitos globais ou emergentes, a partir da interação de um grande número de elementos densamente interconectados. A outra abordagem alternativa enfatiza que os processos cognitivos decorrem de uma variedade de ações dos organismos no mundo. Tais ações consistem em captar as propriedades particulares do meio ambiente, representá-las internamente, o que através de uma análise objetiva e subjetiva nos permite conhecer o mundo (Varela et al., 1993, p.9).

Os pressupostos finais dizem respeito ao papel desempenhado pelos fatores afetivos, históricos e culturais e da reflexão filosófica no estudo da mente. A ciência cognitiva no seu início procurou desenfaturar, como uma estratégia de pesquisa, a influência que os aspectos culturais e afetivos exercem sobre a cognição. A filosofia, através do seu ramo epistemológico, possui destacada influência no estudo da cognição pois forneceu questões tradicionais sobre a natureza do conhecimento com relação às quais a ciência cognitiva atual se debate.

Resumindo, procuramos ressaltar neste capítulo as principais caracterizações da ciência cognitiva, sua origem histórica e os principais pressupostos que fundamentam este campo. Eles nos revelam a ênfase mecanicista dada por esta ciência no estudo da mente.



No próximo capítulo, apresentamos uma proposta de solução do problema da relação mente-corpo através da teoria funcionalista.

CAPÍTULO II - FUNCIONALISMO: UMA PROPOSTA DE SOLUÇÃO DO PROBLEMA MENTE-CORPO

Apresentação

Na agenda de questões da IA, há um problema permanente suscitando desafios e problemas de natureza essencialmente cognitiva.

A ciência cognitiva, apoiada pela teoria da inteligência artificial, elaborou modelos computacionais para compreender e representar a inteligência humana. O uso de modelos como uma metodologia de análise de problemas e a representação de estruturas de conhecimento são aspectos fundamentais na construção de sistemas de inteligência artificial e na solução de problemas de inteligência artificial.

CAPÍTULO II - FUNCIONALISMO: UMA PROPOSTA DE SOLUÇÃO DO PROBLEMA MENTE-CORPO

Na seção I, fazemos uma breve descrição do problema da mente-corpo e analisamos as críticas de Ryle (1948) ao paradigma da mente-corpo e analisamos as críticas de Ryle (1948) ao paradigma da mente-corpo.

Na seção II, caracterizamos a corrente funcionalista e seus argumentos em relação a uma proposta para a solução do problema mente-corpo.

Na última seção, estudamos os modelos de inteligência artificial e as redes neurais artificiais, procurando descrever os aspectos fundamentais da inteligência artificial e da solução do problema mente-corpo.

Apresentação

Na agenda de questões da ciência cognitiva um problema permanece suscitando desafios: o problema da relação mente-corpo.

A ciência cognitiva, apoiada pela teoria funcionalista, elabora modelos computacionais para compreender a natureza desse problema. O uso de modelos como uma metodologia de análise da atividade cognitiva resultou na formulação de dois paradigmas : o modelo de inteligência artificial e o modelo de redes neurais artificiais.

Neste capítulo, analisamos os principais conceitos e dificuldades de uma proposta de solução do problema mente-corpo apresentada pela perspectiva funcionalista. Desenvolvemos, ainda, um estudo conceitual acerca dos modelos de inteligência artificial e de redes neurais artificiais.

Na seção I, fazemos uma breve descrição do problema da relação mente-corpo e analisamos as críticas de Ryle (1970) ao paradigma cartesiano.

Na seção II, caracterizamos a corrente funcionalista e seus limites no que se refere a uma proposta para a solução do problema mente-corpo.

Na última seção, estudamos os modelos de Inteligência artificial e de redes neurais artificiais, procurando destacar os aspectos utilizados nestes modelos para a análise dos processos cognitivos.

Seção I – O Problema da relação mente-corpo

A filosofia da mente preocupa-se, fundamentalmente, com o entendimento da natureza da mente humana. Neste campo, são discutidos os méritos, as objeções e os aspectos conceituais das diversas teorias sobre a natureza dos estados mentais, com o objetivo de caracterizar, principalmente, a relação entre fenômenos mentais e fenômenos físicos. Ao investigar a natureza da atividade mental, a filosofia da mente se depara com questões filosóficas que há muito tempo causam inquietação aos filósofos, transmitindo-as à ciência cognitiva. Neste sentido, a variedade de respostas oferecidas pelas diversas concepções filosóficas tem influenciado diretamente a configuração atual da ciência cognitiva. Portanto, é fundamental para os cientistas cognitivos com uma preocupação empírica, incorporar as reflexões da filosofia da mente na busca de uma ciência cognitiva consistente e acompanhada de rigor conceitual e metodológico.

Na agenda de questões mais estudadas no campo da filosofia da mente *uma possui importância fundamental na ciência cognitiva*, a saber, o conhecido *problema da relação mente-corpo*, principalmente enunciado no interior da filosofia cartesiana. Esta seção não se propõe a realizar um estudo filosófico da obra de Descartes mas apenas destacar, de maneira descritiva, alguns aspectos fundamentais da distinção cartesiana entre a mente e o corpo e suas implicações para a ciência cognitiva.

Apesar de encontrarmos reflexões a respeito da natureza da relação entre a mente e o corpo nos filósofos clássicos, foi Descartes,

quem iniciou, no século XVII, uma reflexão sistemática sobre a distinção entre o espírito e o corpo. O esforço cartesiano foi empreendido adotando o princípio da dúvida sistemática, no qual considerava-se como falso tudo aquilo de que se pudesse suspeitar ou sobre o qual pairasse a mínima dúvida. Seguindo este princípio, Descartes acreditou ser possível obter as idéias claras e distintas sobre o conhecimento humano, objetivando estabelecer a estrutura de uma ciência universal. Com este objetivo, ele estabeleceu a conhecida distinção entre a mente e o corpo. Tal distinção fundamenta-se nos princípios da razão universal e não depende da transitoriedade e variedade dos elementos materiais.

Em sua *distinção entre a alma e o corpo*, Descartes (1996) sugere que o ser humano é formado por duas substâncias : a material e a espiritual. O corpo ou substância material, cuja propriedade principal é a *extensão (res extensa)*, é divisível, ocupa lugar no espaço e seu comportamento é puramente regido por leis físicas. Em contraste com as propriedades da matéria, o espírito ou alma seria formado por uma substância imaterial, indivisível, cujo desenvolvimento escapa às leis físicas e cuja propriedade principal é o *pensamento (res cogitans)*.

O projeto cartesiano, descrito na obra *Meditações*, destaca vários outros aspectos a respeito da distinção corpo-alma. Primeiro, que a alma seria mais fácil de ser conhecida do que o corpo. Diz Descartes (1996, p. 275) em sua segunda Meditação :

...só concebemos os corpos pela faculdade de entender em nós existente e não pela imaginação nem pelos sentidos, e não os conhecemos pelo fato de os

ver ou tocá-los, mas somente por os conceber pelo pensamento, reconheço com evidência que nada há que me seja mais fácil de conhecer do que meu espírito.

No plano lógico-filosófico do projeto cartesiano, o corpo pode ser visto como um elemento passivo, submetido a um campo de forças mecânicas, destituído de propriedades específicas da alma, como pensar, sentir emoções, desejos e crenças. O corpo é comparável a um mecanismo ou máquina. Em contraste, a alma, além de existir independentemente do corpo, é uma entidade unificada, possuidora de capacidades mentais-comportamentais que a levam agir livremente e a desencadear comportamentos na matéria. A possibilidade desta substância imaterial movimentar ou influenciar o comportamento da substância material é de difícil compreensão na explicação cartesiana.

Apesar de enfatizar a diferença radical entre a natureza da alma e do corpo e a independência entre elas, Descartes (1996, p.328-9) enfatiza a união entre essas substâncias :

A natureza me ensina, também, por esses sentimentos de dor, fome etc., que não somente estou alojado em meu corpo, como um piloto em seu navio, mas que, além disso, lhe estou conjugado muito estreitamente e de tal modo confundido e misturado, que componho com ele um único todo.

Esta espécie de "mistura" corpo-alma parece ser incoerente com o fato de Descartes admitir que as duas substâncias são de natureza absolutamente distintas. Nesse sentido, permanece obscuro

como o divisível (substância extensa) se une ao indivisível (substância pensante). Compreender a natureza dessa união é o ponto central do problema mente–corpo.

A perspectiva dualista do ser humano traz consigo diversas dificuldades para o pensamento filosófico cartesiano. Em primeiro lugar, ao destacar que o corpo comporta-se como um mecanismo, demonstrando o seu interesse pela física em amplo crescimento, naquele momento em que a ciência galileana era fluorescente, como poderia Descartes explicar características humanas como a responsabilidade, a linguagem, a percepção, a ação livre, se estes são controlados externamente por leis físicas.

Por outro lado, a mente, com sua natureza introspectiva recebe uma caracterização negativa pois mente é tudo aquilo que o corpo não é : uma substância não-material, não-visível, escapando à aplicação das leis físicas. Em resumo, a mente é caracterizada como um ente não mecânico e que não pode ser decomposto; uma entidade transcendental existindo separada e independentemente do corpo.

Ao postular a existência de duas substâncias de naturezas distintas (física e mental), Descartes (1996) deixou não resolvidos vários problemas. O primeiro resume a dificuldade de entender como se processa a união entre o físico e o mental. A tentativa de explicação empreendida por esse filósofo para esta questão recorre a uma explicação no campo da fisiologia : a interação inicia-se na chamada glândula pineal, localizada no cérebro, que seria uma espécie de sede da alma. Porém, fica mal explicada

a interação da alma com a glândula pineal. As explicações cartesianas, como o próprio Descartes (1996) admite em *As Paixões da Alma* são pouco convincentes e incapazes de esclarecer a real natureza da relação entre a mente e o corpo. Assim, o problema permanece sujeito a críticas de diversos estudiosos, configurando-se um verdadeiro desafio para os filósofos e cientistas que o sucederam.

Estabelecido o problema da relação mente-corpo, vários pensadores tentaram propor soluções, umas fundadas na redução de uma substância a outra, e outros tentando dissolver o problema.

Entre as críticas mais sofisticadas e sistemáticas à teoria cartesiana estão as desenvolvidas pelo filósofo da mente, Gilbert Ryle, em sua obra *The concept of mind*, publicada pela primeira vez em 1949. Ryle (1970, p.12) denomina o dualismo cartesiano de *doutrina oficial* por sua “preponderância entre especialistas e mesmo leigos”, procurando defender a tese segundo a qual a filosofia cartesiana contém erros de princípio em sua formulação. Segundo esse autor, a teoria cartesiana instaura o *dogma do fantasma na máquina* ao postular que o corpo pode se comportar ou ser governado por uma mente invisível, imaterial e inacessível, *misturada* a ele.

Dentre as objeções levantadas por Ryle (1970), a mais significativa delas é aquela que aponta a existência de um equívoco lingüístico que leva a um erro categorial ao postular a existência de duas substâncias de natureza distintas. Segundo Ryle (1970), o erro categorial presente na formulação do problema mente-corpo consiste em classificar

(categorizar) a mente como pertencendo à mesma categoria lógica que o corpo, quando na realidade, pertence a um outro tipo.

Assim, Descartes (1996) teria cometido uma espécie de confusão conceitual ao classificar a mente em oposição ao corpo, mas mantendo-a na mesma categoria de substância. Ou seja, o autor procurou construir um vocabulário adequado aos eventos mentais opondo estes aos eventos do mundo físico (apresentando caracterização negativa), de maneira que podemos encontrar eventos corporais do tipo movimento de membros causados por leis mecânicas e, ao mesmo tempo, encontrar outros movimentos com causas não-mecânicas originárias de um trabalho da mente.

Ryle (1970, p.20) demonstra na expressão seguinte a preocupação acima com a dificuldade que se apresenta como consequência do erro categorial :

Dado que, de acordo com a doutrina cartesiana, a mente pertence à mesma categoria que o corpo e dado que o corpo é rigidamente governado por leis mecânicas, pareceu acertado a muitos teóricos que a mente fosse similarmente governada por leis rígidas não-mecânicas. O mundo físico é um sistema determinista, tal como o mundo mental o deve ser também.

Na opinião de Ryle (1970), Descartes (1996) ao caracterizar o corpo como uma máquina, se comportando de acordo com leis mecânicas, estabelece um outro problema, qual seja, como explicar a questão da ação

humana livre através de uma teoria mecanicista do corpo. Na concepção de Gonzales (1998, p. 91), para tornar a ação humana possível, Descartes propõe juntar ao corpo uma substância não material, pensante, que não estaria sujeita a causalidade mecânica, ficando responsável pelo agir, pensar, desejar, sentir e outros atributos mentais-comportamentais.

Resumindo, Ryle (1970) pretende dissolver o problema mente-corpo argumentando que este seria fruto do erro categorial cometido por Descartes através de uma análise lingüística dos termos utilizados para definir a mente. O percurso teórico desse autor segue uma abordagem contrária à adotada pelas correntes que ora privilegiam o mental ora o físico, na procura de uma solução para o problema mente-corpo, como fazem as correntes Idealista e Materialista. Para Ryle (1970), tanto a abordagem materialista como a idealista só reforçam a concepção sobre a existência das duas substâncias.

Na seção seguinte apresentamos a teoria funcionalista como uma proposta de solução para o problema da relação mente-corpo.

Block (1980 p.172) exemplifica o que seria o evento de dor, segundo o funcionalismo metafísico:

"... pain might characterize pain in part in terms of its tendency to be caused by tissue damage, by its tendency to raise the desire to be rid of it and by its tendency to produce action designed to separate the damaged part of the body from what is thought to cause the damage."

Seção II – Funcionalismo computacional

O filósofo e matemático Hilary Putnam (1995, p.441) desenvolveu em torno de 1960 uma teoria sobre a natureza dos fenômenos mentais denominada de funcionalismo. Juntamente com o behaviorismo e a teoria da identidade, ela constitui mais uma das correntes filosóficas que se preocupam em investigar a natureza dos processos mentais apoiando-se em uma analogia entre a mente e o computador digital. Essencialmente, o funcionalismo enfatiza que o ponto relevante no estudo dos fenômenos mentais é o seu aspecto funcional e não a matéria da qual seriam constituídos.

De maneira geral, a teoria da identidade dos estados funcionais ou funcionalismo sustenta que estar em um estado mental é estar num estado funcional. O estado funcional é definido em termos de suas relações causais entre : a) estímulos do meio ambiente ('inputs'), b) comportamento do organismo ou sistema ('output') e c) outros estados mentais. Block (1980,p.172) exemplifica o que seria o evento de dor, segundo o funcionalismo metafísico :

"... pain might characterize pain in part in terms of its tendency to be caused by tissue damage, by its tendency to cause the desire to be rid of it, and by its tendency to produce action designed to separate the damaged part of the body from what is thought to cause the damage."

Block (1980, p. 171) identifica, na literatura, três versões de funcionalismo. O primeiro refere-se à noção de *análise funcional*, a qual estuda o comportamento de um sistema através da sua decomposição em partes. Neste caso, o comportamento do sistema é determinado por meio do comportamento das partes e da integração entre elas. O segundo é denominado de *funcionalismo computacional*. Esta linha sustenta que a mente pode ser comparado a um programa de computador e os processos mentais seriam manipulações ou computações de símbolos semelhantes às que ocorrem em um computador digital. Por último, o termo funcionalismo pode referir-se ao chamado *funcionalismo metafísico*. Para Block (1980, p.172), este tipo de funcionalismo é o que realmente está preocupado em responder questões sobre a natureza dos estados mentais, ou seja, questões do tipo : " *What is pain ? - or at least to What is there in common to all pains in virtue of which they are pains ?* " .

Algumas similaridades entre as três versões do funcionalismo podem ser destacadas. A mais significativa delas é a ligação com o conceito de máquina de Turing. No caso do funcionalismo metafísico, postula-se que os estados mentais são idênticos aos estados da tabela de estados ("machine table") de uma máquina de Turing. Portanto, para um melhor entendimento dos conceitos que envolvem o funcionalismo é fundamental entender as propriedades da máquina de Turing.

Desenvolvido por Alan Matson Turing em 1936, o conceito que leva o seu nome – "Máquina de Turing" - consiste em um mecanismo

imaginário que seria formado por : a) uma fita infinita dividida em células discretas, b) um dispositivo para “ler”, “gravar” ou “apagar” símbolos (“0” ou “1”) em uma determinada célula da fita e movimentar-se ao longo dela e c) uma lista finita de estados internos, razão pela qual a máquina de Turing é chamada de máquina de estados finitos ou autômato finito. Funcionalmente, a máquina de Turing comporta-se segundo as instruções contidas em sua “machine table”, ou seja, um conjunto de condições que descrevem as relações entre os “inputs” (leitura de um símbolo) , seus estados internos e “outputs”. Assim, uma máquina de Turing pode se encontrar em um determinado estado; receber uma entrada, mudar para outro estado de sua lista interna ou permanecer no estado em que se encontra e emitir um símbolo de saída.

Esquemáticamente, o aspecto funcional da máquina de Turing pode ser representado através de dois diagramas. O primeiro é denominado de *tabela de transição de estados*, mostrada a seguir:

Lista de Estados da Máquina				
1	2	3	4	5 6 7
Machine Table				
Estado	Leitura/ Gravação	Escrever	Mover p/	Ir p/ estado
1	0	1	Esquerda	3
1	1	0	Direita	2
2	0	0	Direita	1
2	1	1	Esquerda	3

Dispositivo de Leitura e Gravação

0	0	1	1	0	1	1	1	0	1	1
---	---	---	---	---	---	---	---	---	---	---

Fita

FIGURA 2: Tabela de transição de estados de uma máquina de Turing (Stillings, 1987,p.322)

De acordo com a "machine table" da Figura 2, se a máquina de Turing se encontra no estado 1 e "lê", a partir da fita o símbolo 1, seu comportamento será o de escrever o símbolo 0 sobre a célula da fita e ir para um próximo estado (estado 2), movimentando-se para a direita ao longo da fita.

De maneira genérica, Block define a máquina de Turing da seguinte forma : uma máquina que contém uma lista de estados internos ($S_1, \dots, S_n$) em alguns dos quais ela pode se encontrar; uma lista finita de "inputs" que a máquina pode receber ($I_1, \dots, I_m$) e uma lista finita de "outputs" que a máquina pode emitir ($O_1, \dots, O_p$).

A "machine table" da máquina de Turing "especifica o conjunto de condições de forma que : se uma máquina está em um estado S_i e recebe um "input" I_j , ela emite um "output" O_k e vai para um estado S_l . Isto é, dado qualquer estado e um estímulo, a tabela especifica um "output" e o próximo estado da máquina de Turing" (Block,1980,p.271).

Esta definição genérica da máquina de Turing pode ser exemplificada por meio da máquina para compra de refrigerante descrita pelo seguinte diagrama :

	S1	S2
Moeda de 5 centavos	Não emite o refrigerante Vai para estado S2	Emite o refrigerante vai para estado S1
Moeda de 10 centavos	Emite o refrigerante Permaneço no estado S1	Emite refrigerante e uma moeda de 5 centavos Vai para o estado S1

FIGURA 3: "Machine table" para uma instância de uma máquina de Turing. Adaptação do exemplo de "Coke-machine" (Block,1980,p.173)

A máquina que se comporta de acordo com tabela acima possui uma lista com dois estados, S_1 e S_2 ; uma lista com dois "inputs" na qual a máquina recebe uma moeda de dez centavos e outro em que a máquina aceita moeda de cinco centavos; uma lista de "outputs" no qual a máquina emite um guaraná e outro no qual emite um guaraná e uma moeda de cinco centavos. Através deste exemplo, nota-se que qualquer que seja o significado dos estados ' S_1 ' e ' S_2 ' eles se originam a partir da *relação* entre um estado e outro, 'inputs' e 'outputs', como especificados pela tabela, não sendo relevante o material da qual a máquina seria constituída.

De acordo com Block (1980), nenhum teórico da visão funcionalista sustentaria que, no exemplo acima, a máquina deseja qualquer coisa; isto seria apenas uma maneira de falar. O que poderia ser reivindicado é que S_1 seria idêntico com desejo de moeda de 10 centavos e S_2 idêntico com desejo de moeda de 5 centavos.

A versão do funcionalismo apoiado em máquina de Turing é chamada de funcionalismo computacional e possui como uma de suas deficiências o fato de que alguma coisa pode estabelecer relações causais do tipo dor, sem estar, por exemplo, em um estado de dor.

A seguir descreveremos algumas dificuldades e objeções mais comuns encontradas na teoria funcionalista. A primeira diz respeito ao liberalismo na sua caracterização dos estados mentais. Uma teoria é considerada liberal no sentido que ela atribui propriedades a certos organismos ou sistemas quando de fato eles não possuem tais

propriedades. No caso do funcionalismo, para um sistema exibir estados mentais, bastaria que ele exibisse certa organização funcional, portanto, os estados mentais poderiam se manifestar tanto em seres humanos quanto em máquinas. A dificuldade reside exatamente na possibilidade de atribuir estados mentais do tipo dor, por exemplo a uma máquina, quando sabemos que ela não sente dor.

Uma outra dificuldade encontrada pelo funcionalismo refere-se ao problema do chauvinismo. De acordo com Block (1980, p.291) a visão funcionalista é considerada chauvinista na medida em que nega a existência de estados mentais em sistemas que, aparentemente, possuem tais estados. Esse autor sugere que os funcionalistas possuem a tendência em definir os 'inputs' e 'outputs' em termos de *nossos* 'inputs' e 'outputs', como por exemplo, quando especificamos 'outputs' em termos de movimentação de braços e 'inputs' através de uma luz incidindo sobre os nossos olhos. Neste sentido, aqueles seres que não têm esta capacidade de movimentação não teriam estados mentais. Em resumo, Block (1980,p.294) sugere que "o funcionalismo, ao tentar evitar o problema do liberalismo, pode levar automaticamente ao chauvinismo, caindo em um dilema, qual seja: existiria uma descrição de 'inputs' e 'outputs' específica o suficiente para evitar o liberalismo e também para evitar o chauvinismo ? Block conclui que o funcionalismo não poderia evitar o chauvinismo e o liberalismo ao mesmo tempo.

Outra dificuldade encontrada pelo funcionalismo é o problema dos qualia. A noção de qualia é utilizada para evidenciar aspectos



intrínsecos que supostamente estariam presentes em alguns processos mentais e que seriam obtidos por meio de um processo introspectivo. Os estados mentais possuidores dessa propriedade seriam denominados de estados mentais qualitativos.

De acordo com Nagel (1980), as teorias filosóficas sobre a mente, em particular o funcionalismo, não conseguem explicar o caráter subjetivo da experiência ou qualia, por meio de relações causais. Esta afirmação é ressaltada por Armstrong (1984, p.169) da seguinte forma : *"But many philosophers have argued that introspection reveals that mental phenomena involve a qualitative component not to be captured by any analysis in terms of mere causal role"*.

Outra argumento utilizado para apontar as limitações do funcionalismo com relação aos qualia é o problema do espectro invertido (*inverted qualia*). Tal noção refere-se à idéia de que um mesmo objeto poderia aparentar a cor vermelha para um observador e a cor verde para outro e vice-versa. Para alguns pesquisadores, uma teoria causal da mente que não permite obter a qualidade experienciada de forma introspectiva não poderia explicar a experiência com cores fenomenologicamente invertidas.

Em resumo, a questão dos qualia constitui um desafio ao funcionalismo, embora existam questionamentos sobre a própria existência desse fenômeno. Supondo a existência de estados mentais com propriedades qualitativas , o funcionalismo encontra-se diante da

difficuldade em apresentar explicações convincentes sobre a natureza de tais estados.

A sua importância no estudo da mente cognitiva pode ser positiva e negativa. O aspecto positivo diz respeito à contribuição para a fixação da noção de organização funcional, que permitiu distinguir uma estrutura abstrata de uma estrutura concreta, sendo que as propriedades relevantes, desta distinção, não são as propriedades físico-químicas. O aspecto negativo do funcionalismo diz respeito à maneira simplificada com que a noção de organização funcional é tratada no estudo da complexidade da mente.

A indagação a ser feita, ao se buscar explicações para a natureza da mente cognitiva deve considerar a seguinte questão: Qual a nossa forma (função) intelectual ? e não se somos constituídos somente de matéria ou alma e matéria (Putnam, 1980).

Na próxima seção mostraremos a relevância dessa concepção funcional na caracterização das duas vertentes da ciência cognitiva.

Seção III - Ramificações da ciência cognitiva: inteligência artificial e redes neurais artificiais

O período de 1950 a 1970 é geralmente conhecido como aquele em que aconteceu um rápido desenvolvimento da ciência cognitiva. Este desenvolvimento foi fortemente alavancado pelo surgimento de um novo campo de pesquisas que se desenvolveu apoiado nas possibilidades científicas abertas a partir do conceito teórico de máquina de Turing, e da invenção do computador com arquitetura Von Neumann em 1941. Este campo de pesquisas foi formalizado com o nome de *Inteligência Artificial* e tinha como proposta fundamental a possibilidade de simular o comportamento inteligente em um computador digital, proposta esta fundada no funcionalismo computacional que apresentamos na seção anterior.

Antigas reflexões filosóficas sobre a natureza do pensamento humano chegaram a conclusões que o caracterizaram como sendo de natureza computacional. Para Dascal (1998, p.1) Hobbes deixa claro, em sua obra, o *Leviatã*, que o ato de raciocinar é semelhante ao ato de computar e enfatiza que a ele pode ser atribuído "... não somente a proposta de uma teoria computacional da mente, mas também uma definição – contemporânea – do espaço de problemas em que atualmente a ciência cognitiva opera".

De acordo com Charniak & McDermont (1987, p.10), George Boole (1815-1864) formalizou em sua obra *An investigation of the*

laws of thought on which are founded the mathematical theories of logic and probabilities, de 1854, uma série de princípios lógicos, a qual representava uma versão mecanizada do pensamento. As proposições lógicas Booleanas podiam assumir apenas dois valores, ou seja, elas são verdadeiras ou falsas e podem ser representadas por um sistema binário. A hipótese da adequação das descrições dos padrões do pensamento humano por expressões que possam assumir valores 0 ou 1 viabilizava a realização de operações aritméticas binárias e foram fundamentais para que nos anos 30 e 40 se construíssem circuitos elétricos que resultaram no computador digital: um pressuposto fundamental na proposta da Inteligência Artificial para simular o comportamento inteligente.

Minsky (1976, p.151) ressalta que especulações teóricas realizadas por pesquisadores do campo da cibernética a respeito da possibilidade de "...converter descrições de processos de pensamento em prescrições para o projeto de máquinas ou, para o projeto de programas", impulsionaram o aparecimento do campo da IA. Esse autor ressalta que Wiener, Rosenbluth e Bigelow trabalharam na produção máquinas capazes de atingirem metas e propósitos e, McCulloch & Pitts sugeriram a possibilidade das máquinas utilizarem conceitos de lógica.

Entretanto, a partir de 1956, o campo da Inteligência Artificial ultrapassou a barreira da especulação teórica e passou a constituir formalmente um programa de pesquisas científicas, quando pesquisadores reunidos no Dartmouth College discutiram a "possibilidade de construir máquinas genuinamente inteligentes" (Minsky, 1977, p.151).



A partir deste momento, um rico aparato conceitual foi desenvolvido para tornar o sonho da Inteligência Artificial uma realidade. Assim, foram concebidas novas tecnologias, tais como: linguagens de programação específicas para manipulação de símbolos e uma variedade de programas, para computadores, com o propósito de simular os mais variados aspectos da inteligência humana, desde programas para provas de teoremas até programas para jogar xadrez.

Na opinião de Bender (1996), o campo da Inteligência Artificial surgiu de maneira autônoma ou independente da ciência cognitiva. Porém, na opinião de Gonzales (1991), a IA compõe o programa interdisciplinar da ciência cognitiva com foco de análise na simulação computacional do comportamento inteligente.

O modelo de IA está fortemente vinculado à noção de procedimento efetivo da máquina de Turing. Segundo tal idéia, o comportamento inteligente, quando descrito por um procedimento efetivo, pode ser programado em um computador digital. A partir de tal pressuposto, os teóricos da IA construíram linguagens de programação específicas para manipulação de símbolos e a consequente simulação dos aspectos da inteligência, tais como, a linguagem LISP, projetada por McCarthy (Gardner, 1995, p.170).

Minsky (1977, p.151) estava interessado em criar máquinas inteligentes sem se preocupar com o comportamento especificamente humano. Segundo Gardner (1995, p. 167) alguns de seus alunos produziram programas para realizar analogias visuais, conhecendo de

antemão alguma relação entre duas imagens representadas por símbolos. Através de relações do tipo A está para B, o programa aplica regras lógicas para concluir que C está para D e programas para solução de problemas de álgebra de segundo grau seguindo os princípios da Inteligência Artificial com o nome de *STUDENT*.

Deve-se a Minsky uma das mais utilizadas definições do objetivo da Inteligência Artificial, qual seja "*artificial intelligence is the science of making machines do things that would require intelligence if done by men*" (Minsky apud Boden, 1977, p.4). De acordo com Boden (1977, p. 4) a definição de Minsky enfatiza a preocupação geral do campo da IA.

Gardner (1995) ressalta que Newell & Simon, foram dois dos precursores e mais ativos pesquisadores da IA. Atuando no nível mais prático, os trabalhos e programas produzidos por eles exerceram forte influência para tornar viável o campo da IA. Segundo Minsky, as investigações realizadas por Newell & Simon enfocavam o desenvolvimento de "modelos do comportamento humano" (Minsky, 1977, p. 151). Basicamente, as pesquisas de Newell & Simon seguem o princípio de que a atividade cognitiva humana pode ser explicada através do processamento simbólico semelhante ao que ocorre em um computador, principalmente as atividades de resolução de problemas. Assim, da grande cooperação científica entre eles resultaram dois dos principais programas que se tornaram paradigmáticos no campo da IA. O primeiro foi o *Logic Theorist*, que destinava-se a demonstrar teoremas lógicos-matemáticos contidos na

obra *Principia Mathematica* de Whitehead e Russell (Bender, 1996, p.27). O *Logic Theorist* foi, basicamente, estruturado através de um conjunto de expressões lógicas do tipo : se A implica B e se B implica C então A implica C; além disso, o programa funcionava “substituindo uma expressão por outra; um método de destacamento, onde o programa trabalha de trás para frente, de algo que já foi provado para algo a ser provado” (Gardner, 1995, p.162).

Explorando a capacidade do computador em manipular símbolos, Newell & Simon realizaram estudos mais profundos procurando identificar quais tarefas tipicamente humanas poderiam ser resolvidas através da manipulação simbólica, tais como o jogo de xadrez e a resolução de problemas geométricos (Gardner, 1995, p.161). O propósito de Newell & Simon era demonstrar que o computador poderia ser usado para exibir aspectos inteligentes particulares dos seres humanos e a confiança nesta possibilidade era tão grande que pode ser resumida nas seguintes palavras de Simon : “Durante o natal eu e Allen Newell inventamos uma máquina pensante” (Gardner, 1995, p.161).

Pesquisando, particularmente, a natureza, a estrutura e a complexidade dos processos humanos de resolução de problemas., Newell & Simon avançaram para criar um programa de alcance mais abrangente que pudesse resolver os mais variados e complexos problemas e foi denominado de *General Problem Solver* . Procurando imitar os seres humanos em suas atividades de resolução de problemas o GPS, como ficou conhecido, segue uma ordem lógica baseada nos seguintes passos;

descritos em Gardner (1995, p.164): inicia-se com uma proposta preliminar de solução do problema e compara-a com o objetivo desejado. Se a solução proposta se assemelha ao objetivo desejado a solução do problema foi alcançada, caso contrário, o GPS procurará diminuir a diferença entre o estado atual e o estado final desejado.

Gardner (1995, p.164) ressalta, ainda, que o GPS pode ser considerado o primeiro programa a simular a característica de manipulação simbólica dos seres humanos fora dos seres humanos.

Em resumo, Newell & Simon viabilizaram a possibilidade de se tornar a IA um campo de pesquisas e também uma teoria da mente concebendo-a como um sistema processador de informações. Neste sentido o comportamento inteligente seria aquele que gera representações e realiza operações sobre estas representações ou símbolos. Assim, a arquitetura de um sistema inteligente seria aquela capaz de receber *inputs*, de realizar um conjunto de operações, de armazenar e recuperar símbolos a partir de ou para localizações específicas, de modificar símbolos e transmitir *outputs* ou padrões de comportamento de acordo com uma ordem lógica previamente estabelecida. Assim, o comportamento inteligente poderia ser incorporado na seguinte estrutura :



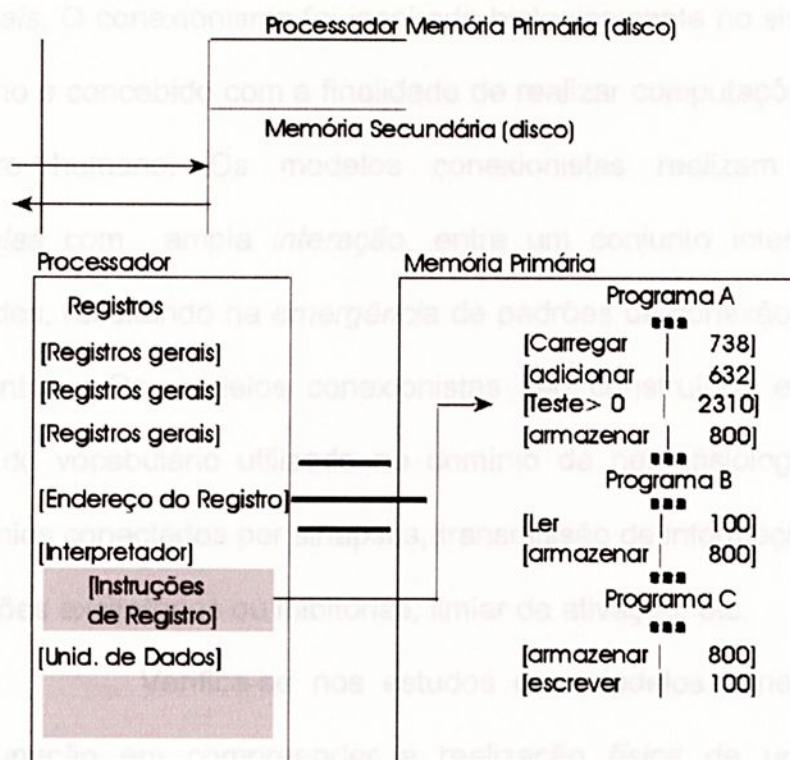


FIGURA 4 : Estrutura de um computador digital simples (Newell, 1990,p.95)

A Figura 4 exhibe os principais mecanismos que compõem a arquitetura computacional na qual os processos inteligentes poderiam ser instanciados aproveitando-se da capacidade para manipular símbolos do computador digital. Esta mesma estrutura computacional serviu de inspiração para a abordagem do problema da relação mente-corpo (conforme veremos no capítulos III e IV) através da relação *software-hardware* contudo, críticas demolidoras a este empreendimento, juntamente com limitações do modelo de inteligência artificial deram lugar a uma nova ramificação da ciência cognitiva.

O modelo cognitivo alternativo ao da IA, amplamente estudado no âmbito da ciência cognitiva, é denominado *Conexionismo* ou *Processamento Distribuído em Paralelo* (PDP) ou ainda, *Redes Neurais*

Artificiais. O conexionismo foi inspirado biologicamente no sistema nervoso humano e concebido com a finalidade de realizar computações no estilo do cérebro humano. Os modelos conexionistas realizam computações *paralelas* com *ampla interação*, entre um conjunto interconectado de unidades, resultando na *emergência* de padrões de conexão entre os seus elementos. Os modelos conexionistas são construídos empregando-se parte do vocabulário utilizado no domínio da neurofisiologia; tal como : neurônios conectados por sinapses, transmissão de informações através de conexões excitatórias ou inibitórias, limiar de ativação, etc.

Verifica-se nos estudos dos modelos conexionistas uma preocupação em compreender a realização *física* de um sistema de informações. Em seguida constroem-se modelos apoiados nas características inspiradas no substrato físico, o cérebro, quando este processa informações.

Do ponto de vista filosófico, os modelos conexionistas encontram sua plausibilidade na teoria da identidade mente-cérebro, particularmente em seu ramo da teoria da identidade de tipos, a qual sustenta que para cada tipo específico de estado mental corresponde um tipo específico de estado físico no cérebro.

Os cientistas cognitivos avaliam alguns modelos conexionistas como uma possível arquitetura para a mente humana, por meio dos quais os processos cognitivos seriam tratados como processos emergentes. A abordagem conexionista é vista como uma forte alternativa aos modelos de processamento serial, digital e de manipulação simbólica



adotados pela inteligência artificial. Ao passo que, em tal abordagem, o estudo da cognição está focalizado em seu aspecto mais abstrato, denominado de nível simbólico, a abordagem conexionista está centralizada no aspecto físico ou microscópico, denominado de nível sub-simbólico. Arbib (1987, p.92-3) enfatiza que o modelo conexionista procura abstrair o caráter computacional do cérebro na tentativa de modelar a mente humana e construir máquinas inteligentes.

Rumelhart (1989, p. 180), cujos trabalhos contribuíram para renovar o interesse surgido durante os anos 50 e 60 pelo modelo de redes neurais, destaca a importância e a mudança conceitual introduzida pela abordagem conexionista :

The architecture that we are exploring is not one based on the von Neuman architecture of our current generation of computers but rather an architecture based on considerations of how brains themselves might function. Our strategy has thus become one of offering a general and abstract model of the computational architecture of brains, to develop algorithms and procedures well suited to this architecture, to simulate these procedures and architecture on a computer, and explore them as hypotheses about the nature of human information-processing system. We say that such models are neurally inspired, and we call computation on such a system brain style computation. Our goal in short is to replace the computer metaphor with the brain metaphor.

Na perspectiva da ciência da computação, o interesse pelos modelos conexionistas se justifica dado o esgotamento e limitações

impostas pela arquitetura computacional do tipo Von Neuman, caracterizada por um processamento serial e digital. Nos modelos connexionistas, a esperança é que ocorra um poderoso aumento da capacidade computacional, que viabilizaria uma série de tarefas de difícil execução na arquitetura tradicional.

Do ponto de vista biológico, conforme já ressaltamos anteriormente, os modelos connexionistas foram inspirados em algumas propriedades do sistema nervoso humano. A propriedade essencial do sistema nervoso explorada pelos teóricos dos modelos connexionistas é a capacidade de receber, processar e transmitir *informação* por meio de uma grande rede de neurônios densamente interconectados. De acordo com Wasserman (1989, p. 12) estima-se em torno de 10^{11} neurônios e 10^{15} o número de interconexões entre estas células. Em sua estrutura as células neuronais são formadas pelos seguintes elementos, mostrados na Figura 5: os dendritos, o soma ou corpo celular, o axônio e as sinapses. Os dendritos estendem-se a partir do corpo celular para outros neurônios. As sinapses efetuam a conexão entre os dendritos e o axônio de outro neurônio por meio da liberação de substâncias denominadas de *neurotransmissores*. Os axônios são responsáveis pela transmissão dos impulsos elétricos a partir do corpo celular até outros neurônios. Estes impulsos tendem a formar conexões inibitórias ou excitatórias dependendo do limiar de ativação ou *threshold* do neurônio. Quando a soma dos impulsos nervosos recebidos por um neurônio ultrapassa o seu limiar de

ativação, um impulso nervoso é formado no axônio e transmitido por meio das conexões até outros neurônios do sistema nervoso.

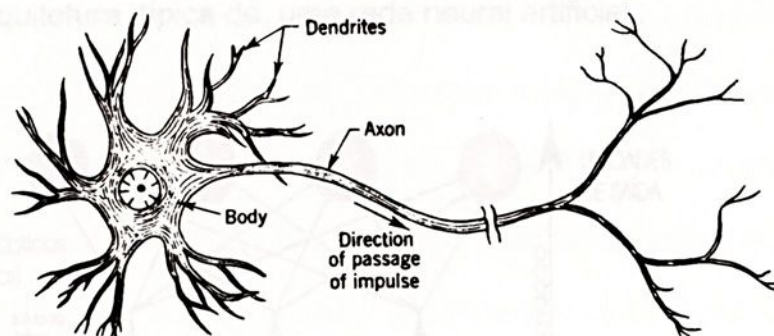


FIGURA 5 - Desenho esquemático de um neurônio (Arbib,1989,p.12)

Conforme ressaltamos, a organização biológica do sistema nervoso serviu de inspiração para a concepção lógica do modelo de redes neurais artificiais. Estruturalmente, os modelos de redes neurais artificiais reúnem um grande número de elementos básicos organizados em camadas. Os elementos básicos, normalmente denominados de neurônios-simile, nódulos, neurônios formais, unidades de processamento ou *neuron-like units* formam as unidades lógicas da rede.

As camadas de uma rede neural artificial são divididas em três classes. A camada de elementos de *entrada*, que recebe os padrões informacionais do meio ambiente, a camada de elementos de *saída* que fornece uma resposta em função do padrão de entrada e a transmite ao meio exterior, e uma ou mais camadas de elementos intermediários ou camadas *escondidas*. A camada escondida não possui contato com o meio externo e efetua a associação entre os elementos da camada de entrada e os elementos da camada de saída. As unidades da camada escondida

integram os sinais recebidos das unidades da camada de entrada após serem mediados pela força ou peso da conexão entre elas. A Figura 6 representa a arquitetura típica de uma rede neural artificial :

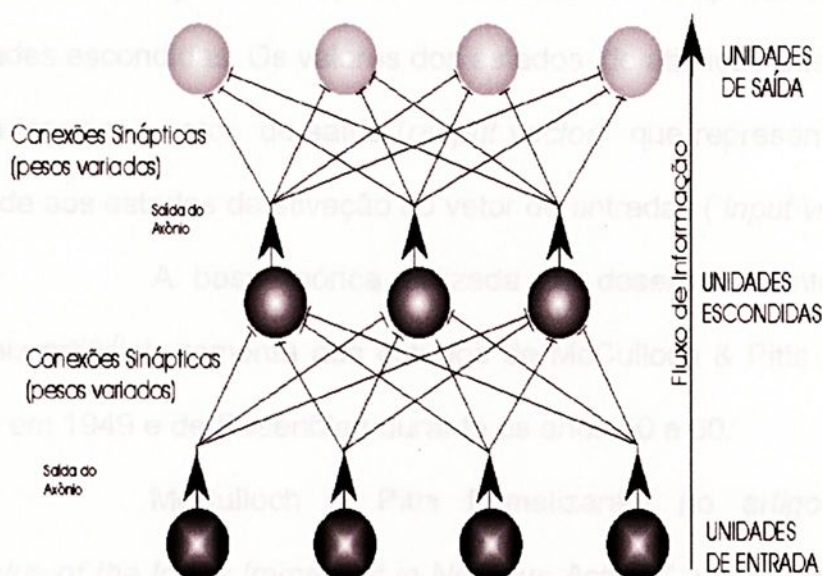


FIGURA 6 - Uma rede neural simples (Adaptado de Churchland (1993, p.162))

Quanto ao fluxo de informação no modelo de rede neural artificial mostrado na Figura 6, Churchland (1993, p. 161) enfatiza que ele se inicia com cada unidade da camada de entrada representando algum aspecto do padrão de *input* global do meio ambiente. O padrão de entrada global ou de *input* é formado através da ativação simultânea de todas as unidades da camada de entrada e se distribui, por meio das conexões sinápticas, na direção de todas as unidades da camada escondida ou *hidden units*. As unidades da camada escondida ao receberem os estados de ativação individuais das unidades de entrada computam seus novos estados individuais apoiando-se nos estados de ativação recebidos e no peso das conexões sinápticas. Os novos estados de ativação, obtidos pelas

unidades da camada escondida, são enviados na direção de todas as unidades da camada de saída. As unidades da camada de saída determinam seus novos estados de ativação individuais apoiando-se nos estados de ativação e nos pesos das conexões sinápticas recebidos das unidades escondidas. Os valores dos estados de ativação das unidades de saída formam o vetor de saída (*output vector*) que representa a resposta da rede aos estados de ativação do vetor de entradas (*input vector*).

A base teórica utilizada no desenvolvimento das redes neurais artificiais remonta aos estudos de McCulloch & Pitts em 1943, de Hebb em 1949 e de Rosenblatt durante os anos 50 e 60.

McCulloch & Pitts formalizaram no artigo "*A Logical Calculus of the Ideas Immanent in Nervous Activity*" um modelo lógico do funcionamento de uma rede neural biológica. O desenvolvimento de McCulloch & Pitts se apoiou na criação de um modelo de neurônio artificial a partir da hipótese de que o neurônio biológico possui um comportamento binário, tipo tudo-ou-nada. Assim, a atividade do sistema nervoso foi modelada de tal maneira que o comportamento da rede neural biológica poderia ser simulada através de um conjunto de neurônios formais capazes de estabelecer relações entre operações lógicas.

As redes neurais artificiais contemporâneas incorporam as propriedades essenciais do modelo de McCulloch & Pitts, tais como o conceito de limiar de ativação ou *threshold*, comportamento excitatório ou inibitório e a idéia de conexões entre neurônios funcionando em rede.

De acordo com Dupuy (1996,p.59), o trabalho de McCulloch & Pitts contribuiu para fundamentar o campo da ciência cognitiva ao postular não somente um modelo idealizado da estrutura anatômica e fisiológica do cérebro, mas também por sugerir que a mente é uma função deste e operando segundo proposições lógicas. Dupuy sugere que McCulloch & Pitts incorporaram no modelo idealizado apenas os aspectos funcionais do sistema nervoso, adotando uma estratégia funcionalista diferente do funcionalismo computacional presente na máquina de Turing e adotada pelo modelo de Inteligência Artificial.

Hebb sugeriu, no seu trabalho *The Organization of Behavior*, um princípio de aprendizado biológico apoiado no aumento da intensidade das conexões sinápticas entre neurônios. Essencialmente, Hebb (1990, p.50) enfatiza que quando a intensidade do impulso da conexão sináptica entre um neurônio A e um neurônio B se mantém constantemente ativo, a intensidade da conexão entre eles se reforça durante este período. O princípio de Hebb foi aproveitado para formular regras de *aprendizado* para as redes neurais artificiais.

Rosenblatt (1962) elaborou uma rede neural artificial, denominada de *Perceptron*, cujo princípio de funcionamento é utilizado nas redes neurais atuais e destinava-se a reconhecer padrões visuais. O *Perceptron* de Rosenblatt consistia de um conjunto de neurônios de McCulloch & Pitts estruturados em camadas : a camada de neurônios de entrada ou *sensory units* responsável por receber os padrões do meio externo, a camada de associação responsável por processar o fluxo de

informação através da rede e a camada de saída ou *response units* que representam o padrão de saída resultante do funcionamento da rede e devolvido ao meio exterior. Contrastando com o modelo de McCulloch & Pitts, Rosenblatt demonstrou que o *Perceptron* poderia aprender a reconhecer padrões através da modificação dos pesos das conexões entre as camadas da rede. Assim, o *Perceptron*, em princípio poderia ser treinado para modificar continuamente o valor das conexões apoiando-se no valor das conexões passadas e apresentando, assim, um aprendizado a partir da experiência.

Outras propriedades incorporadas nos modelos de redes neurais artificiais e inspiradas no funcionamento do cérebro foram a noção de *processamento distribuído* e a idéia de *graceful degradation*. Para Beale & Jackson (1994, p.12) o cérebro é um sistema que funciona de maneira massivamente paralela, cujas funções são compartilhadas entre suas unidades de processamento (neurônios). Este compartilhamento de tarefas é conhecido como *processamento distribuído*. Jackson sugere que o processamento distribuído permite ao cérebro ser tolerante a falhas, durante o decurso de seu funcionamento, já que pode manter-se em operação mesmo com algumas de suas unidades inativas. A propriedade denominada de *graceful degradation* ou *degeneração gradual* permite ao cérebro manter suas funções básicas mesmo quando ocorre um decréscimo em seu desempenho provocado por algum dano em sua estrutura.



Em resumo, analisamos nesta seção os princípios gerais que caracterizam os modelos utilizados no âmbito da ciência cognitiva para explicar a natureza dos processos cognitivos: a inteligência artificial e as redes neurais artificiais. Enfatizamos que o modelo de inteligência artificial se apóia nas noções da máquina de Turing e do funcionalismo para explicar a natureza da atividade cognitiva sem ressaltar seus aspectos físicos. Ao contrário do modelo de inteligência artificial, as redes neurais artificiais procuram ressaltar, além das características funcionais, os aspectos físicos dos processos cognitivos. A análise dos modelos de redes neurais artificiais será completada no próximo capítulo através da descrição de um modelo matemático geral e dos seus métodos de aprendizagem.

Apesar do grande interesse atual despertado pelas redes neurais artificiais, esta área permaneceu durante muito tempo inibida pelas demonstrações feitas por Minsky & Papert (1970) acerca das limitações dos *Perceptrons* na atividade de reconhecimento de padrões. Assim, a ciência cognitiva foi dominada em seus primeiros anos pelo paradigma simbólico da inteligência artificial. O campo das redes neurais artificiais somente foi retomado nos anos 80, com o trabalho de Hinton, Rumelhart e McClelland, entre outros, sugerindo a viabilidade dos modelos dessa área no estudo dos processos cognitivos. No que se segue completaremos a análise dos modelos de redes neurais artificiais.



Apresentação

Neste capítulo analisamos o papel dos modelos na explicação dos processos cognitivos. A prática da modelagem é uma estratégia fundamental empregada pelos cientistas cognitivos para tentar simular, em sistemas artificiais, o comportamento inteligente, na expectativa de poder explicá-lo. A simulação dos aspectos cognitivos, através de modelos, tem por objetivo extrair conceitos e hipóteses que formam a

CAPÍTULO III – O PAPEL DOS MODELOS NA EXPLICAÇÃO DA MENTE

seus limites e alcance». Nos seus primeiros anos, predominou na ciência a suposição acerca da analogia entre a mente e o computador. Tal analogia direcionou a prática de elaboração de modelos para explicação dos processos cognitivos e mais recentemente, em oposição a esta visão, as redes neurais artificiais têm sido aplicadas como um modelo alternativo. O papel e a natureza dos modelos na ciência cognitiva são estudados neste capítulo em três seções.

Na seção I, enfatizamos os aspectos conceituais envolvidos na construção dos modelos.

Na seção II, retomamos o estudo do modelo de Redes Neurais Artificiais apresentado no capítulo anterior, procurando ressaltar seus princípios matemáticos e mecanismos de aprendizagem.

Apresentação

Neste capítulo analisamos o papel dos modelos na explicação dos processos cognitivos. A prática da modelagem é uma estratégia, fundamental, empregada pelos cientistas cognitivos para tentar simular, em sistemas artificiais, o comportamento inteligente, na expectativa de poder explicá-lo. A simulação dos aspectos cognitivos, através de modelos, tem por objetivo extrair conceitos e hipóteses que formarão a base explicativa da mente cognitiva.

Apesar de sua ampla utilização, os modelos centrais utilizados pela ciência cognitiva são freqüentemente criticados quanto aos seus limites e alcances. Nos seus primeiros anos, predominou na ciência a pressuposição acerca da analogia entre a mente e o computador. Tal analogia direcionou a prática de elaboração de modelos para explicação dos processos cognitivos e mais recentemente, em oposição a esta visão, as redes neurais artificiais têm sido aplicadas como um modelo alternativo. O papel e a natureza dos modelos na ciência cognitiva será abordado neste capítulo em três seções:

Na seção I, enfatizamos os aspectos conceituais envolvidos na construção dos modelos.

Na seção II, retomamos o estudo do modelo de Redes Neurais Artificiais, apresentado no capítulo anterior, procurando destacar seus princípios matemáticos e mecanismos de aprendizagem.



Seção I : O papel Na última seção, abordamos os limites e alcances dos modelos de inteligência artificial e de redes neurais artificiais no estudo da mente.

A ciência procura formular teorias que possam fornecer explicações, previsões e controlar o comportamento dos mais diversos fenômenos, sejam eles naturais ou sociais. Para cumprir este objetivo, ela se utiliza de vários recursos e, entre eles, da observação sistemática do comportamento de um organismo, experimentação e elaboração de modelos.

Especialmente no campo da ciência cognitiva, o método dominante e, talvez, o mais fundamental, é o método da modelagem (*modelling*). Basicamente, modelar um fenômeno natural consiste em representar, como afirma Dupuy (1996, p.22), imitar, repetir ou reproduzir aspectos funcionais relevantes desse fenômeno, incorporando-o a alguma teoria. A prática da modelagem na ciência cognitiva consiste em realizar abstrações do fenômeno real, incorporando ao modelo computacional somente as relações funcionais, desprezando-se, assim, a natureza dos elementos fisiológico-biológicos que constituem o processo cognitivo.

Dupuy (1996, p.27), ao analisar a utilização de modelos para gerar hipóteses plausíveis a respeito dos processos cognitivos, enfatiza que o papel desempenhado por eles é equivalente à função de conhecer:

Conhecer é produzir um modelo do fenômeno a partir de dados e manipulá-lo ordenado. Todo conhecimento é reprodução, representação, repetição, simulação. Isso caracteriza o modo científico.

¹ Agradecemos à Profª Carmen B. Miodini pelo auxílio que nos fez refletir sobre o conceito de simulação durante o curso de pós-graduação.

Seção I : O papel dos modelos na ciência cognitiva

A ciência procura formular teorias que possam fornecer explicações, previsões e controlar o comportamento dos mais diversos fenômenos, sejam eles naturais ou sociais. Para cumprir este objetivo, ela se utiliza de vários recursos e, entre eles, da observação sistemática do comportamento de um organismo, experimentação e estratégia de elaboração de modelos.

Especialmente no campo da ciência cognitiva, o método dominante, e que se tornou um dos seus aspectos fundamentais, é o método de modelagem (*modelling*). Basicamente, modelar um fenômeno natural consiste em representar, como afirma Dupuy (1996, p.22), imitar, repetir ou reproduzir aspectos funcionais relevantes desse fenômeno, incorporando-os a algum artefato. A prática da modelização na ciência cognitiva consiste em realizar abstrações do fenômeno real, incorporando ao modelo computacional somente as relações funcionais, desprezando-se, assim, a natureza dos elementos físico-biológicos que constituem o processo cognitivo.

Dupuy (1996, p.27), ao analisar a utilização de modelos para gerar hipóteses plausíveis a respeito dos processos cognitivos, enfatiza que o papel desempenhado por eles é equivalente à função de conhecer :

Conhecer é produzir um modelo do fenômeno e efetuar sobre ele manipulações ordenadas. Todo conhecimento é reprodução, representação, repetição, ¹simulação. Isso caracteriza o modo científico,

¹ Agradecemos à Prof^a. Carmen B. Milidoni pela sugestão que nos fez refletir sobre o conceito de simulação durante o nosso exame de qualificação.



racional de conhecimento. As ciências cognitivas fazem desse modo, o modo único de todo conhecimento. Para elas, todo sistema cognitivo se relaciona com o mundo como o cientista com o seu objeto. Não é de espantar que a noção de representação ocupe um lugar central na ciência da cognição.

Segundo Doshier (1998,p.458), um modelo é uma representação de um fenômeno real, construída com finalidades específicas, para extrair um conjunto de princípios e conceitos que formarão uma teoria sobre um domínio de conhecimento. Tal teoria servirá para explicar e realizar previsões a respeito do comportamento do fenômeno em questão, diante de certas circunstâncias. De acordo com esse autor, os modelos são concebidos também para permitir uma “descrição ou uma analogia usada para ajudar a visualizar alguma coisa que não pode ser observada diretamente”. Doshier sustenta, ainda, que o modelo pode ser formalizado através de um conjunto de postulados, dados e inferências descritos matematicamente. A formalização do modelo permite ao cientista testar a precisão de uma hipótese.

Outro aspecto relativo à construção de modelos diz respeito a necessidade de estabelecermos o que Black (1961, p.218) chama de *convenções de interpretação*. Tais convenções são necessárias para se entender corretamente o que o modelo está querendo “dizer”. As convenções de interpretação permitem, na verdade, que se comuniquem os conhecimentos a respeito do fenômeno modelado.

Para Black (1961, p.218), existem diversos tipos de modelos e, entre eles, os modelos em escala, os analógicos, os teóricos e

os matemáticos. Os modelos em escala têm a finalidade de *imitar*, no sentido de construir uma identidade, uma semelhança visual com o original, de acordo com uma escala constante, mantendo uma relação biunívoca entre o modelo e o objeto representado. O modelo analógico tem a finalidade de reproduzir a *estrutura de relações*, mantendo uma correspondência biunívoca da estrutura de *relações* e não uma correspondência geométrica, como no caso dos modelos em escala. Black (1961, p.220), ressalta que estes modelos analógicos permitem inferir *hipóteses plausíveis* e não gerar demonstrações a respeito do comportamento do objeto modelado. Nos modelos matemáticos, estabelece-se um certo número de variáveis, elaboram-se hipóteses empíricas de acordo com as relações entre as variáveis escolhidas, fazem-se as simplificações necessárias e prossegue-se em busca das soluções para as equações matemáticas. Finalmente, extrapolam-se as conclusões para o comportamento da modelo original.

Aproveitando-se do momento de crescente desenvolvimento da matemática e da teoria da computação, a ciência cognitiva elegeu, como seu modelo predileto, o computador. Gardner (1995, p.20) aponta que o computador é utilizado pelos cientistas cognitivos com duas finalidades: a) realizar simulações e testes de modelos cognitivos e b) como instrumento de imitação do funcionamento da mente humana. Gardner sustenta ainda, que, apesar da ciência cognitiva ter surgido durante os anos 40 foi o sucesso da computação que deu a ela credibilidade. O computador passou a ser encarado como um verdadeiro

paradigma pelos (primeiros) cientistas cognitivos na busca de uma explicação científica para os fenômenos mentais. Assim, o estudo da mente humana passou a adotar a técnica de modelação computacional, dando início à *metáfora do computador inteligente*.

Esta analogia entre a máquina e a mente, se dá através da consideração de que o computador seria um artefato capaz de realizar computações apoiando-se nas capacidades de armazenar, processar e transmitir informações. Tais informações seriam representadas por símbolos, os quais, por sua vez, seriam manipulados por regras. Assim, o computador seria um sistema processador de informações alojado em um sistema material de silício. Paralelamente, em consequência da metáfora computacional, a mente humana seria estudada como um artefato de processamento de informações, funcionalmente semelhante ao processamento de informação de natureza computacional. Assim, as propriedades normalmente atribuídas à mente, tais como memória, aprendizado, inteligência, percepção, linguagem, capacidade intelectual dos seres humanos seriam explicadas segundo uma natureza computacional.

A importância dos computadores é tão forte no campo da ciência cognitiva que a definição do próprio campo ressalta sua importância. Neste sentido a definição de Laird (1993, p. 9) enfatiza :

"This book presents, I hope, a coherent view of the subject, but it is not a view that would be shared by everyone. Its key concept is computation – a notion that derives from the branch of mathematics known as

the theory of computability. Cognitive science, sometimes explicitly and sometimes implicitly, tries to elucidate the workings of the mind by treating them as computations, not necessarily of a sort that lies within this broader framework of the theory of computation."

Esta passagem, notadamente, ressalta a proposta cognitivista de fornecer explicações científicas para o fenômeno mental, através do uso do modelo computacional. Segundo esta visão, a mente poderia ser entendida como um fenômeno de natureza computacional.

Seguindo esta mesma linha de raciocínio, Pylyshyn (1986, p. 51) também enfatiza a forte presença das noções de computação no estudo dos processos cognitivos. Para ele:

"Nobody doubts that computers have had a profound influence on the study of human cognition. The very existence of a discipline called cognitive science is a tribute to this influence. One of the principal characteristics that distinguishes cognitive science from more traditional studies of cognition within psychology is the extent to which it has been influenced by both the ideas and the techniques of computing. It may come as a surprise to the outsider then to discover that there is no unanimity within cognitive science on either the nature (and in some cases the desirability) of the influence or what computing is, or at least on its essential character, as it pertains to cognitive science."

Ao seguir a direção computacional, o estudo das capacidades cognitivas humanas incorpora um vocabulário formado por expressões e conceitos, tais como : algoritmos, programas, linguagens de

programação, símbolos e regras além da dicotomia *hardware-software*. O hardware seria concebido como a instância material e o software seria análogo à instância abstrata da mente, enfatizando a relativa independência entre o processo de informação e sua implementação física.

A noção subjacente à analogia mente-computador é o conceito de *representação*, conhecido como um dos pressupostos da ciência cognitiva. Seguindo esta direção, o estudo das atividades cognitivas seria feito postulando a existência de representações mentais independentemente das influências biológicas e culturais. Os cientistas cognitivos partem do pressuposto de que o conhecimento é concebido como um conjunto de representações na mente humana. Essas representações seriam, segundo Gardner (1995), as regras, imagens, conceitos, idéias e símbolos, entre outros, organizados através de manipulações por algum procedimento mental análogo, em seu aspecto funcional, aos algoritmos computacionais. A analogia mente-computador é resumida por Thagard (1996, p.11) da seguinte maneira:

Programa	=	estruturas de dados	+	algoritmos
Mente	=	representação mental	+	procedimentos computacionais

No esquema acima, a analogia se dá entre estruturas de dados com representações mentais e algoritmos com procedimentos computacionais para se investigar e tirar conclusões ou hipóteses sobre os processos mentais. Assim, a mente seria explicada através de uma

abordagem computacional na qual programas de computador seriam construídos para simular os processos cognitivos. O estudo das representações e processos mentais se apoia, por sua vez, nas definições precisas e em franco desenvolvimento da ciência da computação.

Ao adotar a perspectiva computacional para o estudo da mente, a ciência cognitiva enfrenta vários problemas.

O primeiro consiste em admitir que vários aspectos cognitivos teriam uma natureza serial, seqüencial, à semelhança do que ocorre com o computador; enquanto muitos defendem a natureza paralela e distribuída de alguns processos cognitivos. Uma segunda dificuldade enfatiza a incompletude do modelo cognitivista da mente, que trata a cognição de maneira muito simplificada. Os críticos argumentam que a cognição não pode ser reduzida a operações de tarefas simples como aquelas executadas em um computador. Outra dificuldade diz respeito à natureza material com que é constituído o computador e a natureza biológica (substrato material) da mente. Os críticos como Searle (1997) e Dupuy (1996) argumentam que a mente é um fenômeno essencialmente dependente do organismo e não de qualquer outra matéria inanimada.

Na visão de Gardner (1995, p.405) uma ciência cognitiva, fundada no pressuposto computacional, parece dar conta de processos cognitivos mais simples, racionais, apoiados na lógica matemática. Porém, seria inadequado para explicar processos cognitivos mais complexos carregados de crenças que não se enquadrariam em uma simples manipulação ordenada de símbolos, da maneira que ocorre com o

computador digital serial. Esse autor enfatiza que o ser humano lança mão de outros recursos tais como heurísticas, estratégias, vieses, imagens e outras abordagens vagas e aproximadas. Gardner sugere, ainda, que a ciência cognitiva, ao postular uma explicação racionalista, faz emergir um aspecto contraditório denominado de *paradoxo computacional*, que pode ser assim sintetizado : quanto mais a ciência procura uma explicação para os processos cognitivos apoiada no modelo computacional, mais fica demonstrado o quanto os processos cognitivos são diferentes dos processos computacionais. Em suma, o esforço científico vem mostrando o quanto os seres humanos diferem de um computador. Neste sentido, a ciência cognitiva ainda não produziu a explicação mais geral e apropriada do fenômeno cognitivo.

Gardner (1995, p.408) ressalta, ainda, a dificuldade decorrente da diferença entre sistemas biológicos e sistemas mecânicos como o computador. Para ele, os processos de pensamento são influenciados pelas modificações ou desenvolvimentos biológicos como aqueles que ocorrem ao longo do tempo com os seres humanos que, dificilmente, seriam incorporados ou exibidos por um modelo a imagem do computador. Assim, uma explicação para a cognição teria que incorporar, na opinião de desse autor, tanto o aspecto biológico, fornecido pela neurobiologia, como o aspecto computacional fornecido pela Inteligência Artificial.

Outras dificuldades ressaltadas por alguns pesquisadores, como por exemplo Eckardt (1995, p.98) são conseqüências do fato da

ciência cognitiva analisar as capacidades cognitivas apoiada no uso da metáfora do computador. A autora argumenta que não se estuda a cognição em seus próprios termos e sim com a visão da computação.

Em resumo, a ciência cognitiva explorou, em seus primeiros anos, a capacidade de manipulação simbólica do computador digital beneficiando-se do amplo desenvolvimento da matemática e da ciência da computação para simular processos cognitivos.

Essa metodologia de estudo, no campo da ciência cognitiva, passou a ser formalmente conhecida como modelo de *inteligência artificial*.

Mais recentemente, por volta dos anos 80, uma nova abordagem para o estudo das propriedades cognitivas foi incorporada pela ciência cognitiva, a que também contempla algumas propriedades computacionais em seu funcionamento, ficando conhecida como *Conexionismo*, *PDP* ou *Redes Neurais Artificiais*. A abordagem conexionista, sugere, ao contrário da abordagem simbólica, que os processos cognitivos resultam da interação de elementos interconectados funcionando paralelamente.

O modelo de inteligência artificial foi abordado no capítulo II e o modelo de redes neurais artificiais será estudado com mais detalhes na próxima seção.

Seção II – Redes neurais artificiais e os modelos cognitivos

Nesta seção retomamos a análise do modelo de Redes Neurais Artificiais (RNA), iniciada no capítulo II. Nosso estudo ressalta, particularmente, os princípios matemáticos gerais e os principais mecanismos de aprendizado utilizados na construção da maioria dos modelos de RNA, e será conduzido apoiando-se nas obras de Rumelhart & McClelland (1995, p. 46-56) e Caudill & Butler (1992, p. 3-8).

Rumelhart e McClelland destacam oito aspectos básicos a serem especificados para a construção dos modelos de RNA : a) um conjunto de *neuron-like units* ou *unidades de processamento*; b) o *estado de ativação* das unidades a cada instante de tempo; c) uma *função de saída* para cada unidade; d) um *padrão de conectividades* entre as diferentes unidades; e) uma *regra de propagação* dos *outputs* de cada unidade para as unidades vizinhas; f) uma *regra de ativação* que produz um novo estado de ativação das unidades da RNA; g) uma *regra de aprendizagem* que modifica os padrões de conectividade entre as unidades; h) parâmetros *ambientais* que viabilizem o processo de modelamento.

O primeiro passo na construção de uma RNA consiste em especificar o conjunto de *unidades de processamento* e os aspectos, propriedades, conceitos ou objetos que elas representam. Tais elementos, do mundo externo, podem ser representados de maneira *local*, em unidades específicas, ou *distribuída* pelo conjunto de unidades da RNA de

tal maneira que um padrão global resulta da interação entre as unidades da rede e que não seria encontrado entre nas unidades em separado. Basicamente, as unidades computam os *inputs* recebidos de outras unidades por intermédio das interconexões entre elas, gerando um valor de *output* que se distribui com a mesma intensidade para outras unidades. Conforme sugerem McClelland & Rumelhart, as unidades realizam suas funções ao mesmo tempo, de forma independente e *paralela*. Uma rede neural artificial com seus elementos básicos é mostrada na Figura 7 abaixo:

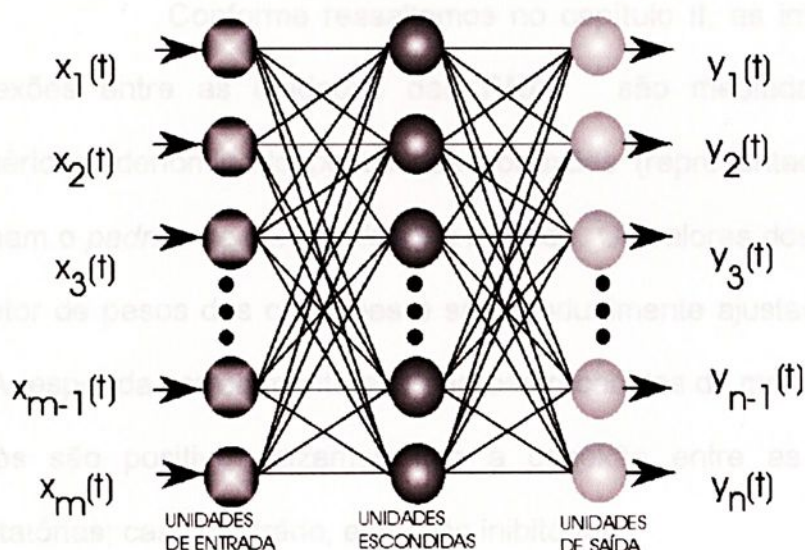


FIGURA 7 - Uma rede reural artificial típica

Durante o funcionamento da rede, as unidades possuem *estados de ativação* individuais $a_i(t)$ que, em conjunto, representam o padrão de ativação global da RNA. O padrão de ativação global, formado pelo conjunto de ativações individuais, é representado, matematicamente, por um vetor composto dos valores de ativação (contínuos ou discretos) de cada unidade.

Cada unidade da rede neural artificial é caracterizada também por uma *função de saída*. A função de saída transforma o estado de ativação atual da unidade em um valor de saída que será transmitido para as unidades vizinhas conforme elas interagem. Normalmente, esta função costuma ser especificada de acordo com limiar o de ativação ou *threshold* da unidade. Representando o sinal de saída da unidade por $y_i(t)$ temos : $y_i(t) = a_i(t)$ se $a_i(t)$ for maior que o *threshold* T da unidade, onde $a_i(t)$ é o estado de ativação da unidade no tempo t .

Conforme ressaltamos no capítulo II, as intensidades das conexões entre as unidades das RNA são mediadas por valores numéricos, denominado pesos das conexões (representado por w_{ij}) que formam o *padrão de conectividades* da rede. Os valores dos pesos formam o vetor de pesos das conexões e são gradualmente ajustados para que a RNA responda corretamente aos padrões recebidos do meio exterior. Se os pesos são positivos dizemos que a conexão entre as unidades são excitatórias; caso contrário, elas são inibitórias.

Os valores de saída ou *output* das unidades das RNA se propagam através da rede quando elas interagem. Assim, as RNA possuem uma regra de propagação que concilia o conjunto de *outputs* das unidades com o conjunto de pesos das conexões, resultando no *net input* ou efeito total recebido por cada unidade a partir das unidades vizinhas. O valor do *net input* I para uma unidade i é dado por :

$$I_i = \sum_{j=1}^n w_{ij} * x_j$$

onde x_j representa a entrada ou *input* da unidade i a partir da unidade j e w_{ij} o peso da conexões entre as unidades i e j . A Figura 8 representa o modelo matemático das unidades que formam uma RNA :

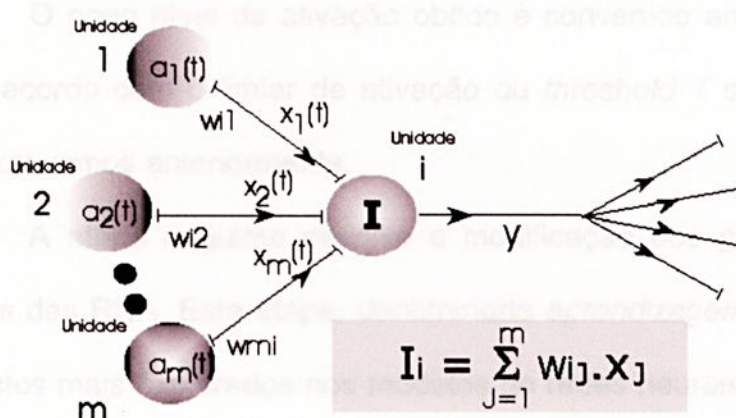


FIGURA 8 - Modelo matemático de uma unidade das RNA

(Adaptado de Churchland (1993, p.160))

Outro passo realizado, durante a construção das RNA, é a definição de *regra de ativação* para as unidades da rede. Através da regra de ativação, cada unidade alcança um novo estado de ativação apoiando-se no seu estado atual $a_i(t)$ e no *net input* I_i recebido das unidades próximas. O novo estado de ativação é obtido transformando o *net input* I_i da unidade no instante t por meio de uma função f que fornece resultados dentro de uma faixa de valores de ativação máximos e mínimos, como por exemplo a função sigmóide :

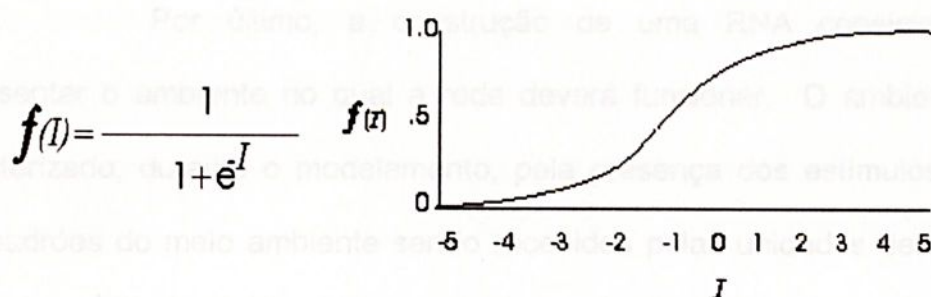


FIGURA 9 - Função sigmóide (Churchland, 1993, p. 161)

O novo nível de ativação obtido é convertido em um sinal de saída, de acordo com o limiar de ativação ou *threshold* T da unidade conforme descrevemos anteriormente.

A etapa seguinte envolve a modificação dos padrões de conectividades das RNA. Esta etapa, denominada *aprendizagem*, constitui um dos aspectos mais explorados nos modelos de redes neurais artificiais. De acordo com Rumelhart & McClelland (1995), a modificação dos padrões de conectividade corresponde ao ajuste dos pesos das conexões entre as unidades. O ajuste dos pesos pode dar origem a novas conexões, eliminar ou modificar conexões existentes. Existem vários métodos de modificação dos pesos das conexões entre as unidades, porém o método de Hebb, descrito no capítulo anterior, foi o primeiro a ser utilizado na construção de uma RNA. Formalmente, a regra de Hebb pode ser dada por :

$$\Delta w_{ij} = \eta a_i y_j$$

A expressão acima sugere que a modificação do peso ou energia da conexão entre a unidade i e a unidade j é proporcional à taxa de aprendizado η , ao valor de saída da unidade j e da ativação da unidade i .

Por último, a construção de uma RNA consiste em representar o ambiente no qual a rede deverá funcionar. O ambiente é caracterizado, durante o modelamento, pela presença dos estímulos ou dos padrões do meio ambiente sendo recebidos pelas unidades de *input*. Os padrões de *input* mantêm uma independência entre si, podendo ser identificados numericamente formando os elementos de um vetor.

Retomando a questão da aprendizagem nos modelos de RNA, analisamos, a seguir, os aspectos relativos a esta propriedade, com maior profundidade. De acordo com Caudill & Butler (1993, p.119), o *aprendizado* em uma rede neural artificial é um conjunto de procedimentos que informam às unidades da rede como modificar os pesos de suas conexões em resposta aos estímulos recebidos do meio ambiente. A estabilização da resposta da rede, em um valor correto, depende apenas do ajuste dos pesos das conexões entre as unidades, pois os padrões de entrada são fornecidos pelo meio externo e as unidades não possuem autonomia para modificar sua própria função de ativação. Assim, o conhecimento ou o padrão informacional do meio ambiente é armazenado nas conexões entre as unidades. Durante o processo de aprendizado, as modificações nos pesos das conexões provocam a propagação de uma cadeia de excitações e inibições entre as unidades que compõem a rede.

Para que uma RNA aprenda a reconhecer os padrões recebidos é necessário que ela passe por um processo de *treinamento*. Durante o treinamento, a RNA recebe, repetidamente, os padrões de *input* do meio ambiente procurando se estabilizar em um padrão de

conectividades para reproduzir o padrão de *input* recebido. Este procedimento pode ser realizado de duas maneiras: através do treinamento *supervisionado* ou do *não supervisionado*, também conhecido por *auto-organização*.

No treinamento supervisionado, a rede *aprende* a *associar* o padrão recebido à classe a qual pertence. Neste caso, para cada padrão de *input* apresentado à rede informa-se, também, o padrão de *output* correspondente. A classe ou padrão de *output* associado ao padrão de *input* é estabelecido por um *professor* ou *supervisor* durante o processo de treinamento. O treinamento supervisionado pode ainda ser implementado através de um procedimento de correção de erros entre o padrão de *output* desejado e o padrão de *output* real obtido. Neste caso, um supervisor apresenta o padrão de *output* desejado para ser comparado com o padrão real fornecido pela rede e, em seguida, os erros são computados para que a diferença entre os dois padrões possa ser reduzida, reajustando os pesos das conexões.

No treinamento não supervisionado ou por *auto-organização*, o padrão de *output* desejado não é fornecido à rede por um *supervisor externo*. A rede procura detectar o padrão de *output* apoiando-se somente nas características ou regularidades do conjunto de padrões de *input* do meio ambiente. De acordo com Caudill & Butler (1994, p.121), a falta de supervisão é utilizada nas situações em que não se conhece, de antemão, a resposta correta a ser fornecida pela rede e, quando a própria RNA deve decidir quais regularidades presentes nos padrões de entrada

serão utilizadas para classificar os padrões recebidos. Matematicamente, a ausência de supervisão é realizada através da adoção de uma medida de similaridade entre dois vetores: o vetor de entradas e o vetor de pesos das conexões entre as unidades.

Conforme ressaltamos, as RNA têm despertado um grande interesse nos cientistas cognitivos. Aleksander (1989, p.329) aponta alguns motivos para tal interesse.

Primeiro, as RNA oferecem uma abordagem alternativa ao modelo de Inteligência Artificial para simular os processos cognitivos. Se no modelo de Inteligência Artificial, a idéia central reside na construção de programas apoiados na manipulação de símbolos por meio de regras abstratas, nas RNA o foco de análise, dos processos cognitivos, está centrado na sua base física.

Em segundo lugar, as hipóteses biológicas adotadas para a construção das RNA e entre elas, o processamento de informação paralela e distribuída, o grande número de elementos interconectados que se excitam ou se inibem mutuamente e a propriedade de aprendizado através da modificação da intensidade das conexões sinápticas.

Em terceiro lugar, os cientistas cognitivos estão interessados, além das propriedades de aprendizagem e reconhecimentos de padrões das redes neurais artificiais, na propriedade de memória demonstrada por estas redes. De acordo com Caudill & Butler (1994, p.37), o armazenamento da informação nos modelos de RNA ocorre por associação entre informações. Em contraste com o modelo de IA, no

qual a informação é localizada em *endereços fixos*, o armazenamento da informação na memória da rede artificial é compartilhado por um grande número de unidades. Tzanakou (1995, p.220) ressalta que a memória nas RNA é um processo *dinâmico* no qual a informação armazenada é reorganizada no decorrer do tempo. Ainda, segundo este autor, a informação não consolidada na memória é descartada e um reajuste das conexões entre as unidades geram novas representações.

Por último, o interesse pelos modelos de RNA também foi despertado pela possibilidade de emular processos de auto-organização. Neste caso, os teóricos da auto-organização analisam a emergência de padrões globais a partir da interação entre as partes componentes de modelos específicos de RNA. O processo de emergência deste padrões nas redes neurais artificiais será estudado através da teoria da auto-organização apresentada nos capítulos IV e V.

Em resumo, procuramos analisar, nesta seção, o modelo matemático e os principais tipos de treinamento compartilhado pela maioria das redes neurais artificiais. O treinamento não supervisionado é de particular interesse para este trabalho, pois é o recurso utilizado pelo modelo de rede neural artificial de Kohonen que será objetos de estudo no capítulo V.

Na próxima seção examinamos algumas críticas aos modelos de inteligência artificial e de redes neurais artificiais.



Seção III – Limites e alcances dos modelos cognitivos

Os modelos de Inteligência Artificial (IA) e de Redes Neurais Artificiais (RNA) se tornaram os instrumentos paradigmáticos da ciência cognitiva no estudo da mente. Conforme analisamos na seção anterior, o modelo de Inteligência Artificial é o resultado do cruzamento dos conceitos de máquina de Turing e do funcionalismo. O modelo de IA, dominante nos primeiros anos da ciência cognitiva, parte do pressuposto que o funcionamento da mente é análogo ao de um computador digital. Sustentar a idéia que processos mentais podem ser descritos de forma algorítmica e reproduzidos, mecanicamente, em uma instância da máquina de Turing enfrenta várias dificuldades quanto aos seus limites e alcances na explicação do comportamento inteligente.

Dreyfus (1975, p.41) sugere que a IA ao estudar o comportamento inteligente, como aquele capaz de processar informação no estilo de um computador digital, serve apenas para, paradoxalmente, evidenciar que as pessoas quando se comportam de maneira inteligente não processam informações como um computador digital. Esse autor aponta vários fracassos do modelo de inteligência artificial ao tentar realizar simulações cognitivas do tipo: tradução automática de idiomas, solução de problemas e reconhecimento de padrões. As críticas de Dreyfus se apoiam fundamentalmente na diferença entre o processamento de informação no computador e nos seres humanos. Para ele, várias tarefas que são solucionadas pelos seres humanos não são resolvidas através da

manipulação de símbolos por meio de regras, como por exemplo a capacidade de lidar com os aspectos contextuais envolvidos na tradução de idiomas. Além disso, Dreyfus (1975, p.270) ressalta que nem todo comportamento inteligente humano pode ser programado para seguir regras rígidas, ao contrário; em muitas situações os seres humanos "...parecem estar utilizando uma organização perceptual global, fazendo distinções pragmáticas entre operações essenciais...", que não se reduzem a simples manipulação de regras.

Searle (1997), crítico severo da corrente da IA forte, não concorda com o princípio de que os estados mentais possam ser caracterizados independentemente do seu substrato físico. Para esse autor, os aspectos biológicos influenciam os estados mentais contestando a idéia que sugere uma equivalência entre a mente e um programa de computador. Searle sugere que manipular símbolos por meio de regras ressalta apenas o aspecto formal ou sintático dos programas enquanto que a natureza da atividade mental envolve aspectos semânticos. Assim, a tentativa de explicar a natureza da mente apoiando-se na analogia mente-computador não caracteriza a verdadeira natureza do fenômeno mental.

Outra limitação do modelo de Inteligência Artificial diz respeito à sua capacidade de exibir propriedades de auto-organização. Como esses modelos são estritamente governados por regras preestabelecidas, a possibilidade de ocorrência de aspectos de auto-organização é extremamente reduzida.

Nos modelos de RNA são apontadas algumas críticas e limitações. Dawson & Shamanki (1997, p.1) sugerem que o modelo conexionista não reproduz apropriadamente os processos neurofisiológicos, por isso não fornece uma explicação adequada para o fenômeno cognitivo. Esses autores ressaltam que o estudo dos fenômenos cognitivos, no âmbito das RNA, são mais guiados para a resolução de problemas de engenharia do que para aspectos cognitivos ou neurofisiológicos.

Fodor & Pylyshyn (1989) elaboraram uma série de críticas aos modelos de RNA confrontando-os com o modelo de IA no tocante à natureza explicativa dos processos cognitivos. Apesar desses modelos se apoiarem na idéia de representação mental, a natureza de tais representações são distintas. No modelo de IA, ressaltam Fodor & Pylyshyn (1989, p.12), a representação através de símbolos possui uma estrutura sintática e semântica combinatorial. Tal estrutura permite compor representações mais sofisticadas (moleculares), a partir de representações básicas ou atômicas. As representações são sintaticamente estruturadas de tal maneira que o conteúdo semântico das representações compostas é o resultado da estruturação sintática dos elementos atômicos. Para esses autores, os modelos de RNA não representam de forma consistente os processos mentais pois eles não reconhecem nenhuma dependência da semântica-sintaxe combinatorial exigida pelas representações mentais. Nos modelos de RNA suas unidades básicas se relacionam somente através de relações causais entre pares de unidades não formando uma cadeia de relações estruturadas. Por último, Fodor & Pylyshyn (1989, p.3) consideram

apenas a possibilidade do modelo conexionista servir como uma abstração do aparato neural no qual as representações simbólicas seriam instanciadas.

Em resumo, ressaltamos nesta seção que os modelos da IA e RNA possuem limitações quanto ao alcance de suas explicações a respeito da natureza dos processos cognitivos.

No próximo capítulo, que constitui a segunda parte do nosso trabalho, discutiremos o tema da auto-organização destacando os elementos conceituais necessários para analisar as propriedades de auto-organização em um modelo de RNA específico denominado *rede de Kohonen*.



Apresentação

Compreender a ampla variedade de elementos que se apresenta na natureza e se espelhar nela para compreender fenômenos biológicos, físicos e socio-culturais parece ser o objetivo da ciência. Em particular, no caso humano, desvendar comportamentos e conhecer a sua razão de ser, constituem tarefa que exige enorme esforço do cientista. Entender um fenômeno ou o aparato cognitivo humano é, com certeza, um grande desafio. Para ajudar neste esforço surgiu durante os anos 40, no seio do movimento cibernético, uma proposta de estudar a mente humana apoiando-se nas teorias da auto-organização.

Esta é também, em parte, a proposta deste capítulo, ou seja, a de analisar o tema da auto-organização aplicado ao âmbito dos modelos cognitivos das redes neurais artificiais, para no capítulo V reconhecer os limites e as possibilidades da auto-organização nestes modelos. Os aspectos de auto-organização serão delineados a partir das idéias filosóficas de Debrun (1996), Von Foerster (1960) e Ashby (1962) e Gonzales(1996).

O capítulo está dividido em quatro seções: na primeira, apresentamos um breve relato histórico do surgimento das teorias de auto-organização; na segunda, analisamos os aspectos conceituais do tema da auto-organização; na terceira, caracterizamos duas modalidades de auto-organização, ou seja, a auto-organização primária e a auto-organização

secundária. Finalmente, na quarta seção, examinamos a relação entre a ciência cognitiva e os estudos de auto-organização.

CAPÍTULO IV - AUTO-ORGANIZAÇÃO



Seção I – Resumo Histórico das Ideias de Auto-organização

O estudo sobre os aspectos de auto-organização começou durante a década de 40 e 50, com a criação de modelos matemáticos e físicos para descrever a organização de sistemas vivos. Os primeiros modelos foram baseados em conceitos de física, como a teoria da difusão e a teoria da reação-difusão. Esses modelos foram utilizados para descrever a organização de sistemas vivos, como a formação de padrões de crescimento e a organização de redes de comunicação.

CAPÍTULO IV – AUTO-ORGANIZAÇÃO

Assim, vários trabalhos científicos foram publicados, mostrando a importância da auto-organização na formação de sistemas vivos. A auto-organização é um processo pelo qual um sistema se organiza espontaneamente, sem a intervenção de um agente externo. Esse processo é observado em sistemas físicos, químicos, biológicos e sociais. A auto-organização é um fenômeno muito comum na natureza, e é um dos principais mecanismos pelos quais a vida se organiza.

De acordo com Yule (1937), o termo "self-organization" foi utilizado pela primeira vez em 1951 por Farley e Clark no Laboratório Lincoln. Porém, a ideia de auto-organização esteve presente em pesquisas anteriores, como as de Prigogine (1966). Prigogine, ao estudar a termodinâmica dos sistemas abertos, descobriu que a auto-organização é um fenômeno muito comum em sistemas abertos. Ele mostrou que a auto-organização é um processo pelo qual um sistema se organiza espontaneamente, sem a intervenção de um agente externo. Esse processo é observado em sistemas físicos, químicos, biológicos e sociais.

Seção I – Resumo histórico das idéias de auto-organização

O estudo científico do conceito de auto-organização começou durante a década de 40 e foi impulsionado nos anos 60 e 70 pelos pesquisadores da corrente cibernética conhecida como *cibernética de segunda ordem*. Enquanto o primeiro movimento cibernético se preocupava com o controle do comportamento dos sistemas focalizando o conceito de *feedback*, os principais ciberneticistas da segunda fase, tais como Foerster, Ashby e Shannon, focalizavam sua preocupação nos conceitos de organização, desordem, complexidade, propriedades emergentes e, principalmente, na noção de *autonomia*, tanto em sistemas naturais quanto artificiais.

Assim, vários trabalhos científicos foram publicados e diversos programas de pesquisas formalizados envolvendo, principalmente, a física, a biologia, a química e as ciências humanas, com a preocupação de estabelecer os princípios básicos que regem o comportamento dos sistemas que se auto-organizam.

De acordo com Yovits (1962), o termo *self-organizing systems* foi utilizado em 1954 por Farley e Clark no laboratório Lincoln. Porém, a idéia de auto-organização estava presente em pesquisas anteriores a esta data. Assim, a partir dos anos 40, Prigogine (1996), atuando no campo da física e da química, estudou os processos termodinâmicos irreversíveis (processos dissipativos), demonstrando a

preocupação com o conceito de auto-organização. Em seu livro, *O Fim das Certezas*, Prigogine (1996, p.59) afirma:

Faz agora cinquenta anos que publiquei meu primeiro artigo sobre termodinâmica do não-equilíbrio. Neste trabalho, eu já ressaltava o papel construtivo da irreversibilidade. Que eu saiba, era a primeira publicação que propunha o problema da auto-organização associado ao afastamento do equilíbrio.

O autor enfatiza que, nos processos dissipativos, as perturbações ambientais – denominadas de *fluctuations* - podem provocar a emergência de fenômenos ordenados que resultariam na formação de novas estruturas denominadas *estruturas dissipativas*. A este princípio foi dado o nome de *order through fluctuation*.

Segundo Dupuy (1996) Von Foerster fundou o Biological Computer Laboratory na Universidade de Illinois, em 1956, enfatizando os estudos sobre auto-organização. Entre seus trabalhos mais importantes está o artigo "*On Self-Organizing Systems and their Environments*", publicado em 1960, o qual se tornou fundamental no campo da auto-organização. Este trabalho descreve a importância do ambiente como elemento necessário e fundamental para que um sistema aberto se auto-organize, descartando, assim, a possibilidade de ocorrência do processo de auto-organização em um sistema fechado. Além disso, Von Foerster (1960) enfatiza que o surgimento de uma organização é possível, não só através do *princípio da ordem a partir da ordem*, sugerido no trabalho *What is Life*,

de Scrodinger, mas também que o ruído pode desempenhar um papel organizador. Este princípio de auto-organização, o autor denominou de *princípio da ordem pelo ruído*.

Atlan (1992), seguindo a linha de pensamento de Von Foerster (1960), reforça os princípios de auto-organização a partir do ruído, aplicando-os ao campo da biologia. Atlan (1992) procura compreender os princípios que levam a um aumento de complexidade das organizações sob efeito do acaso. O estudo de Atlan segue uma linha mais comprometida com a questão da complexidade emergente dos processos de auto-organização. O pensamento de Atlan pode ser sintetizado através do *princípio da complexidade pelo ruído*.

Von Neumann também se preocupou em investigar os aspectos relacionados a autonomia dos sistemas naturais e artificiais, em seu trabalho *"Theory of Self-Reproducing Automata"* de 1966. Ele descreve a noção de autômato celular como um conjunto de elementos computacionais (*"cells"*) que se comportam como autômatos de estado finito. A idéia principal deste trabalho era investigar a possibilidade de auto-replicação em sistemas com comportamento semelhante ao de uma máquina. Apesar da presença de regras e da característica determinista do modelo, o autômato celular pode ser visto, de maneira rudimentar, como um processo de auto-organização bastante simples capaz de gerar estruturas definitivas por meio de um processo evolutivo.

Ashby (1962), também um dos fundadores da segunda fase do movimento cibernético, teve uma imensa influência nas pesquisas sobre

auto-organização. Especialmente no seu "*Principles of the Self-Organizing Systems*" de 1962, ele estabelece os princípios básicos, os limites e as contradições do termo auto-organização quando aplicado aos sistemas naturais e artificiais. Iniciando-se, a partir de um princípio de condicionalidade, entendido como um princípio de dependência, entre os elementos de um sistema, o processo de auto-organização somente é possível com a participação do meio ambiente, através de uma troca de informação entre o sistema e sua vizinhança. Para Asbhy (1962) somente é possível falarmos propriamente em auto-organização considerando o sistema mais o meio ambiente como um todo (acoplados). Neste sentido, o termo auto-organização sugere uma contradição em relação ao prefixo auto. Esse autor conclui ainda que é impossível a existência de processo auto-organizador nos sistemas artificiais, e que a utilização do termo tende a desaparecer no âmbito destes sistemas.

Na mesma linha de atuação, Maturana e Varela (1980), na área de biologia, conceituaram a chamada teoria da *autopoiese*, palavra grega que significa auto-produção. A teoria da autopoiese enfatiza que os organismos vivos são sistemas autopoieticos, ou seja, capazes de autoproduzirem seus componentes e as relações que formam sua estrutura. Os organismos com esta propriedade são capazes de se auto-referenciarem, desenvolvendo um processo de *auto-renovação* ("*self-renewal*") para manter sua organização. Segundo Stichweh (1990, p.195) a teoria da autopoiese difere da teoria da auto-organização pela ênfase relativa que cada corrente atribui ao grau de autonomia dos sistemas. Para

ele, um sistema deve ser pensado primeiro como *auto-organizante* para depois passar à fase da autopoiese.

Após este resumo histórico a respeito do surgimento das teorias sobre auto-organização, veremos, na seção seguinte, os aspectos conceituais mais relevantes para a compreensão das propriedades de auto-organização presentes nas redes neurais artificiais.



Seção II - Auto-Organização: aspectos conceituais

A idéia central deste trabalho é analisar os aspectos de auto-organização que, supostamente, estariam presentes em alguns modelos de redes neurais artificiais, em particular no modelo conhecido como *rede neural de Kohonen*. Assim, guiados por esse objetivo, abordaremos, nesta seção, aspectos conceituais da auto-organização apoiando-nos em alguns trabalhos que tiveram grande impacto nos estudos deste tema. Tomaremos como referência, os trabalhos de Von Foerster (1960), Ashby (1962), Debrun (1996) e Gonzales (1996), embora existam outras abordagens extremamente amplas e ricas no contexto da auto-organização. Apesar da diversidade de caracterizações deste termo, tentaremos identificar os elementos básicos do conceito de auto-organização, relevantes para a posterior análise do modelo conexionista de Kohonen. Neste contexto, a questão central deste trabalho pode ser assim representada: *Pode um sistema artificial, particularmente uma máquina conexionista, se auto-organizar?* Nossa pretensão, ao responder a esta questão, é extremamente moderada diante da dificuldade em fornecer uma resposta precisa. Assim, nos limitaremos a apontar as principais alegações de ordem técnica, os limites e alcances dos modelos conexionistas diante da possibilidade de auto-organização.

Embora ainda haja dúvidas sobre a existência (ou não) do fenômeno da auto-organização em sistemas artificiais, diversos fatores são apontados como motivadores para o estudo deste fenômeno. Na área dos

sistemas artificiais, podemos citar as limitações demonstradas pelos computadores ao tratarem informações não numéricas, como por exemplo, o reconhecimento de padrões. Na área dos sistemas biológicos, os pesquisadores, como Varela (1997), partem da tese de que os organismos são sistemas auto-organizadores. Esses organismos constituem-se em uma rica fonte para os estudos sobre auto-organização, principalmente por apresentarem características tais como as capacidades de adaptação, aprendizagem, auto-regulação, etc., e a habilidade em delinear seus próprios objetivos e propósitos. Debrun (1996) julga que uma forte motivação para falarmos de auto-organização ocorre quando o surgimento de certas "formas", que aparentemente não obedecem a "leis de construção", não podem ser explicadas pela visão clássica (Newtoniana) ou determinista da ciência, mas que poderiam ser explicadas através da alternativa da auto-organização. De acordo com Debrun essas "formas" surgem através de um processo de *ajuste* entre elementos, ajuste esse que culmina com o estabelecimento de relações de dependência entre elementos sem a presença de um condutor dos destinos do processo.

Conforme mencionamos a idéia de auto-organização tem sido estudada sob várias perspectivas. Uma revisão na literatura mostra que o assunto é tratado ora sob a perspectiva epistemológica, ora ontológica. Percebe-se, também, que o tema da auto-organização possui um caráter multidisciplinar fazendo parte, portanto, da agenda de questões de várias áreas do conhecimento. Neste sentido, Yovits & Cameron (1960, p.v) sugerem algumas áreas que envolvem temas da auto-organização :



"On the one hand the psychologist, the embryologist, the neurophysiologist and others involved in the life sciences were attempting to understand the self-organizing properties of biological systems, while mathematicians, engineers, and physical scientists were attempting to design artificial systems which could exhibit self-organizing properties."

A esta citação podemos acrescentar que os estudos de auto-organização abrangem, também, os sistemas sociais, estendendo-se a grupos de indivíduos que, através das interações sociais, produzem permanentes ações e comportamentos coletivos, sem a ação de líderes permanentes.

Vários obstáculos se apresentam diante da tentativa de compreender o conceito de auto-organização. Primeiro, a ausência de definições consensuais das noções de informação, ordem e desordem, ruído, organização, evolução e da noção fundamental do prefixo auto. Segundo, o caráter multidisciplinar do campo da auto-organização, ao mesmo tempo em que contribui para desvendar o assunto, dificulta a formação de um vocabulário único capaz de fornecer conceitos que permitam explicar o fenômeno da auto-organização. Por último, destacamos as múltiplas denominações encontradas na literatura para conceitos sutilmente diferentes do conceito de auto-organização, como por exemplo, os conceitos de *self-referential*, *self-reproducing* e *autopoiese*. A busca de definições unificadas, completas e profundas com um caráter

universal, que resulte em uma teoria da auto-organização exige um enorme esforço, supostamente impossível de ser realizado em estudos isolados.

A presença do tema da auto-organização, tendo como núcleo comum o aspecto da autonomia do comportamento de certos fenômenos na literatura, caminha em duas direções: uma diz respeito ao estudo do fenômeno da auto-organização propriamente dito, buscando-se uma teoria explicativa para uma classe de fenômenos naturais e artificiais; outra segue a linha de uma teoria da autopoiese mais ligada à biologia e à organização das células. Nosso interesse será pelos fenômenos do primeiro grupo.

De acordo com Debrun (Debrun et al., 1996), os trabalhos sobre auto-organização seguem quatro linhas básicas. A primeira é a linha seguida por Prigogine, que enfatiza os estudos dos sistemas termodinâmicos longe do equilíbrio, com flutuações que se manifestam e são capazes de desenvolver alguma forma de organização. Esses trabalhos seguem o princípio da ordem a partir de flutuações.

A segunda linha explica a noção de auto-organização ao enfatizar a participação de elementos distintos que interagem desenvolvendo uma forma ou organização sem a presença de um elemento responsável para supervisionar o processo. Podemos incluir nesta linha os estudos teóricos do próprio Debrun (Debrun et al., 1996) e de Ashby (1962) centralizados nos conceitos de organização, condicionalidade e dependência entre elementos.

A terceira abordagem preocupa-se com a natureza do processo auto-organizador, destacando os seguintes aspectos : o tempo como elemento relevante do processo de auto-organização, a dinâmica do processo organizador visto como uma interação entre elementos distintos, e o surgimento de um atrator. Esta linha, seguida por Debrun, procura reforçar o caráter "auto" do processo de auto-organização, enfatizando sua autonomia em relação ao meio ambiente, o papel do sujeito, do acaso e das condições de partida.

A quarta abordagem dos processos de auto-organização apontadas por Debrun (Debrun et al.,1996) segue as idéias de Von Foerster (1960) e Atlan (1986) que, ao contrário do que tradicionalmente se imagina, atribuem ao ruído um papel organizador. Este princípio pode ser resumido por *princípio de ordem através do ruído*.

Na opinião de Debrun, uma boa definição de auto-organização inclui os seguintes critérios : 1) deve permitir identificar os processos verdadeiramente auto-organizadores; 2) deve indicar os mecanismos desses processos; 3) não deve mostrar uma lei de construção do processo de auto-organização e 4) não deve se distanciar das definições mais técnicas encontradas nos trabalhos dos anos 50.

Passaremos, a partir deste ponto, a enfocar os aspectos relevantes desenvolvidos por Debrun a respeito da sua visão de auto-organização. No trabalho deste autor o estudo conceitual da auto-organização inicia-se com uma caracterização da natureza mais geral da auto-organização e prossegue detalhando o papel dos diversos



componentes encontrados em sua caracterização. Formalmente, Debrun (Debrun et al. 1996, p.4) faz a seguinte caracterização do processo de auto-organização :

Há auto-organização cada vez que o advento ou a reestruturação de uma forma, ao longo de um processo, se deve principalmente ao próprio processo – às características nele intrínsecas -, e só em grau menor às suas condições de partida, ao intercâmbio com o ambiente ou à presença eventual de uma instância supervisora.

Na caracterização acima ficam ressaltados vários aspectos que, ao serem analisados, correm o risco de, eventualmente, enquadrarem-se na categoria de hetero-organização, já que o processo de auto-organização é uma questão de mais ou menos e não de tudo ou nada (Debrun et al., 1996, p.5), de maneira a preservar o aspecto “auto” do fenômeno da auto-organização. Debrun ressalta que, ao falarmos de auto-organização, estamos enfatizando um *processo* no sentido de que ele se desenvolve ou evolui e se consolida através de uma sucessão de estados (mudanças temporais) que recorrem a eles mesmos, ou seja, às suas características intrínsecas. Esse autor afirma (Debrun et al., 1996, p.4) : *o novo no paradigma da auto-organização se origina no próprio processo.*

Segundo, os *inputs* do meio ambiente exercem uma influência mínima durante o processo, o suficiente para não determinar ou controlar o crescimento espontâneo de uma “forma” ou organização.



Terceiro, Debrun (Debrun et al., 1996, p.6) associa dois significados ou justificativas para o uso do prefixo *auto*. Estas justificativas estão associadas à autonomia do processo e ao seu desenrolar. Assim, o processo é auto-organizador no sentido que ele preserva uma autonomia parcial em relação às suas condições de partida. Neste sentido, Debrun (Debrun et al., 1996, p.8) ressalta que o processo “ se desenvolve através de um trabalho de si sobre si “. Uma possível condição de partida citada por Debrun é a existência de elementos não redundantes (sem nenhuma afinidade, ligação ou relação de condicionalidade entre eles). Em suma, elementos com uma distinção real e não analítica que se encontram através de acasos ou por decisões de terceiros (decisões de grupos, de entidades, etc.).

O termo *distinção real* deve ser entendido excluindo a existência de elementos redundantes participando do processo; ou seja, na auto-organização não deve preponderar a existência de elementos amarrados por relações causais, lógicas e nem por uma relação de condicionalidade entre eles. Porém, a falta de afinidades entre os elementos não pode ser total de tal maneira que ficaria inviabilizado o encontro entre estes elementos. Portanto, uma ínfima redundância é necessária para que, como afirma Debrun (1996, p.33) o processo possa “pegar”. Assim uma caracterização que incorpore o aspecto acima é sugerida por Debrun (Debrun et al., 1996, p.13) :

Há auto-organização cada vez que, a partir de um encontro entre elementos realmente (e não analicamente) distintos,



desenvolve-se uma interação sem supervisor (ou sem supervisor onipotente) – interação essa que leva, eventualmente, à constituição de uma “forma “ ou à reestruturação, por “complexificação”, de uma forma já existente.

Mas o que coloca em movimento este processo que se pretende autônomo ? De acordo com Debrun, o motor do processo auto-organizador é a *interação* entre os elementos distintos ou entre partes semi-distintas de um organismo. Ele (1996, p.10) enfatiza ainda que “ a causalidade eminente do processo é a causalidade da interação”. O processo interativo é o responsável pelo aspecto criativo da auto-organização e esta interação não pode ser ditada por regras estabelecidas a priori.

O processo de auto-organização é classificado por Debrun em três estágios. No primeiro estágio, o processo se inicia realizando um *rompimento com o passado*, não sendo um prolongamento nem um desdobramento de processos anteriores. Este começo do processo não é, segundo Debrun (1996, p. 35), dirigido por regras inatas e nem por regras dadas de antemão.

O segundo estágio é denominado, por Debrun (Debrun et al., 1996, p.35), *endogenização*. Nesta fase, o processo se “responsabiliza” por seu desenvolvimento e os eventuais objetivos dos elementos se subordinam ao objetivo global do processo. Debrun aponta ainda três condições para que a endogenização ocorra. Primeiro, o processo não pode ser coordenado nem guiado por um “centro absoluto”, o que o transformaria em um processo hetero-organizado; segundo, durante o

processo de interação um atrator deve ser formado para apontar a direção a ser seguida pelo processo de auto-organização.

O último estágio do processo de auto-organização é aquele em que se cristaliza uma “forma” ou organização caracterizada pela presença de elementos com um certo grau de dependência, formando uma relação permanente entre eles. Nesta etapa, segundo Debrun, ocorre um *ajustamento* entre as partes, que é mantido por uma espécie de “acordo” entre elas.

Após a introdução das características gerais da auto-organização, Debrun, seguindo Ashby (1962), propõe duas modalidades para a auto-organização: a primeira, denominada *auto-organização primária*, é aquela na qual uma interação entre elementos distintos leva à constituição de uma forma ou organização; a segunda, denominada *auto-organização secundária*, é aquela em que ocorre uma reestruturação ou um aumento de complexidade de uma forma ou organização já existente. Estas duas modalidades de auto-organização serão tratadas com mais detalhes na seção seguinte.

Após descrevermos os aspectos gerais do pensamento de Debrun (1996) sobre a idéia de auto-organização, vamos caracterizar o pensamento de Von Foerster (1960) a respeito deste assunto.

Como enfatizamos na primeira seção deste capítulo, Von Foerster foi um dos principais fundadores do movimento cibernético de segunda ordem, cujo tema central foi o fenômeno da auto-organização. Em seu trabalho, publicado em 1960, denominado “On Self-Organizing

Systems and Their Environments” ele desenvolve, no âmbito da física, um conjunto de conceitos sobre os denominados sistemas *self-organizing*.

A análise de Von Foerster (1960) apoia-se nos resultados da concepção estatística da segunda lei da termodinâmica, segundo a qual qualquer sistema termodinâmico fechado evolui, inevitavelmente, para um estado de máxima desordem (entropia). Em oposição a este princípio, Von Foerster argumenta que os sistemas “auto-organizantes” evoluem no sentido de aumentar sua ordem interna, diminuindo, portanto, a sua entropia. Ele argumenta que este aumento de ordem interna não seria possível sem uma estreita interação com o meio ambiente. Portanto, para Von Foerster, somente é possível falarmos em sistemas “auto-organizantes” se considerarmos, conjuntamente o sistema e o ambiente no qual ele se acha inserido. Assim, Von Foerster (1960, p.33) argumenta :

“In spite of this suggested proof of the non-existence of self-organizing systems, I propose to continue the use of the term “self-organizing system,” whilst being aware of the fact that this term becomes meaningless, unless the system is in close contact with an environment, which available energy and order, and with which our system is in a state of perpetual interaction, such that it somehow manages to “live” on the expenses of this environment”.

Dessa forma, de acordo com o pensamento de Von Foerster, não poderia haver auto-organização em um sistema fechado, ou seja, naquele que não efetua troca de informação, energia ou matéria com o meio ambiente.

Ao considerar um sistema *auto-organizante* como aquele capaz de aumentar sua ordem interna, dois pontos preocuparam Von Foerster (1960, p.36) : o significado do termo interno e o problema de como medir a ordem interna gerada. No primeiro, o problema consiste em definir a fronteira do sistema contendo propriedades de auto-organização, já que esses limites podem mudar para regiões que não possuem tais propriedades obrigando a novas escolhas. A saída, segundo Von Foerster, seria considerar aqueles limites que, em um determinado instante de tempo demonstrariam um aumento de ordem. Porém, nessa situação, enfrentamos o problema de como obter algum dispositivo que possa indicar se estamos conectados em uma região "auto-organizante" ou "auto-desorganizante".

Von Foerster (1960) aponta, ainda, uma segunda dificuldade na análise dos sistemas "auto-organizantes", relacionada à capacidade desses sistemas em moverem-se no espaço e mudarem sua forma, além da capacidade de aparecer e desaparecer espontaneamente em locais diferentes, exigindo um dispositivo para que o aumento de ordem seja medido.

Para isso, Von Foerster desenvolve uma formulação quantitativa da noção de ordem interna de um sistema "auto-organizante" apoiando-se no conceito de redundância da teoria da informação de Shannon. Segundo Von Foerster, o conceito de ordem só faz sentido se entendido numa concepção relativa à máxima desordem alcançada pelo sistema, ao invés da desordem absoluta. Assim, seus valores estariam

entre 0 (máxima desordem) e 1 (máxima ordem). Para o cálculo do nível de redundância Von Foerster apóia-se na definição de Shannon dada por $R = 1 - (H / H_{\max})$. Nesta expressão H representa a entropia de uma fonte de informação e $H_{\max}$ o valor máximo dessa entropia. A expressão R representa a medida de ordem, sugerindo que se um sistema se encontra em seu ponto de máxima desordem $H = H_{\max}$, a redundância do sistema é igual a 0; se $H = 0$, temos que a redundância é igual a 1 (ordem perfeita).

Na sua análise dos sistemas "auto-organizadores", Von Foerster procura encontrar princípios de auto-organização. Porém, antes de chegar a sua formulação qualitativa, Von Foerster salienta que um destes princípios foi estabelecido por Schrodinger, no âmbito da física, em sua obra ***O que é vida***, conhecido como o *princípio da ordem a partir da ordem*. A análise de Von Foerster acrescenta o *ruído* como mais um elemento que pode contribuir para disparar um processo auto-organizador. Assim, Von Foerster estabelece o *princípio da ordem a partir do ruído*. Este princípio tem um caráter inovador na medida em que considera o ruído capaz de gerar estruturas organizadas já que, tradicionalmente, o ruído é considerado como uma perturbação indesejável e um fator de desorganização. A proposição de Von Foerster foi exemplificada através da reunião de cubos imantados, que quando agitados ao acaso, produzirão uma disposição ordenada como mostra a Figura 10.



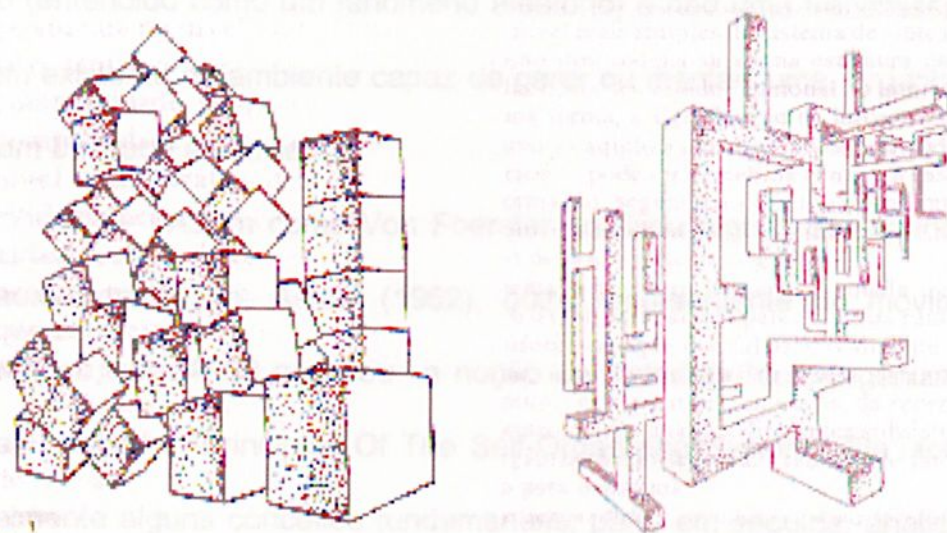


FIGURA 10 – Ímãs do experimento de Von Foerster. a) antes da agitação b) após a agitação. Extraídos de Von Foerster em Yovitz & Cameron (1960, p. 44-5)

Von Foerster (1960, p.45) argumenta que, no exemplo acima, nenhuma ordem foi transmitida do ambiente para o sistema, referindo-se ao princípio da ordem a partir da ordem, mas que *"noise were selected which contributed to the increase of order system"*. A emergência da ordem interna no exemplo acima foi possível graças ao ruído e à propriedade intrínseca (a força de atração entre elementos magnéticos de polarização contrárias) dos elementos.

Apesar de não analisar a noção de ordem, Von Foerster contribuiu para sedimentar a noção de ruído (com efeitos positivos ou construtivos) como um princípio de auto-organização e propondo uma formulação quantitativa para a idéia de ordem.

A geração de ordem é associada ao caráter imprevisível do ruído (entendido como um fenômeno aleatório) e não uma transmissão de ordem existente no ambiente capaz de gerar ou manter uma organização (ordem baseada em ordem).

Assim como Von Foerster analisou alguns dos princípios de auto-organização, Ashby (1962), outro representante do movimento cibernético, em 1962 analisou a noção de sistemas "auto-organizantes" em seu trabalho "Principles Of The Self-Organizing System". Ele formula inicialmente alguns conceitos fundamentais, para, em seguida, analisar as possibilidades de auto-organização tanto nos sistemas naturais como nos sistemas artificiais.

Em primeiro lugar Ashby (1962, p.256) analisa o conceito de *organização*. Para ele, a noção de organização compreende vários aspectos, porém o aspecto central deste conceito reúne os conceitos de *condicionalidade* e *separabilidade*. Uma organização emerge quando uma relação de condicionalidade se estabelece entre seus elementos. Em oposição, uma organização não se forma quando um forte grau de separabilidade se estabelece. Fundamentada na idéia de condicionalidade, a noção de organização seria passível de tratamento quantitativo através de métodos matemáticos para calcular o grau de condicionalidade presente na organização, particularmente pela teoria das funções matemáticas de mais de uma variável. Para Ashby (1962, p.257) uma condicionalidade se estabelece quando uma *comunicação* ocorre entre elementos distintos; neste momento dizemos que eles estão organizados.



Além disso, uma comunicação entre partes se sedimenta quando *restrições (constraints)*, se estabelecem, ou seja, quando comportamentos em um elemento disparam comportamentos em outros elementos. Enfim, a noção de organização é entendida como uma *restrição*, a qual está apoiada na *lógica da comunicação*. Ashby (1962, p.258) enfatiza, ainda, que a noção de organização é dependente do observador, de tal maneira que as restrições que caracterizam a organização abarcam tanto o fenômeno real quanto o observador. Assim, a noção de organização para Ashby possui a peculiaridade de que ela emerge da relação entre o observador e a coisa observada não se prendendo exclusivamente nas propriedades intrínsecas das coisas.

Outro conceito analisado por Ashby (1962, p.261) é o conceito de *máquina*. Para ele uma máquina pode ser definida através de seu comportamento, ou seja, seu comportamento deve ser ditado por uma lei ou regra. Esta lei, obedecida pela máquina, é responsável por levá-la a atingir um estado futuro apoiando-se no estado atual e nos estados anteriores. Devido à necessidade de se caracterizar a noção de máquina de tal forma que possa incorporar eventos contínuos, descontínuos e discretos, Ashby (1962, p.261) sugere um conceito mais geral de máquina denominado de "*machine with input*". O autor enfatiza que uma máquina pode ser caracterizada como uma função que relaciona um conjunto de entradas (I) e um conjunto de estados internos (S), resultando em novo estado interno (S'). Matematicamente, a máquina poderia ser representada por $f : I \times S \rightarrow S$, onde f representa o funcionamento ou a organização



funcional da máquina. A esta noção de máquina não é relevante sua constituição material e nem se existe ganho ou perda de energia; o que importa é a regularidade de seu comportamento.

Outra idéia importante na análise de Ashby (1962, p.262) é a idéia de *boa organização*. Ashby procura identificar critérios que possam distinguir boas organizações daquelas que exibem um comportamento inadequado. Uma boa organização, para Ashby, é aquela capaz de manter adequadamente os parâmetros fundamentais para sua sobrevivência diante de perturbações ambientais. Este princípio é geral, sendo válido tanto para sistemas naturais quanto para artificiais. Ashby argumenta que não existem organizações adequadas no sentido absoluto. Isto é, o conceito de boa ou má organização é relativo ao contexto ou ao ambiente em que uma organização se acha inserida. Entretanto, podemos dizer que uma boa organização deve ser capaz de se auto-modificar quando submetida às perturbações ou mudanças que ocorrem em um determinado ambiente, de modo a preservar a sua existência.

A partir das noções de organização, máquinas e boa/má organização, Ashby caracteriza um processo de auto-organização. Neste sentido ele identifica duas manifestações das propriedades dos sistemas "auto-organizantes". A primeira ocorre quando elementos separados, sem nenhum grau de condicionalidade entre eles, interagem, estabelecendo conexões e criando relações de dependência entre si. Ashby (1962, p. 267) enfatiza que o sentido do termo auto-organização está associado à formação, neste processo, da emergência de uma estrutura

que se altera de partes independentes e desorganizadas para partes conectadas e organizadas. Em seu trabalho Ashby (1962, p.266) argumenta: *Such a system is "self-organizing" in the sense that it changes from "parts separated" to "parts joined"*. Ele argumenta ainda que este tipo de auto-organização poderia ser melhor caracterizado como *self-connecting* já que o que ocorre nesta fase do processo é o estabelecimento de condicionalidades entendidas como um tipo de conexão entre elementos, que estavam inicialmente separados.

Ashby identifica ainda uma segunda manifestação do processo de auto-organização. Neste caso, um sistema já constituído muda, por meio de seus recursos próprios, de um comportamento inadequado para um comportamento adequado, de acordo com condições ou perturbações aleatórias do meio ambiente. Sistemas que incorporam mecanismos de *feedback* e capacidades de aprender são exemplos de sistemas que conseguem se auto-modificar, alcançando comportamentos adequados que garantam a sua sobrevivência relativamente ao seu meio ambiente.

Analisando a aplicação desse segundo sentido de auto-organização, no campo das máquinas, Ashby (1962, p.267) conclui que é logicamente impossível que estes sistemas exibam a propriedade de auto-organização. Dado que o comportamento das máquinas, como definidas por Ashby, é regido por uma função pré-estabelecida que determina seu estado futuro com base no seu estado presente e em *inputs* recebidos, ela não poderá auto-organizar-se. Para que a máquina pudesse exibir este

segundo aspecto de auto-organização, ela teria que ser capaz de modificar a função f que a caracteriza, apenas com base em seu estado interno. Para Ashby esta mudança é impossível, a não ser que seja transmitida de fora do sistema por algo que agisse como *input* organizador, comprometendo, assim, o sentido estrito do termo *auto*. Além disso, Ashby enfatiza que as máquinas não teriam, em princípio, a capacidade de filtrar os estados mais apropriados para sua adaptação e desprezar os estados inadequados para manter a sua organização.

Em resumo, Ashby acredita que o emprego do termo *auto-organizante* seria problemático se usado em sistemas isolados como as máquinas, uma vez que o meio ambiente é importante para desencadear um processo de auto-organização. Como as máquinas são caracterizadas em termos de funções para a alteração de estados seguindo regras preestabelecidas, elas não computariam atribuições fundamentais do meio ambiente em sua estrutura organizacional.

Na próxima seção apresentamos duas modalidades de auto-organização, tal como estabelecidas por Ashby (1962) e Debrun (1996).

Neste tipo de auto-organização denomina-se de "elementos organizados" os elementos organizados que se combinam para a construção de uma forma (Debrun, 1996, p.28) e não de uma organização já existente. Formalmente, Debrun (1996, p.32) dá a seguinte caracterização para o processo de auto-organização primária:

Há auto-organização primária toda vez que um processo de montagem e integração entre elementos realmente distintos, sendo

Seção III – Auto-organização primária e auto-organização secundária

Conforme analisamos na seção anterior, o processo de auto-organização caracteriza-se pela dinâmica da interação entre elementos, dinâmica esta que leva à formação espontânea de padrões, comportamentos ou estruturas organizadas. Enfatizamos, também, que a formação ou o desenvolvimento destes padrões não é fruto de um agente ou elemento externo, de um elemento interno e nem de leis ou regras que controlam ou supervisionam o processo.

As caracterizações do conceito de auto-organização procuram ressaltar os aspectos acima citados; porém, algumas perspectivas identificam dois estágios no processo de auto-organização. Debrun (1996) e Ashby (1962), por exemplo, enfatizam que a noção de auto-organização pode ser analisada sob duas modalidades. A primeira, denominada por Debrun (1996) de *auto-organização primária*, é aquela que não parte de uma forma ou organização já existente. Ao contrário, ela se inicia a partir de um conjunto de elementos independentes que, através de uma interação espontânea, produzem a geração de uma forma ou estrutura organizada. Neste tipo de auto-organização caminha-se de “elementos avulsos para a constituição de uma forma” (Debrun, 1996, p.28) e não de uma organização já existente. Formalmente, Debrun (1996, p.32) dá a seguinte caracterização para o processo de auto-organização primária :

Há auto-organização primária cada vez que um processo de encontro e interação entre elementos realmente distintos tende, sem

intervenção de um supervisor, pelo menos sem supervisor onipotente, à constituição de uma forma global (ou "Todo"), que resulta essencialmente da própria interação, e só em grau menor das condições de partida ou do intercâmbio com o ambiente.

É durante a auto-organização primária que o aspecto criador do processo fica evidente pois dele resulta um ser ou organismo ou, ainda, uma *forma* entendida como uma tendência. Debrun (1996, p.14) destaca alguns pontos que impulsionam o aspecto criativo da auto-organização primária. Inicialmente, o encontro de elementos distintos pode dar lugar a um aspecto criativo desde que ocorra um rompimento com os eventos bem estabelecidos que porventura provocaram este encontro. Assim, por exemplo, pode ser que uma nova organização surja de um encontro marcado, pré-fixado. Contudo, esta organização apenas se caracterizará como primariamente auto-organizada a partir do momento em que ela estabelecer uma ruptura com o passado e adquirir uma dinâmica própria. Em segundo lugar, a natureza dos elementos que interagem durante o processo deve ser tal que o aspecto inovador do processo será tanto maior quanto mais distintos forem os elementos que participam do processo pois, para este autor, a possibilidade de estabelecer novas relações entre estes elementos será maior.

A modalidade de auto-organização primária pode também ser encontrada nos estudos de Ashby (1962). Este autor, apesar de não utilizar o termo *primária* explicitamente, caracteriza a auto-organização primária como o processo no qual um sistema evolui de partes separadas

para partes conectadas (Ashby, 1962, p.266). Durante este processo, Ashby enfatiza que uma organização é produzida quando os elementos, inicialmente com comportamentos independentes, passam a estabelecer conexões vistas como relações de condicionalidades entre eles.

Debrun (1996, p.13) identifica uma segunda modalidade de auto-organização denominada de auto-organização secundária. O aspecto secundário desta modalidade de auto-organização diz respeito ao seu desenvolvimento a partir de uma forma (tendência) ou organização já constituída e se reestrutura para atingir um patamar de complexidade superior. Esta complexidade, como afirma Debrun (1996, p.26), pode ser corporal, intelectual ou existencial. Esta reestruturação, na visão de Debrun, ocorreria através de um processo de aprendizagem exercido pelo organismo sobre ele mesmo com seus próprios mecanismos e com a troca de matéria, energia ou de informação com o meio ambiente. Debrun (1996, p. 26) justifica assim o uso do prefixo auto nesta modalidade de auto-organização :

O que é "auto", aqui, é que tanto o ponto de partida (a "decisão") como o ponto de aplicação (tal parte do corpo, por exemplo) bem como os mecanismos utilizados e parte dos recursos, situam-se dentro de um mesmo organismo .

Ele aponta, ainda, algumas similaridades entre os processos de auto-organização primária e secundária. Entre eles o fato de que nas duas modalidades o que impulsiona o processo de auto-

organização é mais a interação entre as partes, distintas na modalidade de auto-organização primária e semi-distintas na modalidade de auto-organização secundária Debrun (1996, p.270).

Assim como Debrun (1996), Ashby (1962) também identifica um segundo sentido do processo de auto-organização. Este autor enfatiza que na auto-organização pode ocorrer uma variação do grau de complexidade de uma organização: de complexidade inferior para uma de complexidade superior. Esta mudança para Ashby (1962) seria possível nas organizações que incorporam mecanismos de aprendizagem e de *feedback* permitindo a elas adequar seu comportamento de maneira a manter seu funcionamento ou sua sobrevivência. Ashby (1962, p.267) argumenta, ainda, que as máquinas não teriam a capacidade de se auto-organizar secundariamente, pois suas mudanças de estados não poderiam ser executadas com base apenas em seus estados anteriores, mas com a colaboração de um agente externo que determinasse sua organização.

Ashby (1962, p.269) argumenta que a inexistência de um fluxo de energia, matéria ou informação entre a máquina e o meio ambiente impede que ela se auto-organize secundariamente não conseguindo as perturbações do meio externo. À medida que uma máquina, tal como caracterizada por Ashby, é um sistema isolado, ela não teria, em princípio, recursos para se auto-organizar.

Levando em conta tal conclusão, investigaremos, na próxima seção, os possíveis elos de conexão entre a teoria da auto-organização e a ciência cognitiva.

Seção IV – Ciência cognitiva e auto-organização

Conforme vimos nos capítulos anteriores, a ciência cognitiva e a teoria da auto-organização possuem suas origens no campo da cibernética, porém, algumas vezes seguem orientações teóricas distintas.

Assim, Debrun (1996, p.29) argumenta que, enquanto a ciência cognitiva preocupa-se em compreender a natureza do conhecimento, no qual o “sujeito é uma instância cognitiva por excelência” a teoria da auto-organização está centrada nas propriedades que podem emergir a partir das interações de elementos de várias naturezas, sem que o sujeito desempenhe um papel principal.

Mas que sujeito é este a que se refere Debrun, que constitui a preocupação central na ciência cognitiva ? Certamente não se trata do sujeito cartesiano, dotado de corpo e alma. Ainda que não se disponha de uma conceituação precisa da noção sujeito, está claro nos trabalhos dos cognitivistas que não estamos diante de um sujeito transcendental. Nas palavras de Gonzales (1998, p. 9):

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 à concepção humanista – que lhe garantiria o tatus 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.

No lugar do sujeito cognitivo clássico, os cognitivistas apresentam um sistema de processamento de informações. Na vertente da inteligência artificial este sistema será composto de um software e um hardware (diluindo-se, assim, o problema da relação mente-corpo). Temos aí apenas um sistema manipulador de símbolos.

Na vertente conexionista, este sistema de processamento de informação assume a forma de uma rede neural, que se auto-organiza. Tal caracterização do sistema cognitivo possui uma clara proximidade com a caracterização do sujeito no paradigma da auto-organização: "...Situado em um ambiente efetivo, este sujeito, possui uma história ... constrói a sua identidade à medida em que age no meio ambiente..." (Gonzales, 1998, p.10).

Assim, a seguinte questão poderia ser levantada: quais são os pontos (metodológicos) em comum entre a ciência cognitiva e a teoria da auto-organização? Uma resposta a esta pergunta consiste em ressaltar que a convergência ocorre quando a ciência cognitiva trata a atividade cognitiva como um processo emergente da dinâmica da interação entre um conjunto de elementos. Tal postura converge para a definição de auto-organização fornecida por Debrun (1996) e por Ashby (1962). De acordo com Morin (1996, p.280), propriedades emergentes são aquelas que se manifestam "em um todo organizado, e que não existem nas partes isoladas do todo".

Para Debrun (1996), as duas áreas seriam compatíveis a partir do momento em que a teoria da auto-organização analisa a atividade

cognitiva, entendida não em si mesma, mas como um elemento, entre outros, que interage para estruturar alguma forma. Ainda, segundo Debrun (1996), isto ocorre quando se considera apenas um dos modelos estudados pela ciência cognitiva, a saber, o *modelo conexionista*, que supostamente poderia exibir características de auto-organização.

Portanto, uma possível convergência entre as hipóteses dos cientistas cognitivos e dos teóricos da auto-organização ocorrerá a partir do momento em que a ciência cognitiva se tornar “predominantemente conexionista” (Debrun, p.37, 1996).

Um dos modelos conexionistas que incorpora propriedades de auto-organização, e que pode servir como ponto de encontro entre a ciência cognitiva e a teoria da auto-organização, será analisado no capítulo seguinte.



Apresentação

Conforme analisamos no capítulo anterior, as interfaces conceituais da ciência cognitiva e da teoria da auto-organização convergem para os modelos conexionistas. Assim, para demonstrar esse ponto de conexão entre os dois campos, vamos abordar, neste capítulo, um modelo clássico em inteligência artificial denominado rede neural de Kohonen, no qual supostamente existem elementos do processo de auto-organização.

CAPÍTULO V – APLICAÇÕES DA TEORIA DA AUTO-ORGANIZAÇÃO

processo de aprendizagem. Padrões de conectividade emergem neste rede sem a necessidade de referências externas além dos padrões informacionais do meio ambiente. Do ponto de vista biológico, este foi inspirado em um fenômeno denominado "inibição lateral", no qual circuitos inibitórios inibem uma área mantendo áreas vizinhas excitadas recebidas do meio ambiente, particularmente das células vizinhas.

O presente capítulo divide-se em três partes: na primeira, mostramos alguns aspectos do fenômeno clássico de rede de Kohonen se inspirou para construir seu modelo; na segunda, apresentamos o modelo de Kohonen, destacando sua estrutura e suporte matemático, bem como suas propriedades de auto-organização; na última parte, discutimos a formação do mapa de características auto-organizadas, através dos resultados obtidos por simulação computacional do algoritmo de Kohonen.

Apresentação

Conforme analisamos no capítulo anterior, os interesses conceituais da ciência cognitiva e da teoria da auto-organização convergem para os modelos connexionistas. Assim, para demonstrar este ponto de conexão entre as duas áreas, iremos abordar, neste capítulo, um modelo connexionista específico, denominado rede neural de Kohonen, no qual supostamente estariam elementos do processo de auto-organização.

A rede neural de Kohonen é caracterizada como possuidora de um tipo especial de treinamento denominado treinamento *não supervisionado* ou por *auto-organização*. Neste treinamento nota-se a ausência de um supervisor ou controlador responsável pela condução do processo de aprendizado. Padrões de conectividade emergem nesta rede sem a necessidade de referências externas adicionais além dos padrões informacionais do meio ambiente. Do ponto de vista biológico, esta rede foi inspirada em um fenômeno denominado *inibição lateral*, no qual circuitos inibitórios laterais contribuem para uma melhor definição dos padrões recebidos do meio ambiente, particularmente dos padrões visuais.

O presente capítulo divide-se em três partes: na primeira, mostramos alguns aspectos do fenômeno biológico no qual Kohonen se inspirou para construir seu modelo; na segunda, analisamos o modelo de Kohonen, destacando sua estrutura e suporte matemático, bem como suas propriedades de auto-organização; na última parte avaliamos a formação do mapa de características auto-organizadoras, através dos resultados obtidos por simulação computacional do algoritmo de Kohonen.



Seção I – Plausibilidade biológica da rede de Kohonen

A rede neural artificial de Kohonen foi biologicamente inspirada em alguns princípios do funcionamento do cérebro quando este processa informação. Uma das linhas de estudo do cérebro segue o princípio da divisão de funções, onde regiões cerebrais desempenham tarefas específicas, tais como: o controle da fala e das funções motoras, o reconhecimento de padrões visuais e auditivos, entre outros etc. Ao elaborar uma rede neural artificial, Kohonen partiu da constatação que, ao processar informação, o cérebro forma “representações neurais que são imagens de experiências sensoriais, projetadas sobre várias áreas corticais” (Kohonen, 1997, p.79). Tais representações foram denominadas de mapas auto-organizadores. A Figura 11. ilustra o *tonotopic map* representado nas regiões cerebrais, responsáveis pelas funções auditivas.

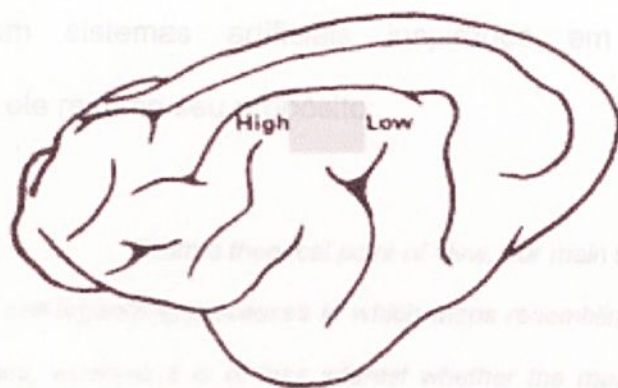


FIGURA 11 - *Tonotopic map*. (Kohonen, 1988, p. 6)

O mapa auditivo corresponde a uma disposição espacialmente ordenada de agrupamento neuronal, excitável às frequências acústicas específicas, de padrões auditivos recebidos do meio ambiente. Tal mapa é um exemplo que ressalta a importância da localização espacial das respostas neurais.

Para Kohonen (1997, p.81), as estruturas cerebrais espacialmente organizadas podem ser explicadas a partir do princípio da *auto-organização*. A propriedade principal que leva à formação de tais organizações deve-se principalmente às *características dos padrões sensórios presentes no meio ambiente*. Assim, áreas específicas do cérebro se auto-organizam em função das características dos padrões sensórios cujo efeito é a formação do que Kohonen (1987, p.7) denomina *mapa de características ordenadas*.

O objetivo de Kohonen (1997, p.82) foi produzir um algoritmo que simulasse o mapeamento das características auto-organizadoras em sistemas artificiais inspirados em propriedades biológicas. Assim ele resume seu propósito:

"From a theoretical point of view, our main task is thus to find abstract self-organizing processes in which maps resembling the brain maps are formed, whereas it is of less interest whether the maps are formed by evolution, postnatal growth, or learning."

O modelo de Kohonen consiste de um conjunto de unidades organizadas de maneira bidimensional, através do mecanismo de

aprendizado por auto-organização, nas atividades de reconhecimento de padrões. Detalhes estruturais e funcionais específicos dessa rede serão analisados na próxima seção deste capítulo, onde procuramos destacar seus aspectos de auto-organização e a posição especial que ela ocupa entre os modelos de redes neurais artificiais.

Outro aspecto funcional do sistema nervoso no qual se apoiou Kohonen para desenvolver seu sistema artificial foi o mecanismo da *inibição lateral*. Kovács (1997, p. 63) sugere que, através da inibição lateral, as transições contidas em padrões espaciais são discriminadas, aumentando o contraste dos estímulos ou padrões recebidos do meio ambiente. As Figuras 12 e 13 ilustram um esquema de inibição lateral em que circuitos neuronais inibitórios limitam a ativação das unidades vizinhas de um agrupamento neuronal mais ativado.

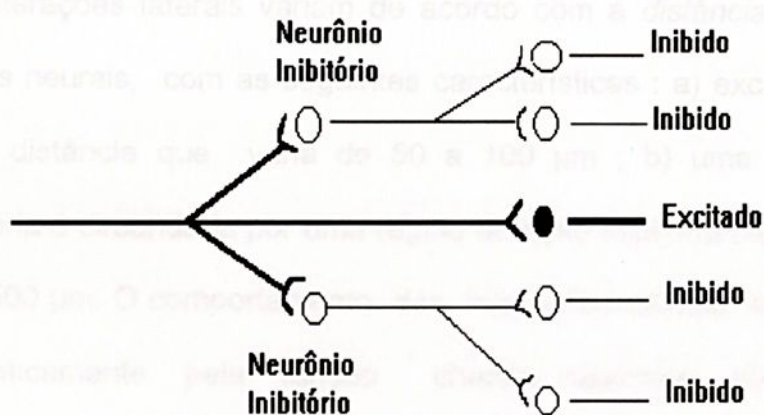


FIGURA 12 - Circuito inibitório lateral através do qual uma fibra excitada de um grupamento neuronal pode determinar inibição de fibras adjacentes. (Guyton, 1989)

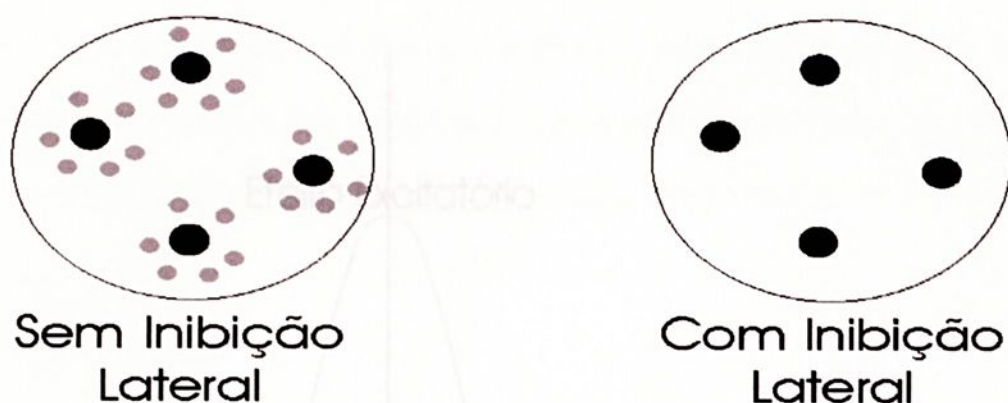


FIGURA 13 - Aumento no contraste do padrão de estímulo determinado pelo circuito inibitório. (Guyton, 1989)

Nota-se, na Figura 13 acima, que, ao final do mecanismo de inibição lateral, circuitos neurais passam a deter um maior nível de ativação, cujo efeito é a formação de padrões ambientais melhor definidos.

Segundo Kohonen (1988,p.122), existem evidências anatômicas e fisiológicas que no cérebro dos mamíferos o comportamento das interações laterais variam de acordo com a *distância* ao longo dos circuitos neurais, com as seguintes características : a) excitações laterais a uma distância que varia de 50 a 100 μm ; b) uma área neuronal excitatória é circundada por uma região de ação inibitória cujo raio varia de 200 a 500 μm . O comportamento das interações laterais é representado matematicamente pela função chapéu mexicano (Kohonen, 1988, p.123), tal como ilustrada na Figura 14.

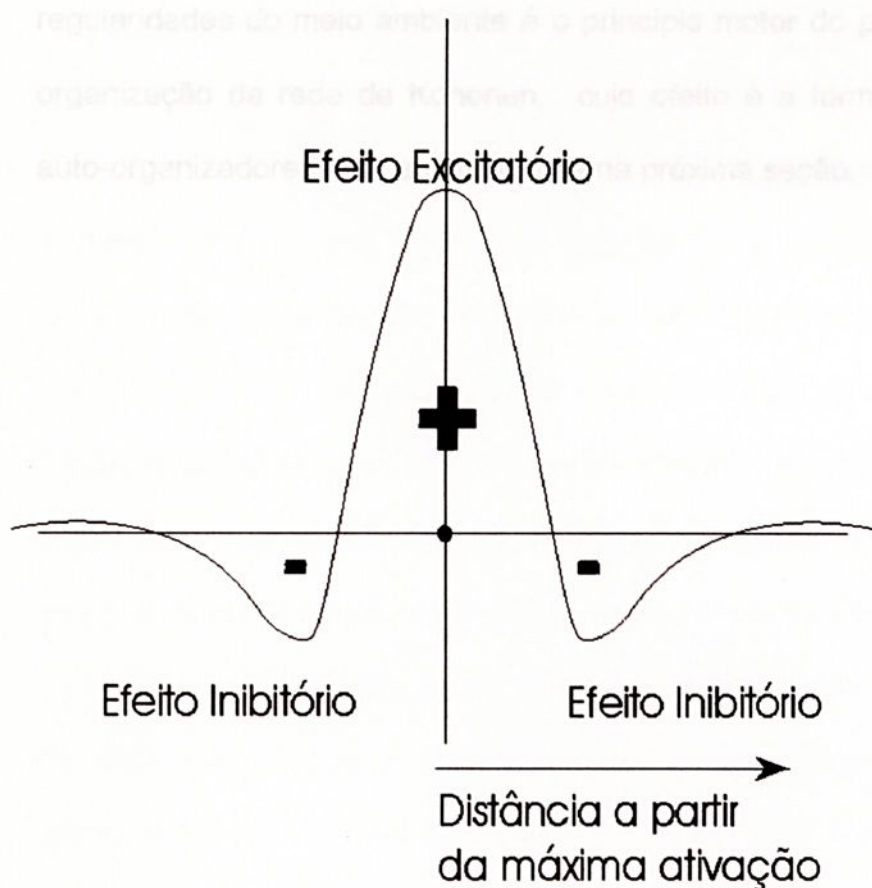


FIGURA 14 - Função chapéu mexicano. Kohonen (1988, p.123)

Na Figura 14, as unidades mais próximas de um centro mais ativado tendem a aumentar seus níveis de ativação através de conexões laterais excitatórias, e as unidades mais distantes tendem a reduzir seus níveis de ativação por meio de conexões inibitórias. Este princípio, segundo o qual unidades vizinhas se ativam em torno de uma unidade com nível de ativação mais elevado, formando um *cluster* de unidades que respondem a específicos padrões do ambiente, foi incorporado na rede de Kohonen para implementar um algoritmo de aprendizado competitivo ou por auto-organização. As interações competitivas entre um conjunto de unidades vizinhas mais afinadas com as

regularidades do meio ambiente é o princípio motor do processo de auto-organização da rede de Kohonen, cujo efeito é a formação dos mapas auto-organizadores, que analisaremos na próxima seção.

Kohonen possui uma memória associativa distribuída, incorporando métodos de aprendizagem denominados Self-Organizing Feature Maps (SOFM). O SOFM é um algoritmo construído para simular aspectos do processo de auto-organização em estruturas em camadas, tais como o mapa de características auditivas (fonotópico) citadas na seção anterior.

Conforme mencionamos, a rede de Kohonen é um exemplo de RNA que utiliza o treinamento não supervisionado ou por auto-organização. Ao contrário das redes neurais artificiais supervisionadas, esta rede realiza um mapeamento dos padrões de entrada sobre uma camada de neurônios em uma única etapa, sem a necessidade de aprendizagem por retropropagação de erros. Ou seja, não é necessária a obtenção de uma saída desejada para a rede, pois a rede é treinada para aprender a reconhecer os padrões de entrada e mapeá-los para uma saída específica.

A rede de Kohonen é formada por uma única camada de neurônios, que são treinados para reconhecer os padrões de entrada e mapeá-los para uma saída específica. A rede de Kohonen é formada por uma única camada de neurônios, que são treinados para reconhecer os padrões de entrada e mapeá-los para uma saída específica. A rede de Kohonen é formada por uma única camada de neurônios, que são treinados para reconhecer os padrões de entrada e mapeá-los para uma saída específica. A rede de Kohonen é formada por uma única camada de neurônios, que são treinados para reconhecer os padrões de entrada e mapeá-los para uma saída específica.

Seção II – Auto-Organização na rede neural de Kohonen

Além das propriedades enunciadas na seção I, a rede de Kohonen possui uma memória associativa distribuída, incorporando métodos de aprendizado, denominados *Self-Organizing Features Maps* (SOFM). O SOFM é um algoritmo construído para simular aspectos do processo de auto-organização, em particular, na produção de mapas auto-organizadores em localizações específicas do cérebro, tais como o mapa de características auditivas (*tonotopic map*) citadas na seção anterior.

Conforme mencionamos, a rede de Kohonen é um exemplo de RNA que utiliza o treinamento não supervisionado ou por auto-organização. Ao contrário das redes neurais artificiais supervisionadas, ela realiza um mapeamento topológico dos padrões de entrada sobre uma camada bidimensional de elementos sem a necessidade de um supervisor externo. Ou seja, não é apresentado à rede nenhum padrão de entrada ou referência externa que informe, a priori, a resposta ou o comportamento desejado.

A estrutura da rede de Kohonen é formada por duas camadas de unidades ou *neuron-like units*: a *camada de entrada* e a *camada de Kohonen*, conforme ilustrado na Figura 15. A camada de entrada recebe os estímulos ou padrões do meio ambiente e os transmite para cada unidade da camada de Kohonen formando uma coleção de *fan-outs*. A camada de Kohonen configura-se por um *array* bidimensional de unidades que recebem, integralmente, o padrão de entrada funcionando,

cada unidade, como um nó de saída da rede. Como as conexões entre as camadas de entrada e de Kohonen são mediadas pelo peso das conexões, cada unidade da camada de Kohonen recebe o padrão de entrada total com diferentes intensidades.

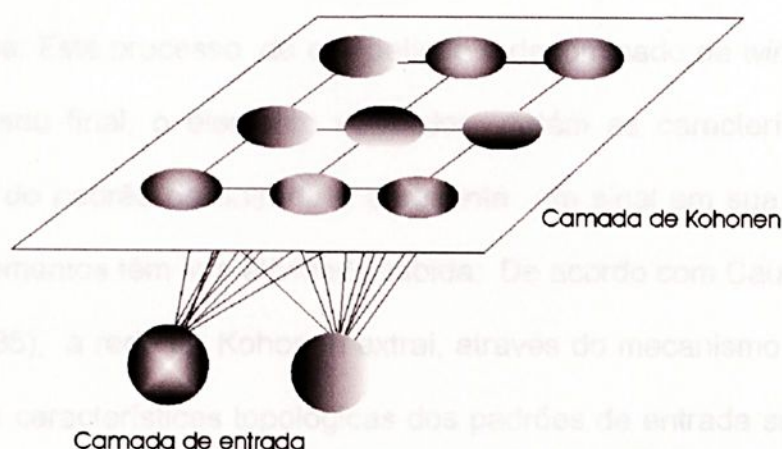


FIGURA 15 - Kohonen feature map. (Beale & Jackson, 1994, p. 110)

Além das conexões recebidas das unidades de entrada, as unidades da camada de Kohonen se conectam lateralmente com as unidades fisicamente vizinhas. As conexões laterais intracamadas modelam a característica do fenômeno da inibição lateral, dando origem a um processo de *competição* no qual uma unidade se ativa fortemente, desencadeando uma série de inibições e ativações em suas unidades vizinhas. A unidade com nível de ativação mais alto é denominada de elemento *vencedor*. Este elemento, juntamente com as unidades vizinhas competem entre si, para extrair as regularidades dos padrões de entrada e fornecerem uma classificação para esses padrões. Na rede de Kohonen as interações laterais se comportam de acordo com a função chapéu

mexicano, ilustrada na seção anterior. Esta incrementa, através de conexões excitatórias, o nível de ativação do elemento vencedor e das unidades mais próximas a ele, e reduz, através de conexões inibitórias, o nível de ativação das unidades mais distantes. O grupo de elementos da vizinhança é aquele que melhor responde às características dos padrões de entrada. Este processo de competição é denominado de *winner-talk-all*, pois, ao seu final, o elemento vencedor detém as características mais próximas do padrão de entrada e apresenta um sinal em sua saída e os outros elementos têm sua atividade inibida. De acordo com Caudill & Butler (1992, p.85), a rede de Kohonen extrai, através do mecanismo da inibição lateral, as características topológicas dos padrões de entrada sem recorrer a um *supervisor externo* para determinar arbitrariamente o vencedor do processo de competição.

O aprendizado de uma rede neural típica envolve o ajuste gradual dos pesos de todas as conexões entre seus elementos. No aprendizado por auto-organização (competitivo) da rede de Kohonen nem todos os elementos participam do processo de aprendizado. Apenas o *elemento vencedor* e os *elementos confinados em sua vizinhança* modificam ou ajustam os pesos de suas conexões durante a apresentação dos padrões de entrada. Os elementos cujos pesos serão ajustados são definidos de acordo com o raio e a disposição espacial daqueles em uma geometria em torno do elemento vencedor. A Figura 16 ilustra possíveis arranjos dos elementos da vizinhança:

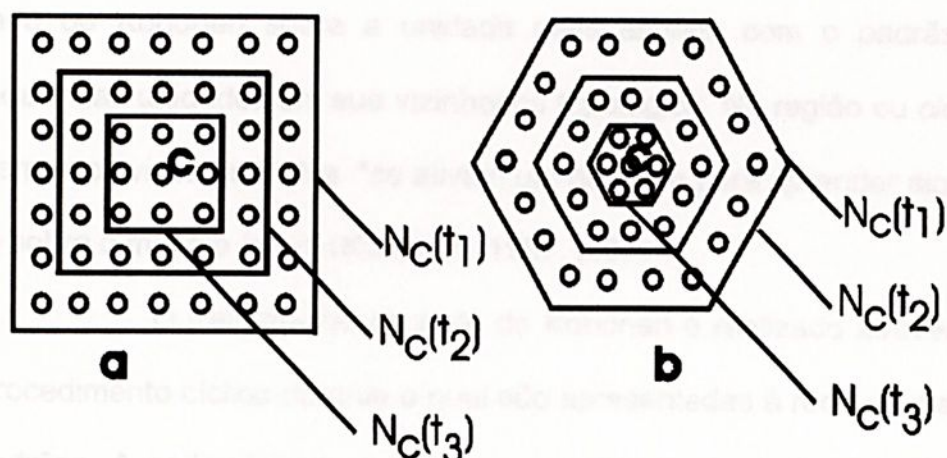


FIGURA 16 - Dois exemplos de vizinhança topológica: a) retangular, b) hexagonal

(Kohonen, 1997, p.87)

Durante o processo de ajuste dos pesos, a vizinhança é gradualmente reduzida, restando, ao seu final, apenas o elemento vencedor. Kohonen (1997, p.54) enfatiza que, se o raio da vizinhança adotado for muito pequeno não ocorre a formação do mapa de característica auto-organizadora. Para se alcançar tanto a formação do mapa de características quanto a convergência do processo, deve-se adotar, segundo Kohonen, um grande raio de vizinhança inicial e produzir sucessivas reduções até a convergência ser alcançada.

A seguir, descrevemos mais detalhadamente o algoritmo básico de aprendizado da rede de Kohonen, que permite mapear as características topológicas do espaço de entradas sobre um *array* de elementos bidimensionais, produzindo mapas espacialmente ordenados. De acordo com Kohonen (1990, 1466), mapas auto-organizadores são formados essencialmente devido à concentração espacial da atividade da

camada de Kohonen sobre a unidade mais afinada com o padrão de entrada e das unidades em sua vizinhança topológica. Na região ou *cluster* de elementos vizinhos, estes "se ativam um ao outro para aprender alguma coisa sobre o mesmo *input*" (Kohonen, 1997, p.87).

O treinamento da rede de Kohonen é realizado através de um procedimento cíclico durante o qual são apresentadas à rede uma série de padrões. A cada ciclo de aprendizagem ocorre um ajuste *espontâneo* dos pesos das conexões de um grupo de elementos vizinhos, espacialmente localizados, sensíveis às regularidades do padrão de entrada corrente, recebido do meio ambiente.

A determinação da unidade com o nível alto de ativação e suas unidades vizinhas são os pontos centrais do algoritmo de Kohonen. Assim, o seu modelo matemático incorpora mecanismos que realizam uma busca seqüencial sobre os elementos da camada de Kohonen para verificar a unidade mais ativa e a topologia de sua vizinhança.

No primeiro passo do algoritmo são definidos alguns parâmetros básicos para dar início ao processo de treinamento. Os pesos das conexões entre a camada de entrada e a camada de Kohonen recebem valores iniciais aleatórios, que ampliam ou reduzem os níveis de ativação recebidos pelos elementos da camada de Kohonen. Um raio e uma geometria são estabelecidos para selecionar os elementos que participarão do processo de aprendizado. Uma taxa de aprendizado ou ganho indica que fração da diferença entre o vetor de entrada e o vetor

Kohonen.

peso das conexões, em um determinado instante, será utilizada para ajustar os pesos das conexões no instante seguinte.

No passo seguinte, para cada padrão de entrada apresentada à rede, calcula-se a distância euclidiana D_j entre o vetor de entrada $\mathbf{x} = \{x_1, x_2, x_3, \dots, x_n\}^T$ e o vetor de pesos $\mathbf{w} = \{w_{11}, w_{12}, w_{13}, \dots, w_{1n}, w_{21}, w_{22}, \dots, w_{2j}, w_{j2}, \dots, w_{ij}\}^T$ das j unidades da camada de Kohonen, para verificar o grau de similaridade entre os dois vetores :

$$D_j = \sum_{i=1}^n (x_i(t) - w_{ij}(t))^2$$

FIGURA 17 - Rede de Kohonen

O elemento j , da camada de Kohonen, que apresentar a menor distância euclidiana será escolhido como o vencedor do processo e seu vetor de pesos, bem como dos elementos vizinhos são ajustados de acordo com a seguinte regra de aprendizado:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t) (x_i(t) - w_{ij}(t)) \text{ para todo } j \in N_c(t), 0 \leq i \leq n-1$$

De acordo com Kohonen (1997, p.87), $\alpha(t)$ e $N_c(t)$ são funções decrescentes com o tempo e representam, respectivamente, o ganho ou taxa de aprendizado (variando entre 0 e 1) e o raio da vizinhança em torno do elemento vencedor. O processo é repetido com a apresentação de novos padrões de entrada. A Figura 17 ilustra a rede de Kohonen.

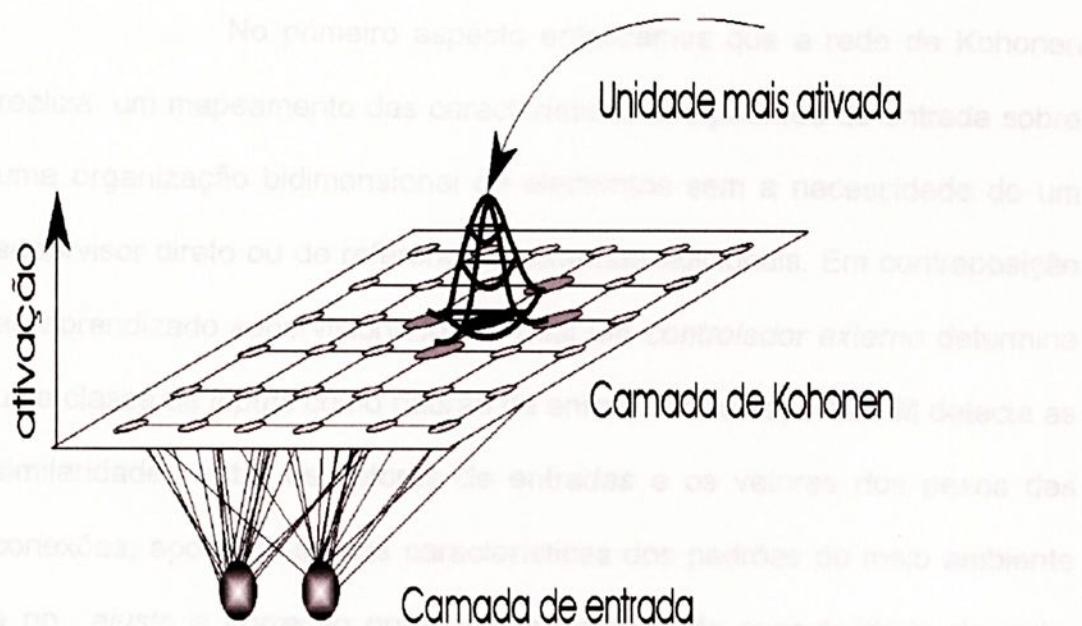


FIGURA 17 - Rede de Kohonen

A partir da descrição e análise do algoritmo de aprendizado por auto-organização ou SOFM de Kohonen, destacamos alguns aspectos funcionais com implicações relevantes para o estudo do processo da auto-organização. Conforme ressaltamos em parágrafos anteriores, a proposta subjacente ao SOFM é de simular a formação do mapa de características auto-organizadoras em sistemas artificiais. Mais precisamente, a rede de Kohonen simula a “criação de *representações internas*, especialmente organizadas, das características de vários padrões sensórios do meio ambiente” (Kohonen, 1990, p. 1464). Assim, através do mapeamento das regularidades do meio ambiente, uma representação topográfica é formada sobre um conjunto de unidades vizinhas que respondem similarmente aos padrões ambientais.

No primeiro aspecto enfatizamos que a rede de Kohonen realiza um mapeamento das características dos padrões de entrada sobre uma organização bidimensional de elementos sem a necessidade de um supervisor direto ou de referências externas adicionais. Em contraposição ao aprendizado supervisionado, no qual um *controlador externo* determina uma classe de *inputs* como padrão de entrada corrente, o SOFM detecta as similaridades entre os vetores de entradas e os vetores dos pesos das conexões, apoiando-se nas características dos padrões do meio ambiente e no *ajuste e correção* gradual dos padrões de conectividade da rede. Como o ajuste e correção dos pesos é realizado sem um *professor*, a rede realiza uma espécie de *aprendizado autônomo*.

Cabe ressaltar um segundo aspecto, relativo à natureza *local* e também *global* do processo de aprendizado na rede de Kohonen. A camada de Kohonen responde às regularidades do meio ambiente concentrando sua atividade em torno de elementos vizinhos, conferindo um caráter local à sua atividade. Simultaneamente, os elementos vizinhos desenvolvem interações coletivas cujo efeito global é a formação do mapa de características auto-organizadoras.

Um terceiro aspecto diz respeito à influência da natureza dos espaços de padrões de entrada e da aleatoriedade dos pesos iniciais das conexões. A rede de Kohonen é construída a partir da hipótese de que no espaço de entradas (conjunto de todos os padrões do meio ambiente apresentados à rede), os padrões compartilham de características comuns formando uma estrutura de relações entre eles (Beale & Jackson, 1993,



p.107) . O algoritmo de Kohonen procura descobrir os detalhes particulares específicos dos padrões que formam tal estrutura. Tattersall (1989, p.41) ressalta que esta estrutura focaliza o relacionamento *perceptual* entre os padrões. Para este autor, quando padrões relacionados perceptualmente são apresentados à rede, eles são capazes de disparar ativações em posições espaciais que refletem o relacionamento *perceptual* entre os padrões. Tal relacionamento é exemplificado por Tattersal como se segue : quando vetores de entradas, cujos elementos representam os sons de vogais, são apresentados à camada de Kohonen, diferentes elementos em posições espaciais distintas se tornam ativados. Assim, o relacionamento perceptual entres os diferentes sons das vogais reflete-se na posição relativa dos diferentes elementos ativados. A existência de relações entre os padrões poderia ser pensada como o ambiente "dirigindo" o processo de auto-organização na rede de Kohonen. Porém, ressaltamos que a estrutura de relações dos padrões de entrada não funciona como uma lei de construção, vinda do meio exterior, que guia o processo e sim na detecção de similaridades, na dinâmica das interações competitivas e nos ajustes de parâmetros internos ao processo.

Como vimos, no início do processo de aprendizagem, as conexões entre os elementos de entrada e os da camada de Kohonen possuem pesos atribuídos aleatoriamente. Isto resulta em uma ativação diferenciada dos elementos da camada de Kohonen, o que torna estes elementos mais distintos. À medida que o processo avança com a apresentação dos padrões, suas condições internas são ajustadas, e a

rede se estabiliza em padrões de conectividades que refletem as regularidades dos padrões de entrada. A dinâmica que leva o processo a evoluir de estados com níveis de ativação independentes para estados de ativação que guardam uma relação entre eles, assemelha-se um sistema que passa de um estado desorganizado para um estado organizado, de acordo com um dos princípios da auto-organização.

No quarto aspecto, ressaltamos que o motor do processo de formação dos mapas auto-organizadores da rede de Kohonen é a interação competitiva e colaborativa, sem supervisor, realizada pelos elementos da camada de Kohonen. No início do processo, os elementos da camada de Kohonen *competem* para se tornarem o elemento mais ajustado ao padrão apresentado. Na etapa seguinte, um conjunto de elementos em torno do elemento mais afinado colabora com interações excitatórias para que tal elemento se ative ainda mais, tornando-se vencedor do processo. As interações competitivas e colaborativas não ocorrem sob o controle de um supervisor, conferindo ao processo uma certa autonomia no ajuste dos pesos cuja representação bidimensional forma o mapa de características auto-organizadoras.

Na seção seguinte complementamos a nossa análise ilustrando a formação dos mapas de características auto-organizadoras, resultantes de diversas simulações do algoritmo de Kohonen.



Seção III – Ilustrações do modelo computacional da rede neural de Kohonen

Na presente seção analisamos o desenvolvimento dos mapas de características auto-organizadoras formados pela rede de Kohonen. A idéia principal é acompanhar, graficamente, a evolução desses mapas, obtidos por simulação computacional do algoritmo de aprendizado competitivo de Kohonen.

Na seção anterior, ressaltamos que o algoritmo de Kohonen realiza duas funções básicas: a) ele pesquisa, sequencialmente, em uma estrutura bidimensional, o elemento mais ativado (*vencedor*), utilizando-se de um critério de similaridade entre os vetores de entrada e os vetores pesos das conexões; b) ele ajusta os pesos das conexões, associados ao elemento vencedor e seus vizinhos, para deslocar, gradualmente, a orientação dos vetores pesos em direção ao vetor de entrada. A Figura 18 ilustra os parâmetros básicos envolvidos no algoritmo de Kohonen.

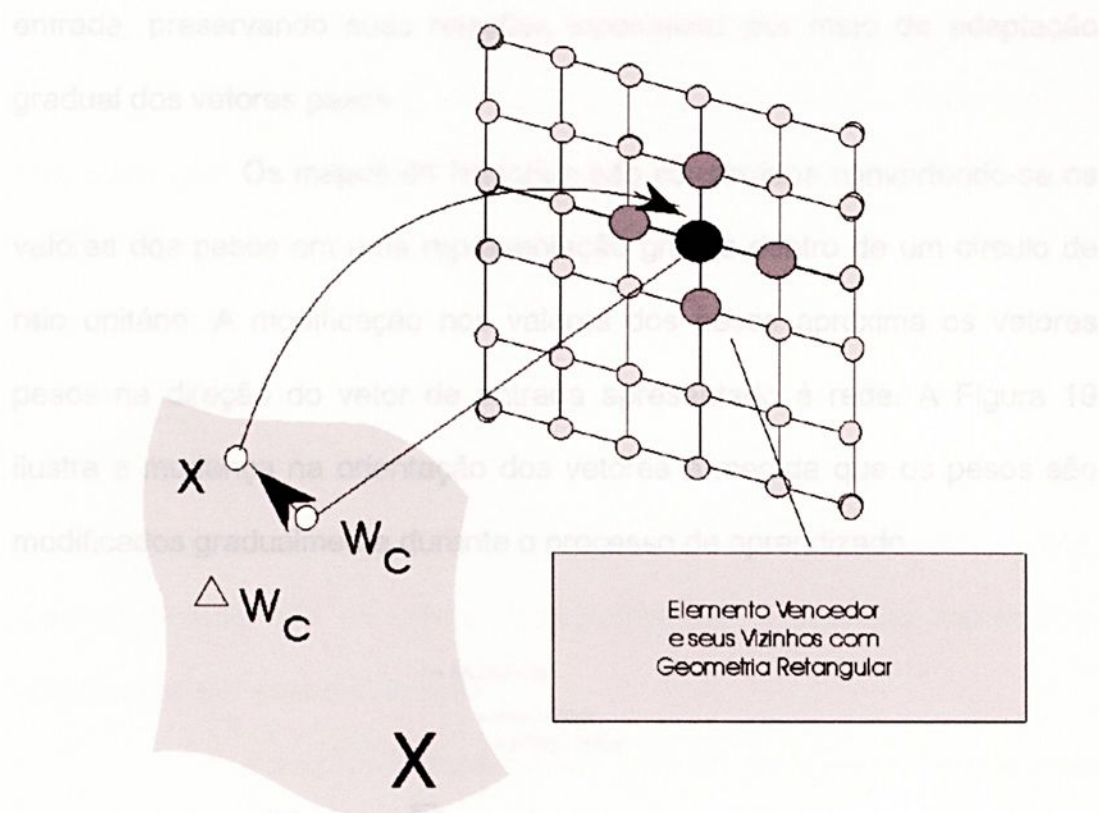


FIGURA 18 - Rede de Kohonen. Ritter et al. (1992, p.71)

De acordo com a figura acima, o algoritmo de treinamento procura reduzir a diferença Δw_c entre os vetores pesos w_c do elemento vencedor e seus vizinhos e o padrão de entrada x , de acordo com a regra de aprendizado apresentada na seção anterior, ou seja : $\Delta w_c = \alpha(x - w_i)$. O espaço de entradas X é formado por padrões cujas características são distribuídas estatisticamente por meio de uma função que estabelece a densidade de probabilidade. O algoritmo de Kohonen detecta as regularidades mais relevantes presentes no conjunto de padrões de

entrada, preservando suas relações topológicas por meio da adaptação gradual dos vetores pesos.

Os mapas de Kohonen são construídos convertendo-se os valores dos pesos em uma representação gráfica dentro de um círculo de raio unitário. A modificação nos valores dos pesos aproxima os vetores pesos na direção do vetor de entrada apresentado à rede. A Figura 19 ilustra a mudança na orientação dos vetores à medida que os pesos são modificados gradualmente durante o processo de aprendizado.

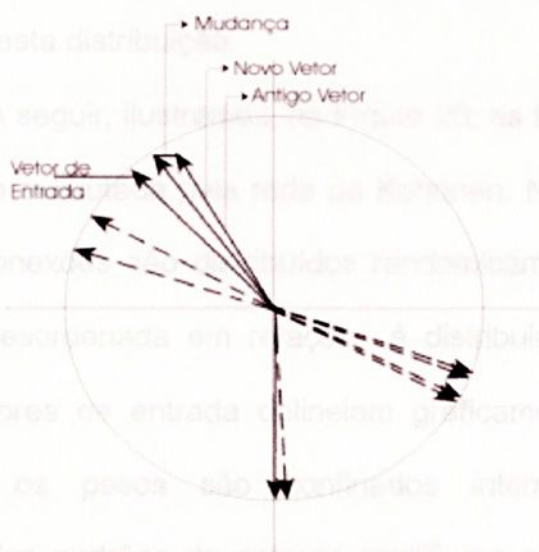


FIGURA 19 - A mudança do valor dos pesos das conexões rotaciona o vetor de pesos para a proximidade do vetor de entrada (Caudill & Butler, 1993, p.. 94)

De acordo com Caudill & Butler (1993, p.94), após a apresentação repetitiva de todos os padrões do espaço de entradas, estatisticamente distribuídos, todos os pesos representarão aproximadamente a mesma distribuição estatística dos padrões de entrada. Assim, Caudill & Buttler (1993, p.96) enfatizam que, se um determinado

percentual dos padrões de entrada aponta em uma direção específica, o mesmo percentual de vetores pesos apontará nesta direção. Estes autores ressaltam que a distribuição dos padrões de entrada é modelada pelos vetores pesos, mesmo *sem a rede ser informada externamente qual distribuição ela deverá reproduzir*. Neste sentido, eles enfatizam que a rede de Kohonen se *auto-organiza* modelando a distribuição dos padrões de entrada, apoiando-se apenas nas características destes padrões. A rede é capaz de reproduzir a distribuição estatística dos padrões de entrada, mesmo quando *não se conhece explicitamente* a equação matemática responsável por esta distribuição.

A seguir, ilustramos, na Figura 20, as fases do processo de auto-organização executada pela rede de Kohonen. No início do processo os pesos das conexões são distribuídos randomicamente, resultando em uma estrutura desordenada em relação à distribuição dos padrões de entrada. Os vetores de entrada delineiam graficamente uma estrutura retangular e os pesos são confinados internamente. Repetidas apresentações dos padrões de entrada modificam os vetores pesos das conexões do elemento vencedor e seus vizinhos, provocando um movimento desses vetores na direção do vetor de entrada. À medida que a quantidade de padrões apresentados à rede aumentam, o algoritmo de Kohonen realiza um *mapeamento ordenado* das características topológicas do espaço de entradas, aproximando os vetores pesos dos vetores de entradas.

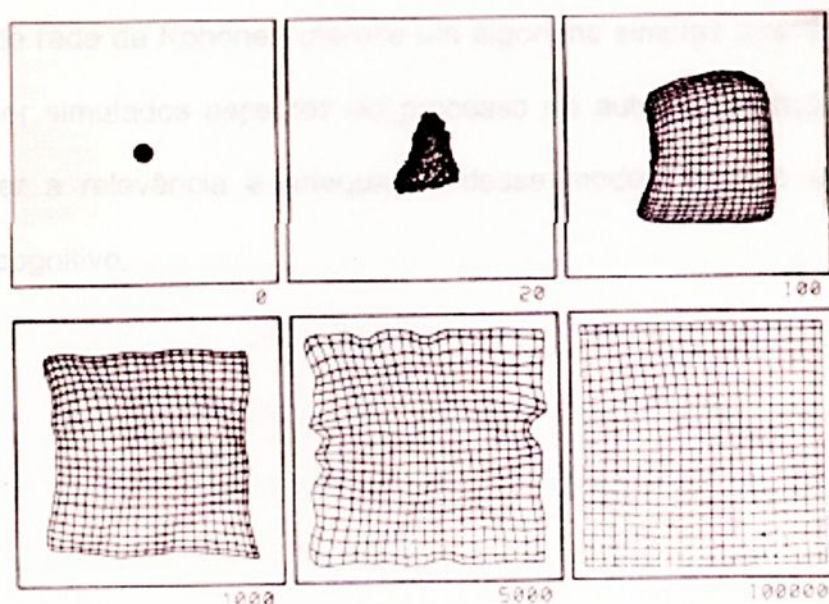


FIGURA 20 - Representação gráfica da evolução dos vetores pesos durante os ciclos de treinamento da rede de Kohonen. Kohonen (1998, p. 90)

A seqüência de ilustrações da Figura 20 mostra fases intermediárias do processo de auto-organização. O mapeamento das características se torna visível a partir de uma certa quantidade de padrões apresentados à rede. O ordenamento espacial dos vetores pesos mostram o aspecto dinâmico do processo de auto-organização no qual a rede passa por diferentes estados de atividade para se adaptar às regularidade dos elementos do meio ambiente, até atingir uma formação mais homogênea.

Kohonen (1997,p.90) destaca que a simulação computacional do seu algoritmo não visa reproduzir um mapeamento fiel das propriedades estatísticas dos padrões de entrada e, sim, detectar *automaticamente, domínios e dimensões* do conjunto formado por uma quantidade significativa desses padrões.

CONCLUSÃO Em resumo, procuramos mostrar neste capítulo que o modelo de rede de Kohonen oferece um algoritmo simples através do qual podem ser simulados aspectos do processo de auto-organização. Restamos saber a relevância e adequação desse modelo para o estudo do aparato cognitivo.

CONCLUSÃO

No decorrer deste trabalho procuramos relacionar os paradigmas da ciência cognitiva e da auto-organização. Este esforço foi empreendido com o objetivo de contribuir para o estudo da natureza do fenômeno cognitivo. Iniciamos caracterizando o campo da ciência cognitiva do ponto de vista histórico, seu domínio de atuação, seus principais pressupostos metodológicos e filosóficos, bem como alguns dos seus modelos cognitivos. Neste contexto, na segunda parte da dissertação, descrevemos os princípios gerais do processo de auto-organização, através do estudo do modelo cognitivo da rede neural artificial de Kohonen.

Assim, verificamos que a ciência cognitiva possui suas raízes no interior do movimento cibernético, herdando deste sua preocupação com o entendimento da natureza da mente inteligente, tanto em sistemas naturais quanto em sistemas artificiais, sistematizando seus conceitos e princípios apoiando-se na construção de modelos computacionais.

Vimos que a ciência cognitiva, ao unificar o estudo da cognição, se organiza em torno de três pressupostos básicos. Primeiro, assume-se, em geral, que o fenômeno cognitivo possui uma natureza representacional. Segundo, que a ciência cognitiva é o resultado da interação interdisciplinar entre as áreas de filosofia, psicologia, antropologia, inteligência artificial, lingüística, neurociências e a física.

A participação destas áreas no estudo da cognição parece variar em importância de acordo com as dificuldades que surgem durante a



tentativa de explicá-la. Se nos primeiros anos da ciência cognitiva a inteligência artificial possuía um papel fundamental, pois fornecia os elementos para explicar a mente, entendida como possuindo uma natureza computacional, o surgimento de teorias como a da auto-organização e das redes neurais artificiais enfatizam o valor da física através das noções de entropia e informação. O mesmo começa a ocorrer com a antropologia quando tentamos, atualmente, explicar as influências histórico-culturais no fenômeno cognitivo. Por último, ressaltamos que a ciência cognitiva faz do computador um instrumento fundamental, tanto para a simulação de alguns processos cognitivos quanto ao sugerir, mais radicalmente, que a mente funciona como um computador digital. Esta concepção da mente prometia, na década de 70, uma solução para o problema da relação mente-corpo.

Neste contexto, analisamos, resumidamente, o problema da relação mente-corpo e as principais propostas de solução formuladas pelas correntes dualistas e materialistas. Constatamos que os debates entre estas correntes fornecem importantes desafios para a ciência cognitiva na busca de uma explicação para tal questão. Verificamos que predomina ainda, no âmbito da ciência cognitiva, uma explicação funcionalista para o problema da relação mente-corpo, que dispensa a preocupação com a natureza material dos estados mentais e compromete-se apenas com seus aspectos funcionais. Neste sentido, a investigação promovida pela ciência cognitiva traduz a relação mente-corpo em termos da dicotomia *software-hardware*, na qual o *software* corresponde à mente abstrata e o *hardware* seria à sua instância material.



Indicamos, ainda, na primeira parte, que a ciência cognitiva se utiliza de dois modelos paradigmáticos para caracterizar os processos cognitivos : o modelo de inteligência artificial e o conexionismo ou redes neurais artificiais. No modelo da IA, a mente cognitiva é entendida como um sistema de manipulação de símbolos, através de regras dispostas de forma algorítmica, seguindo um procedimento serial, à semelhança do processamento de informação que ocorre em um computador digital. Críticos dos modelos da inteligência artificial supõem que nem todo fenômeno cognitivo é passível de uma descrição algorítmica, não sendo portanto passível de simulação computacional.

Em oposição ao processamento de informação serial, proposto pelo modelo de IA, analisamos os princípios gerais dos modelos conexionistas utilizados para o estudo dos processos cognitivos. Nestes modelos, ressaltamos várias propriedades interessantes, destacando-se: 1) a sua inspiração no cérebro humano, enfatizando-se a relevância da estrutura física na caracterização dos processos cognitivos; 2) o foco de análise dos processos cognitivos está centrado no nível microscópico ou subsimbólico e não na construção de programas para simular o comportamento inteligente; 3) o modelo interage diretamente com o meio ambiente, criando representações de padrões informacionais presentes no meio; 4) a atividade cognitiva seria uma propriedade emergente de uma ampla *interação* e *cooperação* entre um conjunto, densamente interconectado em rede, de unidades; 5) os padrões globais de atividade da rede resultam do *ajuste* dos *pesos das conexões* entre as unidades; 6)

no modelo conexionista ocorre um processamento simultâneo de informação, o que leva a um aumento de eficiência e de flexibilidade; 6) a rede aprende através do processo de ajuste dos pesos das conexões entre as unidades.

Conforme procuramos mostrar, o modelo conexionista é mais robusto, flexível e adaptável às variações dos padrões informacionais do meio ambiente, do que o modelo de IA, pois dispõe de mecanismos para correção de erros através da modificação de parâmetros internos (ou pesos das conexões) de maneira dinâmica, em resposta às mudanças no ambiente (plasticidade). Tal modelo é tolerante a falhas, pois sua arquitetura, constituída de um grande número de unidades, se mantém estável, mesmo quando ocorre uma perda de certas unidades que respondem às propriedades do padrão apresentado.

Uma vez que a propriedade de auto-organização é constantemente mencionada nos estudos conexionistas, passamos a analisar, na segunda parte da dissertação, as propriedades gerais da teoria da auto-organização para, no último capítulo, investigar as propriedades de auto-organização de um modelo conexionista específico. Percebemos que, assim como as hipóteses mecanicistas da ciência cognitiva, muitas das hipóteses da teoria da auto-organização estavam presentes no interior do movimento cibernético, especialmente na sua segunda fase, que focalizou o tema da auto-organização, conforme demonstram os estudos de Von Foerster e Ashby.

maneira formal. Ressaltamos, por exemplo que, a teoria da auto-organização procura compreender o *processo* de surgimento espontâneo de padrões, formas ou organizações a partir das interações entre elementos, sem a necessidade de um supervisor. O impulsionador deste processo é a dinâmica da interação entre seus elementos, que caracteriza a noção de *feedback* circular nos estudos cibernéticos. Indicamos também que, de acordo com Ashby, o processo de auto-organização é caracterizado pela formação de condicionalidades entre elementos que se encontram, sem uma determinação preestabelecida. Uma vez que as máquinas não satisfazem tal condição, ele argumenta em defesa da tese da impossibilidade lógica da ocorrência de auto-organização nas máquinas. Tal conclusão decorre do fato das máquinas terem seus comportamentos, segundo Ashby, ditados por funções preestabelecidas que não são auto-modificáveis sem a participação de eventos externos. *o ambiente sobre um*

conjunto Verificamos que a teoria da auto-organização possui um extenso domínio de atuação, que incorpora não somente os processos cognitivos mas, também, a natureza do surgimento espontâneo de certas organizações tanto nos sistemas biológicos e físicos quanto nos sistemas artificiais. Argumentamos que os modelos de redes neurais artificiais, caracterizados pela formação de padrões a partir da dinâmica da interação entre unidades interconectadas, podem constituir um ponto importante de ligação entre os estudos de auto-organização e da ciência cognitiva. A partir deste encontro, os processos cognitivos poderiam ser estudados como processos que se auto-organizam. Acreditamos que, ainda que de

maneira tênue, as hipóteses da teoria da auto-organização começam a permear os modelos cognitivistas, principalmente aqueles ligados à ciência cognitiva dinâmica, que faz amplo uso de redes neurais artificiais.

Com o intuito de verificar o emprego do conceito de auto-organização nas redes neurais artificiais, analisamos, no último capítulo, a rede de Kohonen já que tal modelo apresenta o esquema de aprendizado por auto-organização. Mostramos que na rede de Kohonen, padrões informacionais são formados a partir de uma ampla interação de elementos interconectados, onde se verifica a presença de propriedades dos processos de auto-organização, tais como: a dinâmica de cooperação e competição que se estabelece entre seus elementos, bem como a ausência de um elemento supervisor no ajuste de parâmetros internos.

Procuramos deixar claro que na rede Kohonen ocorre um mapeamento das características dos padrões do meio ambiente sobre um conjunto de elementos, espacialmente ordenados, sem o auxílio de um *supervisor externo*. Neste sentido, a rede se auto-organiza por não se utilizar de um padrão informacional externo, previamente definido, que serviria de referência durante a formação do padrão global da rede.

Durante a dinâmica de auto-organização da rede de Kohonen, os padrões internos de conectividade são modificados apoiando-se apenas na detecção de regularidades presentes nos padrões recebidos do exterior. O *ajuste* gradual dos padrões de conectividades (pesos das conexões) provoca a formação de mapas de características auto-organizadoras. Neste sentido, a rede produz a geração de uma “forma”

ou organização (mapa auto-organizado), simulando, de maneira simplificada, aspectos da auto-organização primária ressaltados por Debrun e Ashby.

O aprendizado por auto-organização, exibido pela rede de Kohonen, se inicia com a atribuição aleatória dos pesos das conexões produzindo, assim, um estado inicial estruturalmente não organizado. Com a apresentação repetitiva dos padrões do meio ambiente, ocorre uma absorção desta aleatoriedade e um aprimoramento da resposta (padrões são criados a cada instante) da rede. Julgamos que este processo pode ser interpretado como se a rede estivesse passando de um estado inicial, não organizado, para um estado final organizado sem a necessidade de um regulador externo dos seus parâmetros de funcionamento. Neste sentido, a rede de Kohonen apresenta também, de maneira simplificada, aspectos da auto-organização secundária descrita por Debrun. Além disso, se considerarmos os pesos iniciais, aleatoriamente distribuídos, como perturbações aleatórias, sendo que a rede ao invés de desorganizar-se passa a reconhecer padrões de maneira ordenada, então estaremos diante de um exemplo de formação de ordem a partir do ruído, tal como defendido por Von Foerster.

Por outro lado, a apresentação repetitiva dos padrões de entrada que possuem uma estrutura topológica, previamente definida, pode ser entendida como uma imposição do meio ambiente sobre o efeito global que a rede apresenta. Neste sentido, o processo de auto-organização ficaria comprometido. Pois, uma vez que, de certa forma, o meio impõe a

configuração do efeito global apresentado pela rede, o processo deixaria, neste momento, de ser auto-organizado.

Assim, ainda que a rede de Kohonen simule alguns aspectos do processo de auto-organização, uma vez que não há uma programação prévia preestabelecida nem um supervisor absoluto controlando os estágios do processo, ela não parece se auto-organizar na sua totalidade. Julgamos que dado o papel fundamental do meio ambiente nas representações da rede, o emprego do termo auto-organização, no plano macroscópico (do comportamento da rede) deverá ser utilizado com certa cautela.

Após as discussões realizadas acima poderíamos especular sobre a contribuição a ser dada pela teoria da auto-organização à ciência cognitiva no estudo da mente. De acordo com a teoria da auto-organização, a mente pode se manifestar na dinâmica da interação entre o organismo e seu meio ambiente. Parte desta dinâmica é caracterizada pela detecção de regularidades, ou informação, dinâmica esta que provoca um ajuste estrutural no sistema auto-organizado contribuindo para este se adaptar ao meio ambiente frente a presença de elementos novos ou desestruturadores.

Uma vez que habilidades de detecção de regularidades e ajuste poderiam estar presentes em um sofisticado termostato (e dado que o termostato não possui mente) que elementos adicionais seriam necessários para a caracterização de uma mente? Uma resposta possível

para esta questão seria o elemento da *consciência* que deveria ser adicionado à tais propriedades dos sistemas auto-organizados.

especulação q Mas o que é uma consciência² ? Esta pergunta exigiria um novo trabalho (que começa a se esboçar palidamente no nosso horizonte). Uma permitida especulação provisória, e certamente prematura, no contexto da teoria da auto-organização, poderia ser feita em termos da noção de *autonomia*. Uma vez estabelecido um sistema através do processo de auto-organização, nos moldes sugeridos acima, nos deparamos-nos com sistemas do tipo redes neurais artificiais, cuja organização não se distingue, em geral, do meio, pois, como vimos, a rede no final do processo acaba por se moldar ao meio. Se uma certa autonomia fosse estabelecida neste sistema, criando-se, nas palavras de Debrun “um dentro e um fora” e, se, além disso, um processo perceptivo fosse desenvolvido, de tal forma que houvesse o reconhecimento da alteridade, então possivelmente teríamos aí um sistema consciente. Estamos sugerindo que consciência parece estar presente em sistemas que preservam uma certa autonomia em relação ao meio ambiente e que, ao mesmo tempo, possuem a habilidade de reconhecer a presença do “outro” como diferente de si mesmos. Neste sentido primitivos de consciência, os homens não seriam os seus únicos portadores; mesmo animais muito simples poderiam ser considerados conscientes. As máquinas, entretanto, tal como se configuram atualmente, necessitariam evoluir muito para atingir tal estágio de desenvolvimento. Os passos para

²Agradecemos ao Prof. Piqueira pela instigante questão que nos fez refletir sobre esse tema durante o nosso exame de qualificação



se adquirir autonomia e reconhecimento da alteridade parecem estar ainda bastante distantes do nosso entendimento. Esta é, contudo, uma especulação que fica para um futuro desenvolvimento.

ARMSTRONG, D. A. L., MALCOLM, N. *Consciousness & Causality*. Oxford: Blackwell Publishers, 1994. 222p.

ARRIB, M. A. *Brains, machines, and mathematics*. 2 ed. New York: Springer-Verlag, 1997. 762p.

ASHBY, W. R. *Introdução à cibernética*. São Paulo: Ed. Perspectiva, 1970. 345p.

ASHBY, W. R. Principles of the self-organizing system, in : VON FOERSTER, H. ZOPF, G. W. (Orgs.) *Self-organizing systems*. Londres: Pergamon, 1962. p. 255-78.

ATLAN, H. *Entre o cristal e a função: ensaio sobre a organização do ser vivo*. Rio de Janeiro: Jorge Zahar, 1992. 268p.

BEALE, R. JACKSON, T. *Neural computing: an introduction*. Bristol: Institute of Physics, 1994. 240p.

REFERÊNCIAS BIBLIOGRÁFICAS

ALEKSANDER, I. (ed.) *Neural computing architectures: the design of brain-like-machines*. Oxford : North Oxford Academic, 1989. 401p.

ARMSTRONG, D. A ., MALCOLM, N. *Consciousness & Causality*. Oxford :Blackwell Publisher, 1984. 222p.

ARBIB, M. A. *Brains, machines, and mathematics*. 2.ed. New York: Springer- Verlag, 1987. 202p.

ASHBY, W. R. *Introdução à cibernética*. São Paulo: Ed. Perspectiva, 1970. 345p.

ASHBY, W. R. Principles of the self-organizing system, In : VON FOERSTER, H. ZOPF, G. W. (Orgs.) *Self-organizing systems*. Londres: Pergamon, 1962. p. 255-78.

ATLAN, H. *Entre o cristal e a fumaça: ensaio sobre a organização do ser vivo*. Rio de Janeiro: Jorge Zahar, 1992. 268p.

BEALE, R. JACKSON, T. *Neural computing: an introduction*. Bristol: Institute of Physics, 1994. 240p.



BECHTEL, W. *Philosophy of mind : an overview for cognitive science*.

Hillsdale: Lawrence Erlbaum, 1988. 161p.

BENDER, E. A. *Mathematical methods in artificial intelligence*.

Washington: IEEE Computer Society, 1996. 635p.

BLACK, M. *Modelos y metáforas*. Madrid: Editorial tecnos, 1961. 257p.

BLOCK, N. Introduction: what is functionalism ? In: _____. *Reading in the philosophy of psychology*, v. 1. Cambridge : Harvard University Press, 1980, p. 171-84.

BLOCK, N. Troubles of Functionalism. In: _____. *Reading in the philosophy of psychology*, Cambridge : Harvard University Press, 1980. v.1. p. 268-305.

BUTTON, G., COULTER, J., LEE, J. R. R., SHARROCK, W.

Computadores, mentes e conduta. São Paulo: Unesp, 1998. 304p.

CAUDILL, M., BUTLER, C. *Understanding neural networks: computer explorations*, Cambridge: Mit Press, 1992. 2v.

CAUDILL, M., BUTLER, C. *Naturally intelligent systems*, 3 rd.ed.

Cambridge: Mit Press, 1993. 2v.

CHARNIAK, E. , MCDERMOTT, D. *Introduction to artificial intelligence.*

Addison Wesley: Massachusetts, 1987.

CHURCHLAND, P. M. *Matter and consciousness: a contemporary introduction to the philosophy of mind.* 3 rd ed. Cambridge: Mit Press, 1988. 183p.

CHURCHLAND, P. M. *A neurocomputational perspective: the nature of mind and the structure of science.* Cambridge: MIT Press, 3.ed. 1993. 321p.

DASCAL, M. *Hobbes challenge.* Tel Aviv: Tel Aviv University, 1998. 33p.

DAWSON, R. W., SHAMANSKI, K. S. Connectionism, confusion, and cognitive science [online]. http://web.psych.ualberta.ca/~mike/Pearl_Street/Papers/Confuse/confuse.htm [02 mar 1997].

DEBRUN, M., GONZALES, M. E. Q., PESSOA JUNIOR., O. (Orgs.) *Auto-organização: estudos interdisciplinares em filosofia, ciências naturais e humanas, e artes.* Campinas: UNICAMP, 1996. 455p. (Cle,18).

DEBRUN, M. M. Auto-organização e ciências cognitivas. In: ENCONTRO COM AS CIÊNCIAS COGNITIVAS, 1., 1995, Marília. Anais... Marília: Faculdade de Filosofia e Ciências, 1995. p. 29-38.

DREYFUS, H. L. *O que os computadores não podem fazer: crítica da razão artificial*. Rio de Janeiro: Casa do Livro Eldorado, 1975. 314p.

DESCARTES. São Paulo: Nova Cultural, 1996. 296p (pensadores).

DOSHER, B. Models of visual search: finding a face in the crowd. In: SCARBOROUGH, D., STERBERG, S. *Methods, models, and conceptual issues*. Cambridge: Mit Press, 1998. v.4.

DUPUY, J. P. *Nas origens das ciências cognitivas*. São Paulo :Ed. Unesp, 1996. 228p.

DUTRA, V. L. O dualismo mente-corpo : implicações para a prática da atividade física. Rio Claro, 1996. 194p. Dissertação(Mestrado em Ciências da Motricidade) – Intituto de Biociências, Universidade Estadual Paulista.

ECKARDT, B. V. *What is cognitive science*. Cambridge: Mit Press, 1995. 466p.

FODOR, J. A. PYLYSHYN, Z. N. *Connectionism and cognitive architecture: a critical analysis*. In: PINKER, S., MEHLER, J. *Connections and symbols*. Amsterdam : Elsevier Science, 1989. 255p.

GARDNER, H. *A nova ciência da mente: uma história da revolução cognitiva*, São Paulo : Editora da Universidade de São Paulo, 1995. 454p.

GONZALES, M. E. Q., BROENS, M. C. (Orgs.) *Encontro com as ciências cognitivas*. Marília: Unesp-Marília-Publicações, 1998. v.2, 280p.

GONZALES, M. E. Q. *Redes neurais e representação mental: um ensaio sobre harmonia e racionalidade*. *Transformação*, v.14, p. 93-108, 1991.

GONZALES, M. E. O. ALLEN, A . G. *Self-organization and the mind-brain problem*. *Cienc. Cult. J. Braz. Assoc. Adv. Sci.* v.50, n. 2/3, p.90-4, 1998.

GONZALES, M. E. Q. *Ação, causalidade e ruído nas redes neurais auto-organizadas*. In: DEBRUN, M. GONZALES, M. E. Q., PESSOA JUNIOR, O. *Auto-organização: estudos interdisciplinares*. Campinas: UNICAMP, 1996. p.273-97. (Cle,18)

GUYTON, A. C. *Fisiologia humana e mecanismos das doenças*. 4.ed. Rio de Janeiro: Guanabara Koogan, 1989. 572p.

HEBB, D. O. *The first of perception: growth of the assembly*. In: ANDERSON, J. A ., ROSENFELD, E. *Neurocomputing: foundations of research*. 15. ed. Cambridge: MIT Press, 1990. p. 45-56.



JANTSCH, E. *The Self-Organizing Universe: scientific and human implications of the emerging paradigm of evolution*. Oxford: Pergamon Press, Oxford: Pergamon, 1980. 343p.

KOHONEN, T. *Self-organization and associative memory*. 2.ed. Berlin: Springer-Verlag, 1988. 312p.

KOHONEN, T. *Self-organization maps*. 2.ed. Berlin: Springer-Verlag, 1997. 426p.

KOHONEN, T. The self-organization map. *Proc. IEEE*, v.78, n.9, p. 1464-80, set. 1990.(a)

KOHONEN, T. Self-organized formation of topologically correct feature maps. In: ANDERSON, J. A ., ROSENFELD, E. *Neurocomputing: foundations of research*. 15. ed. Cambridge: MIT Press, 1990. p. 509-21.(b)

KOVÁCS, Z. L. *O cérebro e sua mente: uma introdução à neurociência computacional*. São Paulo: Edição Acadêmica, 1997. 214p.

KROHN, W., KUPPERS, G., NOWOTNY, H. *Selforganization: portrait of a scientific revolution*. Dordrecht : Kluwer Academic Publishers, 1990. 270p.

LAIRD, P. J. *The computer and the mind: an introduction to cognitive science*. FontaPress, 1993. 450p.

MATURANA, H. M., VARELA, F. J. *Autopoiesis and cognition: the realization of the living*. Boston: Reidel, 1980, 141p.

MCCULLOCH, W. S., PITTS W. A Logical Calculus Of The Ideas Immanent In Nervous Activity . Bull. Math. Bioph. v . 5, p.115-33, 1943.

MINSKY L. M. Inteligência Artificial. In: FENICHEL, R. R., JOSEPH, W. *Computadores e computação*. São Paulo: Perspectiva, 1977.

MINSKY L. M., PAPERT A. S. *Perceptrons : an introduction to computational geometry*, Cambridge: Massachusetts Institute of Technology, 1988.

292p. PYL, W. Z. *Computation and cognition : toward a foundation for cognitive science*. Cambridge : MIT Press, 1980. 251p.

MORIN, E., *Ciência com consciência*. Rio de Janeiro: Bertrand Brasil, 1996. 336p.

ROGERS, J. *Principles of cybernetics*. New York : Sparger Books, 1967.

NAGEL, T. What is like to be a bat. In : BLOCK, N. *Reading in the philosophy of psychology*. Cambridge: Harvard University Press, 1980. v.1, p.159-68.

POSNER, M. I. *Foundations of cognitive science*. Cambridge: MIT Press, 1989.

PRIGOGINE, I. *O fim das certezas: tempo, caos e as leis da natureza*. São Paulo: Ed. UNESP, 1996. 199p.

PUTNAM, H. Why of Functionalism didn't work. In: _____. *Words & life*. 2.ed. Cambridge: Harvard University Press, 1995. p. 441-59.

PUTNAM, H. *Philosophy and our mental life*. In: BLOCK, N. (Ed.) *Reading in the philosophy of psychology*. Cambridge: Harvard University Press, 1980. v.1, p. 134-43.

PUTNAM, H. *Renewing philosophy*. Harvard University Press, 1992. 234p.

PYLYSHYN, W. Z. *Computation and cognition : toward a foundation for cognitive science*. Cambridge : MIT Press, 1986. 291p.

ROSENBLATT, F. *Principles of neurodynamics*. New York: Spartan Books, 1962.



RUMELHART, D. E. The architecture oh mind: a connectionist approach. In: *Foundations of cognitive science*. Cambridge: MIT Press, 1989.

RUMELHART, D. E., MCCLELLAND, J. L. *Parallel distributed processing: explorations in the microstructure of cognition*. 11. ed. Cambridge: MIT Press, 1995. v.1.: Foundations, 457p.

RITTER, H. MARTINETZ, T, SCHULTEN, K. *Neural computation and self-organizing maps*. New York: Addison-Wesley, 1992. 306p.

RYLE, G. *The concept of mind*. Harmondsworth: Penguin Books. 1970. (Ed.) *Neural computing architectures*. London: North Oxford Academic, 1989.

SEARLE, J. *A redescoberta da mente*. São Paulo: Martins fontes. 1997. 379p.

SHANNON, C. E., WEAVER, W. The mathematical theory of communication. Urbana: University of Illinois Press, 1971. 125p. http://web.psych.ucla.edu/~mike/Psych_Street/INTD554/Week1/control.html

SIMON, H. A, KAPLAN, C. A. The architecture oh mind: a connectionist approach. In: *Foundations of cognitive science*. Cambridge: MIT Press, 1989.

TL. 1989. A. M. On Computable numbers, with an application to the entscheidungsproblem. *Proc. Lond. Math. Soc.*, v. 42, p.230-35, 1936

SLOMAN, A. *The computer revolution philosophy: philosophy, science and models of mind*. Sussex: Harvester Press, 1978. 304p.

STICHWEH, R. Self-organization and autopoiesis in the development of modern science. In: KROHN, W., KUPPERS, G., NOWONTNY, H. *Selforganization: portrait of a scientific revolution*. Dordrecht: Kluwer Academic, 1990, p.195-207.

STILLINGS, A. N. *Cognitive science: an introduction*. Cambridge : MIT Press, 1987. 521p.

TATTERSALL, G. Neural map applications. In: ALEKSANDER, I. (Ed.) *Neural computing architectures*. London: North Oxford Academic, 1989. p.41-73.

THAGARD, P. *Mente: introdução à ciência cognitiva*. Porto Alegre: Artes Médicas, 1998. 209p.

THE COFFEE room and cognitive science. [online]. http://web.psych.ualberta.ca/~mike/Pearl_Street/INTD554/Week1/control.html [10 mar 1997].

TURING, A. M. On Computable numbers, with na application to the entscheidungsproblem. *Proc. Lond. Math. Soc.*, v .42, p.230-35, 1936.

- TZANAKOU, E. M. Nervous system. In: BRONZINO, J. D. *The biomedical engineering handbook*. Connecticut: CRC Handbook Published, 1995. p. 22-32.
- VARELA, F. J., THOMPSON, E., ROSCH, E. *The embodied mind: cognitive science and human experience*. 3.ed. Cambridge: MIT Press, 1993. 307p.
- VON FOERSTER, H. On self-organizing systems and their environments. In: YOVITS, M. C., CAMERON, S. (Orgs.) *Self-organizing systems*. Oxford: Pergamon, 1960. p. 31-50.
- WASSERMAN, P.D., *Neural computing: theory and practice*. New York: Van Nostrand Reinhold, 1989. 230p.
- WINOGRAD, T. *Understanding computers and cognition*. New Jersey : Ablex Publishing Corporation, 1986. 207p.
- YATES, F. E. *Self-organizing systems: the emergence of order*. New York: Plenum Press,
- YOVITS, M. C. *Self-organizing systems*. Washington : McGregor & Werner, 1962, 563p.

OLIVEIRA, M. Q. *Ciência cognitiva e auto-organização: um encontro no modelo conexionista da mente*. Marília, 1999. 158p. Dissertação (Mestrado em Filosofia – linha de pesquisa : ciência cognitiva e filosofia da mente) – Faculdade de Ciências e Letras, Campus de Marília, Universidade Estadual Paulista “Júlio de Mesquita Filho”.

RESUMO

Nesta dissertação analisamos um possível encontro de interesses explicativos da atividade cognitiva entre dois paradigmas científicos: a ciência cognitiva e a teoria da auto-organização. Oriundos dos estudos da cibernética, esses paradigmas fornecem uma estrutura conceitual relevante para compreender a natureza dos processos cognitivos. Inicialmente, caracterizamos o campo da ciência cognitiva, ressaltando seu domínio de atuação, seu histórico, seus principais pressupostos metodológicos e os modelos computacionais por ela utilizados para estudar a natureza da atividade cognitiva. Dominada em seus primeiros anos pelo modelo de Inteligência Artificial, que estuda o comportamento inteligente apoiando-se na manipulação de símbolos através de regras abstratas, esta ciência mais recentemente, retoma o modelo das Redes Neurais Artificiais como instrumento de análise do pensamento inteligente. Tal modelo postula que a cognição é o efeito da dinâmica da interação entre um conjunto de elementos densamente interconectados. No presente trabalho, analisamos o modelo de Rede Neural de Kohonen, utilizado para emular propriedades de auto-organização. A rede de Kohonen se caracteriza pela sua plausibilidade biológica e, principalmente, por não exigir a presença de um elemento supervisor para controlar seu processo de aprendizado. Com base em um processo de competição entre elementos de uma determinada vizinhança, o modelo de Kohonen consegue detectar regularidades presentes nos padrões informacionais do meio ambiente e mapeá-las sobre um conjunto de elementos bidimensionais, formando os chamados *Self-Organizing Maps*. Finalizamos nosso trabalho destacando os principais aspectos de auto-organização exibidos pelo modelo de Kohonen, bem como seus limites perante as principais hipóteses do paradigma da auto-organização.

Palavras-chave: auto-organização; ciência cognitiva; inteligência artificial; redes neurais artificiais.

Keywords: artificial intelligence; cognitive science; artificial neural networks; self-organization.

OLIVEIRA, M. Q. *Cognitive science and self-organization: an convergence in connectionist model of mind*. Marília, 1999. 158p. Dissertação (Mestrado em Filosofia – linha de pesquisa: ciência cognitiva e filosofia da mente) – Faculdade de Ciências e Letras, Campus de Marília, Universidade Estadual Paulista “Júlio de Mesquita Filho”.

ABSTRACT

In this dissertation we analyze a possible convergence of explanatory interests of the cognitive activity between two scientific paradigms: cognitive sciences and the theory of self-organization. Arising from studies in cybernetics, these paradigms furnish a conceptual structure that is relevant to the understanding of the nature of cognitive processes. First, we characterize the sphere of cognitive sciences, emphasizing its field of action, its background, its main methodological presuppositions, and the computing models used to study the nature of the cognitive activity. Dominated during its first years by the model of the Artificial Intelligence, which proposes to study intelligent behavior by relying on the manipulation of symbols through abstract rules, more recently, cognitive scientists have retrieved the Artificial Neural Nets model, or "Connectionism", as a tool to better analyze intelligent thought. This model claims that cognition is the effect of the dynamic interaction between a set of elements densely interconnected. In this present work, we analyze Kohonen Neural Net model, used to emulate properties of self-organization. Kohonen Net is characterized by its biological plausibility and, mainly, by the fact that it does not demand a supervising element to control its learning process. Based on a process of competition between elements of a given neighborhood, Kohonen model is able to detect regularities present in informational patterns of the environment and map them over a set of bidimensional elements, forming the so-called *Self-Organizing Maps*. We conclude our work by investigations the main aspects of self-organization shown by the Kohonen model, as well as by pointing out its limits in relation to the main hypothesis of self-organization paradigm.

Keywords: artificial intelligence; cognitive science; artificial neural networks; self-organization.

