Ronaldo José Moraca

A CIÊNCIA COGNITIVA E SUAS DUAS GRANDES VERTENTES:

Considerações filosóficas sobre a inteligência artificial e o conexionismo

MARÍLIA
1999

unesp

MFN
12822

1410001003

RONALDO JOSÉ MORACA

A CIÊNCIA COGNITIVA E SUAS DUAS GRANDES VERTENTES:

CONSIDERAÇÕES FILOSÓFICAS SOBRE A INTELIGÊNCIA ARTIFICIAL E O CONEXIONISMO

T. 1.003

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

Orientadora: Profª. Drª. Maria Eunice Q. Gonzales

Fac. de Filosofia e Ciências
BIBLIOTECA
Campus de Marília

Marília
1999

unesp

Nº cham 153.4
M 827c

Nº Tombo 1.003
Aquisição D
Preço R$ 5,00
Data 17/08/00
Proced. Autor

A474m Moraca, Ronaldo José.
A Ciência Cognitiva e Suas Duas Grandes Vertentes: considerações filosóficas sobre a inteligência artificial e o conexionismo. / Ronaldo José Moraca. – 1999.
133 p.; 30 cm.

Dissertação (mestrado) –Universidade Estadual Paulista. Faculdade de Filosofia e Ciências de Marília, 1999.

1. Filosofia da Mente. 2. Ciência Cognitiva. 3. Inteligência Artificial. 4. Conexionismo. 5. Ciência Cognitiva Dinâmica I. Título.

CDD 20 – 153.4

RONALDO JOSÉ MORACA

A CIÊNCIA COGNITIVA E SUAS DUAS GRANDES VERTENTES:

CONSIDERAÇÕES FILOSÓFICAS SOBRE A INTELIGÊNCIA ARTIFICIAL E O CONEXIONISMO

COMISSÃO JULGADORA

DISSERTAÇÃO PARA OBTENÇÃO DO GRAU DE MESTRE

Presidente e Orientadora: Prof. Drª. Maria Eunice Quilici Gonzales,
Departamento de Filosofia – FFC – UNESP.

2º Examinador: Profa. Drª. Adriana Benevides Soares,
Centro de Ciências do Homem. – UENF.

3º Examinador: Profa. Drª. Mariana Cláudia Broens,
Departamento de Filosofia – FFC – UNESP.

4º Examinador (Suplente): Prof., Dr. Lauro Frederico B. da Silveira,
Departamento de Filosofia – FFC – UNESP.

5º Examinador (Suplente): Profa. Drª. Paula Mousinho Martins,
Centro de Ciências do Homem. – UENF.



AGRADECIMENTOS

Sei que estas poucas linhas não conseguiram expressar, completamente, a imensa dívida que tenho com todas as pessoas (e instituição) que me ajudaram a realizar esta dissertação. Ciente desta limitação, quero apenas tornar público meus agradecimentos a todos que me ajudaram na realização deste trabalho.

Quero, primeiramente, agradecer a Profª. Drª Maria Eunice Quilici Gonzales, que aceitou o desafio de transformar aquela simples folha escrita a mão, apresentada em 21 de Abril de 1993, nesta dissertação. Desde que esse processo começou, ela tem sido generosa com suas idéias, críticas e inesgotável paciência para ler e reler as diversas versões deste trabalho.

Um agradecimento, todo especial, a minha banca examinadora composta pela Profª. Drª Mariana Claudia Broens – do departamento de Filosofia da UNESP de Marília – e pela Profª. Drª Adriana Benevides Soares – do Centro de Ciência Homem da UENF, que colaboraram decisivamente, com suas sugestões e críticas no exame da qualificação, na redação final da dissertação. Quero, também, agradecer aos professores: Dr. Lauro Frederico Barbosa da Silveira e Drª Paula Mousinho Martins que aceitaram participar da banca examinadora.

Minha imensa dívida com a Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP – que concedeu uma bolsa para me dedicar exclusivamente a este trabalho. Sem tal auxilio dificilmente conseguiria terminar, com sucesso, esta dissertação. Sou grato, também, ao parecerista desta entidade que, com sua leitura atenciosa, contribui decisivamente no rumo que tomou meu trabalho.



Agradeço a todos os meus professores da graduação e pós-graduação. Em especial, um muito obrigado aos professores Jézio H. Gutierre e Drª Carmen B. Milidoni que, com seus ensinamentos sobre Filosofia da Ciência e Teoria das Ciências Humanas, contribuíram para que minhas idéias tomassem forma.

Obrigado, também, aos meus amigos da graduação e pós-graduação, simbolizados pelos amigos (as): Ademar S. Júnior, bem como toda sua família, pelo caloroso acolhimento no inicio deste trabalho e Patrícia, Eloisa, (e suas "cachorras", bianca e duda) por me suportarem até agora (acolhimento parte II).

Meus agradecimentos finais a meus irmãos, Roberto, Reinaldo, Rêne, Rosemari, Rosemeire, Rogério e Renato, precisam ser de um tipo diferente. Todos contribuíram, através das mais variadas maneiras, com ingredientes para meu trabalho. Porém, ao meu ver, a mais valiosa contribuição foi que insistiram para que minha vontade de estudar fosse levada a diante. Qualquer um que queira realizar o que eu realizei saberá o quanto é fundamental tal estímulo. Honestamente faltam palavras para agradecer-lhes.

Para Antônio Moraca (*em memória*), Vera J. Moraca e Vilma A. Bianchi, pelas valiosas lições sobre a vida.

ÍNDICE

RESUMO..... 01

INTRODUÇÃO GERAL..... 02

Primeira Parte

AS BASES HISTÓRICO-FILOSÓFICAS DA CIÊNCIA COGNITIVA

Capítulo 1 O Surgimento da Ciência Cognitiva: pressupostos, formação e seu forte vínculo com a cibernética

INTRODUÇÃO..... 08

1.1 Dos pressupostos básicos da ciência cognitiva: interdisciplinaridade e o emprego do computador..... 09

1.2 A herança cibernética..... 19

1.2.1 Wiener e o nascimento da cibernética..... 20

1.2.2 John Von Neumann: o paralelo entre o computador e o cérebro..... 22

1.2.3 As idéias de Ashby sobre auto-organização..... 24

1.3 A proposta de Von Eckardt e as semelhanças com o paradigma de Kuhn 27

1.3.1 A ciência cognitiva dinâmica e o paradigma da complexidade..... 35

Capítulo 2 Teria a Ciência Cognitiva, em Sua Vertente Denominada IA, Raízes na Filosofia Cartesiana?

INTRODUÇÃO..... 41

2.1 A inteligência artificial e o dualismo mente/corpo..... 42

2.2 A atualidade da filosofia cartesiana..... 44

2.2.1 O espírito como elemento central da filosofia cartesiana..... 46

2.3 Críticas ao dualismo cartesiano..... 50

2.3.1 As hipóteses de De La Mettrie..... 52

2.3.2 Ryle e a análise dos estados mentais a partir dos estados disposicionais..... 56

2.3.3 O naturalismo biológico de Searle..... 62

Segunda Parte

O ALCANCE E OS LIMITES DO MECANICISMO NA EXPLICAÇÃO DOS PROCESSOS MENTAIS

Capítulo 3 Perspectivas Mecanicistas do Pensamento: da máquina de Turing à inteligência artificial

INTRODUÇÃO..... 69

3.1 A descrição da máquina de Turing..... 70

3.2 Os modelos da inteligência artificial..... 81

3.3 Alcance e limites dos programas da inteligência artificial..... 90

3.3.1 O problema semântico..... 90

3.3.2 Os abusos filosóficos da linguagem ordinária..... 94

Capítulo 4 Modelos de redes neurais

INTRODUÇÃO..... 99

4.1 Os neurônios formais de McCulloch e Pitts..... 100

4.2 Descrição das redes neurais artificiais..... 103

4.3 Modelos de redes neurais artificiais..... 109

4.3.1 A rede de Perceptron..... 109

4.3.2 O modelo de Kohonen..... 113

4.4 Críticas aos modelos conexionistas..... 116

4.4.1 O papel do corpo humano no reconhecimento da padrões..... 116

4.4.2 O problema do reconhecimento da padrões, a partir do treino, em redes neurais artificiais..... 120

CONSIDERAÇÕES FINAIS..... 123

REFERÊNCIAS BIBLIOGRÁFICAS..... 129

RESUMO

O objetivo desta dissertação é tecer considerações, a partir da perspectiva da filosofia da mente, sobre as duas principais linhas de pesquisa da ciência cognitiva: a inteligência artificial e o conexionismo. A ciência cognitiva, enraizada na mentalidade científica e tecnológica dos nossos dias, propõe modelos explicativos dos processos cognitivos guiando-se pela perspectiva neo-mecanicista. Essa perspectiva situa-se entre o simplismo dos mecanismos artificiais e a complexidade dos seres vivos. Investigamos, inicialmente, as bases histórico-filosóficas da ciência cognitiva a fim de caracterizar com precisão os pressupostos básicos, o domínio e as diferenças entre as linhas de pesquisas existentes nesta nova ciência. Constatamos que dois pressupostos são básicos nesse domínio científico, a saber, a interdisciplinaridade e o emprego de máquinas na elaboração de modelos explicativos da atividade cognitiva. Além destes dois pressupostos, assinalamos que no interior desta nova ciência existem dois programas centrais de pesquisa. Um primeiro que pode ser chamado de *'representacionista'* pois, na explicação da atividade mental, a noção de representação é fundamental. Tal programa abriga a inteligência artificial e conexionismo (linhas de pesquisas tratadas nesta dissertação). O segundo programa existente nesta ciência recebe o nome de *'ciência cognitiva dinâmica'* ou *'não representacionista'* e adota como hipótese central de trabalho a rejeição da noção de representação mental na explicação da cognição (este programa de pesquisa será brevemente mencionado nesta dissertação). Em seguida, tecemos algumas considerações à respeito dos modelos da inteligência artificial e do conexionismo. Tais considerações procuram apontar, fundamentalmente, o alcance e os limites destes modelos. Concluímos esta dissertação extraindo o que podemos aprender sobre a natureza dos processos cognitivos a partir da ciência cognitiva.

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

Parte I

AS BASES HISTÓRICO-FILOSÓFICAS DA CIÊNCIA COGNITIVA

CAPÍTULO 1

O SURGIMENTO DA CIÊNCIA COGNITIVA: pressupostos, formação e seu forte vínculo com a cibernética

INTRODUÇÃO GERAL

A presente dissertação situa-se no âmbito da filosofia da mente e tem como objetivo central tecer considerações filosóficas sobre a ciência cognitiva. O interesse em investigar esta nova ciência foi despertado por alguns objetivos que nela estão presentes, a saber, procurar esclarecer como o conhecimento é possível, de que maneira ele se estrutura em nossa mente, qual a relação da mente com o corpo, qual o papel das representações mentais na aquisição do conhecimento, entre outros. Sendo assim, podemos dizer que nosso trabalho investiga este novo empreendimento científico a partir de uma perspectiva externa ao projeto cognitivista.

Foram os filósofos os primeiros a refletirem sobre a natureza e possibilidade do conhecimento. Pensadores como Platão, Descartes, Kant, entre outros, deram-se conta da importância de compreender com profundidade aquilo que julgavam ser uma singular propriedade da mente humana capaz de modificar o mundo.

Grandes dificuldades estão reservadas aqueles que, como Platão, Descartes e Kant, se aventuraram no complexo problema da natureza do conhecimento. De forma resumida, podemos dizer que duas tendências predominaram quando se procurou responder este problema.

A primeira tendência pode ser chamada de *mentalista* pois, na investigação sobre a natureza do conhecimento e dos processos de aprendizagem, entre outros processos cognitivos, empregam-se conceitos como os de mente, representação mental, processos mentais. Como exemplo desta tendência podemos mencionar o monismo anômalo. Esta tendência combina um monismo

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

ontológico com uma espécie de dualismo de propriedades. Ontologicamente falando só existe uma espécie de substância, a saber, a física. Porém, quando descrevemos nossos estados mentais empregamos as noções supracitadas. Esta forma própria de descrição dos eventos mentais nos leva a concluir que, segundo esta concepção, os conceitos mentais são irredutíveis aos conceitos físicos.

A segunda tendência pode ser denominada *materialista* pois, algumas de suas correntes dispensam, na explicação do conhecimento, referência a termos mentais. Como exemplo desta tendência podemos citar o eliminativismo. Os eliminativistas propõem um completo rompimento com categorias do tipo crenças, desejos, etc., consideradas como categorias da psicologia popular. A proposta eliminativista consiste em substituir as categorias da psicologia popular por uma perspectiva teórica com uma nova ontologia e um novo vocabulário extraído diretamente da neurofisiologia.

Em meados dos anos 50 inaugura-se um novo empreendimento científico que visa responder às questões sobre a natureza da inteligência e do conhecimento, bem como de sua forma de estruturação. Tal empreendimento recebeu o nome de ciência cognitiva. Esta pode ser caracterizada como um empreendimento interdisciplinar, que lança mão de máquinas para investigar a natureza e a estrutura da mente cognitiva.

Em nossos dias podemos dizer que essa nova ciência abriga em seu interior dois programas de pesquisas: um 'representacionista' e o outro 'não representacionista'. Nossa dissertação tratará, principalmente, das duas linhas de pesquisas filiadas ao primeiro programa. A primeira linha deste programa é denominada inteligência artificial e a segunda é chamada de conexionismo.

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

A inteligência artificial adota o funcionalismo como perspectiva de trabalho. O funcionalismo, na ciência cognitiva e filosofia da mente, se caracteriza pela tese de que é possível dispensar, nos estudos da cognição, uma particular instanciação material dos processos mentais. Tais processos seriam descritos exclusivamente em termos de uma organização funcional da mente, que se constituiria por exemplo, de módulos. Tais módulos desempenhariam funções especificadas por classes de representações, causalmente relacionadas, frente ao *input* dado e o *output* desejado.

O funcionalismo computacional, que teve entre os seus fundadores o filósofo Putnam, apresentou-se por um bom tempo como uma via intermediária entre o dualismo de substâncias (do tipo cartesiano) e o reducionismo behaviorista. Os estados mentais, segundo esta concepção, podem ser caracterizados apenas por suas inter-relações funcionais. Estes estados, por sua vez, podem ser instanciados em qualquer sistema com capacidade de processar informação estruturada. Tal capacidade não será exclusiva do cérebro humano, mas de qualquer sistema que processe este tipo de informação.

O conexionismo, a segunda linha de pesquisa existente no programa reprentacionista, distancia-se da perspectiva funcionalista radical, subjacente aos modelos da inteligência artificial. Segundo essa linha de pesquisa, a tese funcionalista, que abstrai os elementos físicos da caracterização dos processos cognitivos e toma por objeto exclusivo os elementos estruturais\funcionais dos processos mentais, não esboça uma visão adequada da natureza da cognição. Para os conexionistas, os processos cognitivos pressupõem uma instanciação (como elemento constituinte) em arquiteturas capazes de

processar informação do tipo distribuída e em paralelo. Além de processar informação distribuída e em paralelo, tal processamento dar-se-ia num plano subsimbólico. Para o conexionista, o significado de uma representação está associado não ao par regra/significado, como querem os pesquisadores da inteligência artificial, mas a um estado físico global da rede.

Para abordar o tema mencionado, a saber, considerações filosóficas sobre as correntes representacionistas da ciência cognitiva que inclui a inteligência artificial e o conexionismo, estruturamos nossa dissertação em duas partes. Na primeira (I) intitulada: *As Bases Histórico-Filosóficas da Ciência Cognitiva,* temos os capítulos 1 e 2. Na segunda (II) denominada: *Alcance e Limites do Mecanicismo na Explicação dos Processos Mentais,* temos os capítulos 3 e 4.

No capítulo 1, da primeira parte, nos debruçamos sobre as raízes históricas da ciência cognitiva. Traçamos um histórico do surgimento dessa nova ciência, bem como seus pressupostos básicos. Em seguida, tecemos algumas considerações sobre os fortes vínculos entre a ciência cognitiva e a cibernética. Nossa pretensão foi retomar algumas noções chave presentes no movimento cibernético, em especial, no pensamento de Wiener, Von Neumann e Ashby. Mostramos como noções discutidas no movimento cibernético ainda fazem parte da agenda de discussões dos cientistas cognitivos atuais, particularmente, em seu programa 'não representacionista'. Neste contexto, analisamos a formação desse novo campo de pesquisa enfocando a proposta de Von Eckardt sobre os paradigmas na ciência cognitiva. Por fim, discutimos, brevemente, este novo programa de pesquisa conhecido como *ciência cognitiva dinâmica*. Tal programa



aponta para a necessidade de discutir uma nova conceituação de paradigma, proposta por Morin, que recebeu o nome de *paradigma da complexidade*. O objetivo central deste capítulo é apresentar os fundamentos desta nova ciência da mente.

No capítulo 2, após a análise dos pressupostos básicos e do domínio desse novo campo do conhecimento, bem como dos seus vínculos com a cibernética, abordamos a seguinte questão: teria a ciência cognitiva, em sua vertente denominada IA, raízes na filosofia cartesiana? Essa questão constitui o fio condutor do capítulo em questão. Discutimos inicialmente a aproximação sugerida por autores como Penrose e Searle, entre o projeto de pesquisa da inteligência artificial e a filosofia cartesiana. Tal discussão, ao nosso ver, pode ser proveitosa à medida que recuperamos alguns elementos da filosofia cartesiana para decidirmos se a aproximação sugerida pelos autores mencionados está bem fundamentada. Em seguida apresentamos a disputa que se estabelece entre Descartes e seus críticos, acerca da natureza do conhecimento, a fim de encontrar elementos que julgamos presentes nas pesquisas desenvolvidas na ciência cognitiva.

O capítulo 2, que encerra a primeira parte da nossa dissertação, visa iniciar as considerações filosóficas sobre as duas principais vertentes existentes na ciência cognitiva. Procuramos, com estas considerações iniciais sobre a inteligência artificial e o conexionismo, apontar pressuposições filosóficas que julgamos presentes nestes programas de pesquisa, bem como preparar o terreno para a segunda parte da nossa dissertação que tratará do alcance e limites do mecanicismo na explicação dos processos mentais.

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

A segunda parte (II), intitulada: *Alcance e Limites do Mecanicismo na Explicação dos Processos Mentais,* está também dividida em dois capítulos. No capítulo 3, discorremos sobre a base intuitiva da máquina de Turing para, em seguida, vermos como o funcionalismo computacional adota o modelo mecanicista proposto por Turing e como a inteligência artificial, que segue a tese funcionalista computacional, constrói seus modelos seguindo tais especificações. Feita essa análise, apontaremos o alcance e os limites, dos modelos da inteligência artificial.

No capítulo 4, investigamos se os modelos de redes neurais, que possuem clara inspiração neurofisiológica, apontam (ou não) para uma saída dos problemas encontrados pelo funcionalismo computacional radical e pelos modelos da inteligência artificial na caracterização dos processos cognitivos. Esta investigação sobre os modelos de redes neurais, como no caso da análise dos modelos da inteligência artificial, pretende-se crítica. Isto é, buscamos estabelecer até onde vão e de que natureza são as vantagens oferecidas pelos conexionistas no estudo destes processos.



INTRODUÇÃO

O objetivo deste capítulo é traçar um histórico do surgimento da ciência cognitiva procurando apresentar os fundamentos desta nova ciência. Conforme mencionamos na introdução geral, essa nova ciência agrupa dois programas de pesquisas, a saber, o *representacionista*, do qual fazem parte a inteligência artificial (daqui por diante IA) e o conexionismo ou redes neurais artificiais (daqui por diante RNA) e o *não representacionista* ou *ciência cognitiva dinâmica*. Neste último temos os pesquisadores ligados à psicologia ecológica e às teorias de auto-organização e percepção direta.

Nossa discussão estará centrada nas hipóteses e modelos desenvolvidos pelo programa representacionista. Este programa apresenta duas principais linhas de pesquisa a IA e a RNA. Para os adeptos da IA, o conhecimento dos processos cognitivos deve ser caracterizado através de uma série de operações de *regras abstratas* sobre proposições que são representadas pela mente humana.

Já os conexionistas argumentam que os inúmeros fracassos acumulados pela IA residem, justamente, em sua caracterização simbólico-abstrata, fundada em regras, acerca do conhecimento. Sendo assim, os conexionistas julgam que uma caracterização adequada do conhecimento deve levar em conta não tanto as regras abstratas que o estruturam, mas, sim, as leis físicas que atuam na micro estrutura (preferencialmente neurofisiológica) dos sistemas físico-biológicos que processam informação. Segundo eles, um bom candidato a modelo dos estados mentais deve ter a capacidade de interagir com o

meio ambiente e se auto-organizar frente as novas informações.

Para entendermos plenamente a proposta de cada linha de pesquisa, ou seja, da IA e dos conexionistas, faz-se necessário um breve histórico sobre o surgimento desse nosso campo do saber intitulado ciência cognitiva.

Na seção 1.1, percorremos a caracterização histórica dessa nova ciência, bem como discutimos seus pressupostos básicos. Vencida esta etapa, na seção 1.2, iremos tecer algumas considerações acerca dos fortes vínculos existentes entre a cibernética e a ciência cognitiva. Em seguida, na seção 1.3, analisamos a formação desse novo campo de pesquisa científica, a partir da proposta de Von Eckardt (1995). Investigamos, ainda nesta seção, se essa jovem disciplina científica não se adequaria a proposta feita por Morin (1996) de um paradigma da complexidade.

1.1 Dos pressupostos básicos da ciência cognitiva: interdisciplinaridade e o emprego do computador

De acordo com o psicólogo e historiador Gardner (1995;p.19), a ciência cognitiva foi oficialmente reconhecida como um novo campo de conhecimento em meados dos anos cinqüenta. Esse novo campo, ainda de acordo com o historiador, pode ser definido como:

> ...um esforço contemporâneo, com fundamentação empírica, para responder questões epistemológicas de longa data – principalmente aquelas relativas à natureza do conhecimento, seus componentes, suas origens, seu desenvolvimento e seu emprego.

Gardner, com essa ambiciosa caracterização acerca da natureza e do domínio desse novo campo do conhecimento, acredita estar explicitando os

aspectos mais fundamentais que norteiam o empreendimento cognitivo-científico, a saber, que a discussão da atividade cognitiva humana pressupõe uma análise que deixa de lado os aspectos biológicos e culturais. Tal plano de análise é chamado de *representacional* ou plano das representações mentais. Vemos que a caracterização, por ele sugerida, abarca apenas parte do empreendimento científico cognitivo, uma vez que deixa de lado o programa conhecido como ciência cognitiva dinâmica, que não considera as representações mentais na explicação da atividade cognitiva.

Tendo claro o plano de análise, a fundamentação empírica das hipóteses sugeridas pelos cientistas cognitivos sobre a cognição humana se concretizará através do seguinte pressuposto: as máquinas representam bons modelos do funcionamento da mente humana. Como exemplo destas máquinas temos o computador. Este servirá, principalmente na IA, como modelo para explicar como a mente funciona no processo de aquisição do conhecimento.

Outro aspecto importante presente no empreendimento dessa nova área do saber é a desenfatização no estudo da cognição dos fatores culturais, históricos, afetivos e emocionais. Um quarto aspecto presente nesse empreendimento é que essa nova ciência herda questões epistemológicas que há séculos enfrentam os filósofos.

O último aspecto dessa nova ciência que, a nosso ver, não pode ser depreendido da caracterização proposta por Gardner, apesar do autor mencioná-lo como um dos aspectos fundamentais deste novo empreendimento científico, é a crença no estudo *interdisciplinar*. Para nós, os dois aspectos que tornam singular essa nova área do saber é o emprego do computador e a

interdisciplinaridade no estudo dos processos cognitivos.

Esta nova ciência, que se pretende interdisciplinar, procura abrigar cientistas de várias áreas, que lidam com diferentes aspectos dos processos cognitivos, bem como filósofos, que procuram unificar e formar uma versão coerente dos resultados obtidos em cada área.

A preocupação de designar a filosofia como a disciplina encarregada de unificar e coordenar as diferentes formas de abordagem acerca dos processos cognitivos foi explicitada pelo filósofo Dupuy (1996;p.113):

> O que mantém juntos, hoje em dia, os múltiplos programas agrupados sob o nome de "Ciências Cognitivas" é o trabalho filosófico que é feito acerca deles. Sem a filosofia "Cognitiva" haveria trabalhos nas áreas da Psicologia, da Lingüística, da Neurobiologia, da Inteligência Artificial – não haveria ciência da cognição. A filosofia é que reflete e sistematiza a ou as atitudes de base que constituem o único laço social no interior da área.

A preocupação em designar uma disciplina cuja função é unificar as idéias das diversas áreas do saber advém, como o próprio Dupuy esclarece, da complexidade do objeto de estudo. Dado que muitos problemas acerca dos processos cognitivos situam-se nas fronteiras entre disciplinas como Psicologia, Lingüística, Neurobiologia e a Lógica, corre-se o risco, sem o trabalho de unificação, de termos disciplinas trabalhando isoladamente, sem qualquer comunicação e conhecimento sobre os avanços acerca do mesmo objeto conquistado por outras disciplinas. Em outras palavras, sem o trabalho unificador da filosofia, a atividade científica se voltaria para a abordagem da cognição de maneira não *interdisciplinar*, isto é, o psicólogo isolado tentaria uma explicação, o mesmo ocorrendo com o lingüista, o lógico, etc.

Apesar de ressaltar a importância do trabalho unificador do

filósofo, Dupuy ainda está pensando nas ciências cognitivas. Ou seja, ciências que manteriam sua própria metodologia e que, a partir dos resultados obtidos, se comunicariam para estabelecer uma ponte com as restantes áreas do saber que se debruçam sobre o mesmo objeto.

Contudo, devemos alertar que a pretensão do cientista cognitivo é mais ambiciosa. Acredita-se que a comunicação *interdisciplinar* só será satisfatória se todos os esforços nas diversas áreas do saber convergirem para uma nova área transdisciplinar.

Romper-se-iam as dificuldades acerca dos vocabulários próprios de cada área, bem como da metodologia especializada a cada saber e se inauguraria, não mais uma cooperação *interdisciplinar*, mas um estudo integrado dos diversos aspectos da cognição. Apoiado em uma *única metodologia*, ainda que multifacetada, teríamos, idealmente, um novo campo capaz de, dada a sua transdisciplinaridade, abordar a complexidade do mental.

A pretensão de uma ciência da cognição apoiada, principalmente, na unidade metodológica das ciências, ganha críticos ferrenhos, mas também, desperta entusiastas fervorosos como é o caso de Morin (1996;p.135):

> Sabemos cada vez mais que as disciplinas se fecham e não se comunicam umas com as outras. Os fenômenos são cada vez mais fragmentados, e não se consegue conceber a sua unidade. É por isso que se diz cada vez mais: "Façamos interdisciplinaridade". Mas a interdisciplinaridade controla tanto as disciplinas como a ONU controla as nações. Cada disciplina pretende primeiro fazer reconhecer sua soberania territorial, e, à custa de algumas magras trocas, as fronteiras confirmam-se em vez de se desmoronar.

Na visão de Morin, a complexidade do real impõe um saber complexo. A promessa de que a *interdisciplinaridade* levaria a um saber

complexo redundou em um grande fracasso. Como exemplo desse grande fracasso podemos citar a tentativa de *interdisciplinaridade* levada a cabo pela primeira cibernética. O que se sabe sobre essa tentativa é que toda sugestão advinda das ciências humanas e biológicas era, primeiramente, traduzida para o vocabulário da física para depois ser discutida.

Essa atitude aos poucos afastou os pesquisadores das ciências humanas e biológicas interessados em estudar a cognição, fazendo assim da primeira cibernética um grupo quase que exclusivamente formado por físicos, matemáticos e engenheiros. Sendo assim, a promessa de *interdisciplinaridade* pretendida pela primeira cibernética não passou de uma falsa promessa. Ou se falava no vocabulário de física ou não se participava das discussões.

Essa primeira tentativa de *interdisciplinaridade* acabou isolando ainda mais as disciplinas interessadas em estudar os vários aspectos da cognição. A lição que a ciência cognitiva pode extrair dessa tentativa é que os processos cognitivos, dada a sua complexidade, dificilmente serão compreendidos por disciplinas isoladas.

Somente a explicação do biólogo não esgota a riqueza da cognição. Também, não bastam apenas explicações psicológicas ou lingüísticas. Deve-se, pensa Morin, ir além dos métodos mutiladores do saber especializado. E por isso que o seu recado encaixa-se perfeitamente com os planos de uma ciência cognitiva: "... é preciso ir além, e aqui aparece o termo 'transdiciplinaridade'." (Morin 1996;p.135*)*.

Essa sede de uma metodologia unificadora de trabalho presente na ciência cognitiva é, sem dúvida, herdada da tradição filosófica chamada

analítico-positivista. A chamada epistemologia analítico-positivista foi, e ainda é, como mostraremos, fortemente criticada por muitos filósofos. Uma das críticas mais importantes é que, no anseio de englobar várias disciplinas sob uma única metodologia, acaba-se, ao invés de abarcar a complexidade do objeto, reduzindo-o a apenas alguns conceitos específicos de uma disciplina. Isto ocorre porque o pensamento reinante na epistemologia analítico-positivista não admite a existência de uma diferença essencial entre as ciências naturais (neste caso a física, a química e a biologia) e as ciências humanas, por exemplo, (antropologia, psicologia).

Os conceitos das ciências humanas, na tradição analítico-positivista, são reduzidos aos conceitos das ciências naturais. Pressupõe-se ainda que os métodos das ciências naturais são adequados para abordar, no caso da ciência cognitiva, os aspectos culturais da cognição.

Mas, como sustenta Oliveira em seu artigo 'O que todo cientista cognitivo deve saber sobre a lógica' (1996;p.17): " Existem diferenças essenciais entre as ciências naturais e as ciências humanas, e os métodos daquelas não são adequados a estas. "

Esta será a tese central da crítica de Oliveira (1996) à ciência cognitiva. Ou seja, que esta nova ciência comete equívocos fundamentais, decorrentes de seu naturalismo, ao ignorar as diferenças entre as ciências naturais e as ciências humanas, e acreditar que os métodos das primeiras são adequados às últimas.

Uma profunda diferença entre as ciências naturais e as ciências humanas, apontada por Oliveira a partir das idéias desenvolvidas por Taylor, é

que:

> ...as ciências naturais - mas não humanas – tratam de objetos que são independentes do processo de investigação. Nas ciências humanas não se verifica tal independência em seus domínios, o processo de investigação pode, e costuma influenciar a própria realidade a que a investigação se dirige. (Oliveira 1998;p.2).

O que Oliveira procura apontar é que nas ciências humanas tem-se um impacto direto do processo de investigação sobre a realidade investigada. Na verdade, Oliveira vai além. Ele toma esta tese e a generaliza a todas as etapas da investigação científica. Ou seja, nas ciências humanas nós temos o que ele denomina de: *transformação cognitiva*. Isto é, o fenômeno em que o processo de investigação *científica afeta seu próprio objeto.*

Além deste fenômeno de transformação cognitiva, Oliveira alerta para o perigo de uma ciência que não leva em consideração a diferença entre o cultural e o natural. Em sua opinião, cair no erro de abordar o cultural como natural é condenar as explicações sobre a cognição humana a um modelo explicativo simples e pobre. A solução apontada por ele é desmembrar a ciência cognitiva em dois domínios, *o do Ciência Cognitiva Natural e o da Ciência Cognitiva Cultural.*

Porém, esse desmembramento parece, como sugere Oliveira, muito distante de ser aceito pelos cientistas cognitivos. Se existe algo que a ciência cognitiva busca fortalecer de sua herança 'cultural' é a sua 'tendência naturalizante', como tentamos explicitar no início deste trabalho, enfocando a proposta da transdiciplinaridade.

Referimo-nos aqui à 'tendência naturalizante' como uma 'herança cultural' pois, é sabido que a ciência cognitiva está profundamente

enraizada na mentalidade científica e tecnológica do século XX. Como exemplo desse estreito laço temos a hipótese adotada pelo cientista cognitivo, principalmente aqueles envolvidos com IA, segundo a qual o computador eletrônico constitui um bom instrumento explicativo do funcionamento da mente humana. Ou seja, o computador servirá como um instrumento apropriado para simular a atividade cognitiva.

Lançando mão do computador como instrumento apropriado para a elaboração de modelos que simularão as atividades cognitivas, o cientista cognitivo acredita estar dando um passo gigantesco rumo à uma explicação destas atividades. Para estes cientistas, com uma ciência da cognição pode-se abandonar a velha especulação de gabinete, característica do filósofo quando este debruça-se sobre as questões relativas à natureza do conhecimento, em prol de uma metodologia que permita testar e verificar as hipóteses explicativas sobre estas atividades.

Para o cientista cognitivo é de suma importância recorrer a *métodos empíricos* que permitam tornar suas hipóteses acerca da atividade cognitiva passíveis de verificação. Com esse intuito, não há duvida de que a invenção do computador nos anos 30 e 40 foi uma mola propulsora do ideário verificacionista presente na metodologia adotada pelo cientista cognitivo.

Como comenta Gardner (1995;p.55) à respeito do computador:

> Além de servir como modelo do pensamento humano, o computador também funciona como uma ferramenta valiosa para o trabalho científico cognitivista. A maioria dos cientistas cognitivos o utiliza para analisar os seus dados, e um número crescente destes cientistas tentam simular processos cognitivos nele.

Essa dupla utilidade do computador eletrônico — de análise de

dados e modelagem explicativa — para o trabalho do cientista cognitivo fica, também, claramente explicada na Introdução dos organizadores do livro *Encontro com as Ciências Cognitivas*, volume I:

> O computador funciona não apenas como instrumento explicativo, das estruturas e processos que, a partir de uma perspectiva mecanicista, se supõem presentes na atividade cognitiva, mas, principalmente, como teste das hipóteses subjacentes aos modelos computacionais (as simulações) que se pretendem explicativos dessa atividade. (Gonzales et al.1996;p.10)

Se, para alguns, o computador deve ser visto como um bom instrumento para modelagem do mental, para outros, como é o caso do filósofo Putnam[1] (1980), essa nova invenção pode também revelar importantes idéias acerca da natureza dos estados mentais. Foi fundamentado nesta crença que ele funda uma vertente na filosofia da mente conhecida como funcionalismo computacional.

Os estados mentais, segundo Putnam, devem ser abordados em termos de suas estruturas funcionais, isto é, de suas estruturas de *input* e *output*. Para explicar melhor a sua hipótese funcional, ele propõe em seu artigo — '*The Nature of Mental States*' —, uma análise do estado de dor como um estado funcional de um organismo inteiro.

Para isso, Putnam recorre à noção de máquina de Turing como um modelo para o organismo. Segundo sua abordagem, descrever um estado como sendo um estado de dor, por exemplo, nada mais é que situar sua posição em uma organização funcional:

---

[1] Estamos nos referindo explicitamente à primeira fase de Putnam chamada de funcionalista computacional. Esse filósofo em 1995 lançou um livro criticando a sua antiga postura.

> Uma descrição de S onde S é um sistema, é qualquer afirmação verdadeira para o fato de que S possui distintos estados S1,S2....Sn que estão relacionados para um outro e para as saídas motoras e as entradas sensórias e pelas probabilidades de transição em uma tábua de representação. A tábua de representação mencionada na descrição será então chamada de Organização Funcional de S... (Putnam 1980;p.226)

Uma mesma organização funcional, acaba sendo nada mais que um programa, que pode ser rodado em *hardwares* diferentes. Sendo assim, é indiferente que a matéria instanciando um programa seja de carne e osso como nós, ou de silício e *chips* como os computadores. Putnam chamou essa idéia de 'instanciação múltipla'.

A idéia de instanciação múltipla, decorrente da teoria da organização funcional, logo despertou nos cientistas cognitivos a conhecida metáfora do software e hardware. Se existisse alguma identidade entre computadores e homens, esta não está no hardware, mas sim no software. Como resume Gardner (1995;p.9): "...tanto os seres humanos quanto as máquinas poderiam ser capazes de realizar os mesmos tipos de programa."

O funcionalismo, com as idéias de Putnam na década de 60, ajudou ainda mais o fortalecimento das pretensões da ciência cognitiva. Essa foi (e ainda é) fortemente influenciada pelas teses funcionalistas de Putnam, chegando mesmo a tomar a metáfora *hardware-software* como literal. Os cognitivistas, principalmente na sua vertente da IA, argumentaram que o *hardware* — que corresponderia ao elemento físico do sistema de processamento de informação — não seria determinante para se entender o *software* — que corresponderia ao elemento abstrato, característico dos processos mentais.

Com a visão reinante da IA durante as décadas de 70 e início dos 80, a ciência cognitiva seguiu seu caminho, apostando em uma unidade



metodológica e no computador como instrumento de simulação das atividades cognitivas. A IA só perdeu sua hegemonia com o crescimento de uma nova vertente chamada RNA.

Em resumo, procuramos nessa primeira seção explicitar os pressupostos fundamentais da ciência cognitiva, que possui, conforme veremos, raízes na cibernética, discutindo dois aspectos fundamentais para a consolidação dessa nova área do conhecimento, a saber, a *transdisciplinaridade* e o *emprego do computador* como elemento explicativo da natureza da Inteligência. Continuaremos nosso histórico, na próxima seção, aproximando a cibernética, e alguns de seus conceitos, ao empreendimento científico desenvolvido pela ciência cognitiva.

1.2 A herança cibernética

Nesta segunda seção, do presente capítulo, pretendemos analisar os vínculos existentes entre o projeto cibernético e a ciência cognitiva. Tal análise é importante para uma visão mais completa sobre o surgimento desta nova ciência da mente. Foi através da cibernética que o pensamento cognitivo científico se viu imerso em idéias complexas extraídas, principalmente, da biologia, física e engenharia, no estudo dos sistemas autônomos.

Para abordar a relação existente entre a cibernética e a ciência cognitiva, apresentaremos as idéias de três cientistas do movimento cibernético, a saber, Wiener (1990), Ashby (1962) e Von Neumann (1958). Não iremos apresentar minuciosamente o pensamento desses três autores, apresentaremos



apenas as noções por eles sugeridas e que foram explicitamente retomadas na ciência cognitiva. Tais noções focalizam o conceito de *causalidade circular*, apresentado por Wiener, a questão da *complexidade*, explicitada por Von Neumann e o conceito de *auto-organização*, tal como apresentado por Ashby.

1.2.1 Wiener e o nascimento da cibernética

A cibernética foi batizada pelo matemático norte americano Wiener como a 'ciência do controle e da comunicação'. Tal ciência procurava fundamentalmente se estruturar em torno das noções de *informação* e *comunicação*.

O nome cibernética não foi escolhido por acaso. Essa palavra vinda do grego, servia para designar aquele saber que permite ao piloto conduzir adequadamente sua embarcação. O mesmo termo foi ainda mais enriquecido através da conotação de 'a arte de governar'. Foi com esta última conotação, que a raiz grega da palavra 'cibernética' se difundiu entre nós.

Wiener (1990;p.15) relata a história do termo:

> ...'cibernética', deriva da palavra grega Kubernetes ou 'piloto', a mesma palavra grega de que evidentemente derivamos nossa palavra 'governador'. Descobri casualmente, mais tarde, que a palavra já havia sido usada por Ampère com referência a ciência política e que fora inserida em outro contexto por um cientista polonês...

O interesse da cibernética converge, principalmente, para o modo como se comportam os organismos e as máquinas. No entanto, o estudo do comportamento de máquinas e organismos, nesta área, não pretende seguir o

esquema estímulo-resposta consagrado pelo behaviorismo.

Há um ingrediente muito importante que impede que o estudo do comportamento inteligente realizado pela cibernética seja visto como uma continuidade do esquema behaviorista. Este ingrediente é a noção de *causalidade circular*. Podemos entender essa noção como uma possibilidade que alguns sistemas possuem de alterar a si e ao seu meio, a partir da combinação de elementos que o circundam. Graças a esta relação, que se estabelece coletivamente entre meio e sistema, é que se modifica a relação tradicional entre um *input* recebido e um *output* desejado, em outros termos, entre a relação de estímulo-resposta.

No entanto, só enxergamos a *causalidade circular* de um dado sistema se estabelecermos dois planos de observação. Olhando para o interior deste, podemos entrever a capacidade que o sistema possui de se ajustar e mudar a sua resposta frente a um dado *input*, a fim de alcançar um objetivo determinado. Tal olhar sugere a idéia de que este sistema tem a capacidade de aprender com seus erros, a fim de atingir uma meta. Será nesse plano de observação que se encontrará a riqueza da noção de *causalidade circular*.

O outro plano de observação se concentra apenas no comportamento exterior do sistema. A partir desse olhar, acredita-se que o sistema é guiado apenas pela relação entre *input* e *output*.

Tradicionalmente a relação de causalidade tem sido caracterizada através de cadeias seqüenciais: um dado evento *A* causa um outro evento, seguinte, *B* . Nesse esquema, *A* aparece como causa e *B* como efeito. Em contraste, uma relação de *causalidade circular* um evento é, ao mesmo tempo,



causa e efeito. Assim, *A* causa e é causado pelo evento *B*. Por exemplo, indivíduos formam sociedades que formam indivíduos e assim sucessivamente.

Como indicou Morin, a causalidade circular é importante quando se pretende explicar fenômenos que envolvem processos de 'auto-organização'. Em nossos dias, o programa de pesquisa da ciência cognitiva, denominado *não-representacionista* ressalta a fundamental importância de se explorar esta nova noção de causalidade como substitutiva da pressuposição (da ciência cognitiva ortodoxa) de representações mentais que intermediariam o conhecimento do meio em que vivemos.

Para os pesquisadores da ciência cognitiva dinâmica o conhecimento se estabelece pela interação direta entre o ser e o meio em que este vive. Neste programa de pesquisa a noção de sistema também é alterada. Sistema, nesta tendência, não é somente, por exemplo, a máquina, mas sim, a máquina e o meio em que esta se insere. Tendo agora como ferramenta explicativa a noção de *causalidade circular* pode-se, acreditavam os cibernéticos e os pesquisadores da ciência cognitiva dinâmica, abordar os fenômenos complexos, tais como aqueles que caracterizam os seres inteligentes. Um dos primeiros cientistas a evocar a noção de complexidade foi, como veremos a seguir, Von Neumann.

1.2.2 John Von Neumann: o paralelo entre o computador e o cérebro

Von Neumann (1958), como comenta Gardner (1995;p.25), foi um dos primeiros a apresentar o paralelo entre o cérebro e o computador:

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

> ...o matemático John Von Neumann, forjou uma comparação extraordinária entre o computador eletrônico (no momento uma descoberta que cheirava ficção científica) e o cérebro (que já existia havia algum tempo).

Devemos lembrar que Von Neumann é o pai do computador moderno. Embora sempre prudente, ele via no computador semelhanças com o cérebro humano. Ele julgava que este poderia, entre outras coisas, servir para entender melhor o funcionamento do cérebro, ou para acrescentar ao cientista um auxiliar importante para os testes de suas hipóteses. Sua máquina era encarada, enfim, como diz Breton (1991;p.130): "...como uma espécie de extensão técnica de seu utilizador, uma espécie de ampliação do cérebro humano que deveria ser levada o mais longe possível."

Além de ser um dos pioneiros do paralelo entre o cérebro e o computador, Von Neumann também chamou a atenção dos cientistas de seu tempo para um dos problemas fundamentais do pensamento científico contemporâneo, a saber, a complexidade biológica.

Von Neumann nos apresenta o que parece ser a diferença fundamental entre o autônomo artificial, mesmo o computador mais moderno, e o autômato natural mais rudimentar, o unicelular, que reproduz a complexidade biológica.

De maneira brilhante Von Neumann coloca no centro das discussões a capacidade que o autômato natural tem de se auto-reproduzir e auto-reparar. Esses parecem ter a capacidade de gerar o seu programa e elaborar estratégias ou condutas inventivas que se modificam segundo as condições vividas. Devido a essa capacidade de elaborar estratégias e condutas de acordo



com a situação vivida, é que ele acreditava que os sistemas naturais toleram uma certa dose de *desordem*. Isto é, são capazes de, em situações às vezes desfavoráveis, traçar uma estratégia que lhes permita suportar as adversidades e continuar vivendo.

Por essas razões, a saber, a capacidade de gerar o seu 'programa' e elaborar suas estratégias de acordo com a situação 'vivida' em seu meio, é que Von Neumann julgava que qualquer tentativa de descrição lógica da complexidade devia ser imediatamente precedida por uma reforma na lógica formal, pois esta, como diz Dupuy comentando uma afirmação de Von Neumann: "...em seu estado presente, é rígida demais, combinatória demais para poder colocar-se como a lógica dos autômatos, sejam eles artificiais ou naturais." (Dupuy 1996;p.80).

A idéia de complexidade biológica, isto é, a capacidade de que todo autômato natural tem de se auto reparar e auto reproduzir, nos conduz à idéia de auto-organização introduzida por Ashby.

1.2.3 As idéias de Ashby sobre auto-organização

As idéias do médico e matemático britânico Ashby são consideradas um dos pilares da segunda geração da cibernética, que caracteriza o neomecanicismo. Como médico, levou até as últimas conseqüências os estudos sobre a complexidade biológica. Como matemático não abria mão de tentar explicar a atividade mental humana de maneira mecanicista.

Segundo Gardner (1995;p.41) o objetivo de Ashby foi:

unesp

> ...simplesmente copiar o cérebro vivo. Em particular se o cérebro vivo falha de certas maneiras características, então quero que o meu cérebro artificial também falhe. Tentei deduzir o que é necessário, que propriedades o sistema nervoso deve ter para se comportar ao mesmo tempo de forma mecanicista e adaptativa.

Será a partir de uma experiência de pensamento que Ashby (1962) irá procurar ilustrar uma tese que ele gostaria que fosse revestida de um caráter universal, a saber: longe de serem acidentes extremamente improváveis da evolução, a vida e a inteligência desenvolvem-se necessariamente em sistemas semi-isolados. Estar vivo está intimamente ligado à capacidade de manutenção de um pequeno número de variáveis essenciais no interior de sistemas que interagem com seu meio ambiente.

Assim, para Ashby, basta submetermos um sistema físico qualquer a leis físicas invariáveis para criarmos a possibilidade de geração da vida. A inteligência, nesta perspectiva, será caracterizada como a capacidade que um tal sistema tem de manter os seus parâmetros essenciais, criados a partir de leis físicas invariáveis, inalterados. Em outras palavras, a inteligência de um sistema é medida pela sua capacidade de manter-se vivo.

Contudo, se Ashby acreditava que vida e inteligência podem ser geradas em qualquer sistema semi-isolado por leis invariáveis, o mesmo não ocorre acerca dos processos de 'auto-organização', que ele julga não existir, *em princípio*, nas máquinas. Pois, segundo Ashby, uma máquina pode ser definida como uma função que mapeia um estado interno em outro estado interno. A partir dessa definição de máquina, entendida como uma função, é logicamente incorreto esperar que uma máquina altere sua própria função (que neste caso implicaria em alterar a si próprio).

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

Esse mesmo ponto é mencionado por Dupuy (1996;p.205) acerca da impossibilidade da 'auto-organização' nas máquinas:

> Para que haja auto organização seria preciso, com efeito, que o sistema pudesse, por seu próprio funcionamento, determinar a mudança da função que o rege. Mas, para isso, dever-se-ia imaginar uma outra função, de um nível lógico superior à primeira, que a modificasse. E neste caso o sistema teria sido definido de maneira errada pela função inicial.

Essa impossibilidade lógica, apontada por Ashby, de uma máquina artificial alterar, por si só a função que a rege e, por consequência, se auto-organizar, aguçou ainda mais as investigações cognitivistas acerca desse fenômeno de auto-organização.

Ainda não se chegou a uma conclusão definitiva sobre a possibilidade ou impossibilidade de auto-organização em sistemas artificiais. Porém, o importante é que as idéias de Ashby, lançadas há mais de trinta anos, ainda estão presentes nas mais recentes discussões sobre auto-organização.

Em resumo, a nossa pretensão em retomar algumas idéias de três cientistas ligados à cibernética foi enfatizar os fortes vínculos existentes entre essa ciência e a ciência cognitiva. Para nós, as noções de causalidade circular proposta por Wiener, o paralelo entre o cérebro e computador eletrônico, bem como a problemática da complexidade tratada por Von Neumann desembocam em uma noção central para o programa não representacionista existente dentro da ciência cognitiva, a saber, a noção de 'auto-organização' inicialmente formulada por Ashby.

Nosso panorama histórico procurou estabelecer uma herança que, segundo Dupuy, havia se perdido na ciência cognitiva, a saber, a herança

cibernética. Na visão desse autor, a postura dos adeptos da IA, que reinou na ciência cognitiva até o início dos anos 80, rompia com as influências das noções que apresentamos nesta seção e que foram formuladas pelos cibernéticos.

Ao nosso ver Dupuy estava reivindicando o advento de um programa de pesquisa que reatasse os laços com a cibernética. Tal programa não só assumiria sua herança com a cibernética como também, estabeleceria no futuro um paradigma da complexidade. Tal paradigma, como mostraremos a seguir, tenta incorporar, em seus modelos para simular a cognição humana, noções como: *auto-organização* e *causalidade circular*. A seguir veremos como se forma esta nova ciência na visão de Von Eckardt.

1.3 A proposta de Von Eckardt e as semelhanças com o paradigma de Kuhn

Faremos, nesta seção, algumas breves considerações concernentes à formação desse novo campo do conhecimento intitulado ciência cognitiva. Para tanto usaremos, principalmente, algumas idéias que Von Eckardt desenvolve no seu livro *What is Cognitive Science*?

Von Eckardt (1995;p.2) parte de uma dúvida levantada por Gardner:

> Em A Nova Ciência da Mente, Howard Gardner sugere que 'a ciência cognitiva não mantém um acordo acerca de seu paradigma de pesquisa'- não existe um consenso sobre as hipóteses ou métodos de pesquisa...

Inquietada por essa afirmação de Gardner, segundo a qual não

há ainda na ciência cognitiva um acordo sobre um paradigma de pesquisa e nem mesmo sobre as hipóteses ou métodos, Von Eckardt irá, inspirada na noção de paradigma de Kuhn (1997), propor uma estrutura de compromissos compartilhados pelos cientistas cognitivos (daqui por diante adotaremos a abreviatura ECC, para tal estrutura de compromissos).

Essa ECC, acredita Von Eckardt, pode fornecer à ciência cognitiva uma unidade paradigmática e, com isso, uma unidade metodológica. Pois, segundo essa autora, o enquadramento da ciência cognitiva em uma ECC irá permitir investigar a coerência interna dessa nova ciência, além de definir os compromissos e comprometimentos compartilhados pelas várias disciplinas abrigadas pelo título de ciência cognitiva. Em seguida, a partir de uma metodologia única, será possível a análise dos resultados dos diversos trabalhos desenvolvidos nessa área.

Além dos requisitos imposto pela ECC, Von Eckardt afirma que uma estrutura de pesquisa deve conter os seguintes elementos:

> Uma estrutura de pesquisa consiste de quatros elementos: o conjunto de hipóteses pré-teoricas que especificam o domínio estudado; o conjunto básico da questões empíricas, formulado pré-teoricamente; um conjunto de hipóteses que apresentam as questões básicas e aproximam as possíveis respostas a estas questões, e o conjunto de hipóteses metodológicas. (Von Eckardt.1995;p.18)

Ao indicar esses quatro elementos, a saber, o conjunto de hipóteses pre-teóricas, as questões básicas, as hipóteses que aproximam uma resposta e as hipóteses metodológicas, Von Eckardt acredita que está fornecendo uma estrutura de pesquisa capaz de responder cientificamente as questões de um certo domínio. No caso da ciência cognitiva essa estrutura é tomada para

responder as questões concernentes à natureza dos processos cognitivos.

As especificações pré-teóricas servem para explicitar se o fenômeno estudado é real ou apenas logicamente possível, quais são suas propriedades e a que classe pertence. Cabe as especificações pré-teóricas delinear a que domínio pertence o fenômeno estudado.

Estabelecendo-se o domínio a que pertence o fenômeno, extrai-se um conjunto de questões básicas passíveis de resposta empírica. Em seguida, procura-se hipóteses que respondam aos problemas decorrentes das questões básicas levantadas.

O último passo será estabelecer um conjunto de hipóteses metodológicas que sirvam para indicar os caminhos apropriados para abordar, de maneira eficiente, o fenômeno estudado.

Esses quatro elementos estruturados fornecem uma estrutura de pesquisa que, por sua vez, estabelece um grau de comprometimento participado pela comunidade científica. Ou seja, na visão de Von Eckardt, quando uma estrutura de pesquisa consegue estabelecer esses quatro conjuntos de elementos ela não só edifica um paradigma, como também fornece resultados que serão passíveis de apreciação pela comunidade científica.

O que Von Eckardt pretende é aplicar a noção de paradigma como foi proposta por Kuhn (1997) como uma possível solução para uma definição desse novo campo do saber. Retomando brevemente a noção de paradigma, tal como proposta por Kuhn, veremos que este contém: leis, pressupostos metafísicos, definições, problemas, fatos aceitos, valores de decidibilidade ou escolha, etc. Ou seja, o paradigma irá nos fornecer não só uma

unesp

visão de mundo, como também os problemas e os métodos com que um dado campo do saber tem que lidar para gerar conhecimento.

Von Eckardt, ao propor os quatro elementos que estruturam uma ECC, está sugerindo um paradigma que possa lidar com as questões que surgem no âmbito de uma ciência. Não seria, por acaso, o conjunto de hipóteses pré-teóricas do domínio estudado, algo parecido com os pressupostos metafísicos mencionados por Kuhn? Mas, o conjunto de hipóteses que aproximam respostas às questões básicas, bem como limitam possíveis respostas a essas questões, não seria aquele conjunto de leis contidas no paradigma que responde aos problemas por ele suscitado? A limitação das possíveis respostas, por sua vez, não contém a idéia do que Kuhn designou dogma na ciência?

Bem, essas questões mereciam uma atenção especial. Acreditamos que, em linhas gerais, o conjunto de comprometimentos proposto por Von Eckardt, guarda estreitas semelhanças com a figura do método kuhniano. Se isso é verdade, voltamos à questão inicial: qual o paradigma dominante na ciência cognitiva?

De modo a responder tal questão, seria apropriado lembrar que encerramos a seção 1.1 afirmando que em meados dos anos 80 uma linha de pesquisa denominada RNA vem ganhando força na ciência cognitiva, ameaçando a supremacia da IA. Além disso, como dissemos na introdução, esta nova ciência abriga dois programas distintos de pesquisa: um representacionista (do qual fazem parte a IA e RNA) e um não representacionista ou dinâmico (que engloba as teorias de auto-organização e percepção direta).

Com este segundo programa de pesquisa não temos mais um

unesp

único paradigma predominando na ciência cognitiva. Na visão de Kuhn, teríamos dois programas, a saber, o representacionista e o dinâmico, tentando se tornar o paradigma dominante. Não nos concentraremos, nesta dissertação, neste problema. Procuraremos saber o que podemos aprender com a ciência cognitiva e sua pretensão de simular alguns processos cognitivos em modelos mecânicos.

Podemos apenas dizer, até aqui, que a ciência cognitiva vive, como diz Kuhn, um período de ciência extraordinária. De modo a fundamentar esta afirmação, iremos fazer algumas breves considerações acerca das duas tendências representacionistas existentes no estudo da cognição presentes na ciência cognitiva, a saber, IA e RNA.

Apesar de suas diferenças, podemos destacar alguns pontos em comum entre o projeto da IA e o Conexionista: tanto a IA quanto a RNA postulam um plano representacional de análise no estudo da atividade cognitiva. Ambos procuram estudar o comportamento inteligente através do emprego de modelos mecânicos. Esses, por sua vez, servem para estabelecer hipóteses e teorias acerca do funcionamento e estrutura dos estados mentais. Os aspectos emocionais e sociais que possivelmente originaram os estados mentais são, num primeiro momento, deixados de lado. Porém, esse parentesco quanto a certos pressupostos, não esconde, como veremos, as diferenças entre a IA e RNA na forma de abordar seu objeto de estudo.

O objetivo da IA é simular por meio de máquinas, normalmente computadores eletrônicos, alguns processos cognitivos e, talvez através dessa simulação ou mesmo falhando nessa tentativa, aprender um pouco mais sobre o funcionamento do cérebro.



A proposta na IA é entender a cognição humana, a partir de uma perspectiva de análise independente do físico e do biológico, conhecida como *plano representacional*. Ao optar pela postulação desse plano independente de análise, assume-se que os estado mentais são *abstrações* que podem ser abordadas em termos de símbolos e regras.

Postulando o plano representacional de análise, a IA acredita ser possível tratar a atividade mental como sendo uma seqüência bem definida de operações, denominada algoritmo. Devemos entender nesse contexto o termo algoritmo, como um procedimento de cálculo. Dessa forma, qualquer tipo de atividade mental de um cérebro humano teria que ser um algoritmo, ainda que sua descrição fosse muito mais complexa do que um algoritmo de, por exemplo, um termostato.

Um pressuposto é importante na proposta da IA, acerca do funcionamento e estrutura dos estado mentais, a saber: não existe diferença, em princípio, entre o funcionamento do cérebro humano e de um aparelho como, por exemplo, o termostato. A diferença entre ambos não está no funcionamento de um ou outro, mas, reside apenas na forma de descrever os objetos em questão.

Além de tratar a atividade mental como a realização de um algoritmo, os cientistas que trabalham na IA acreditam que esse algoritmo pode ser instanciado em qualquer matéria. Como já mencionamos, essa idéia de instanciação múltipla, herdada da primeira fase do filósofo Putnam, explicita que tanto faz se um certo algoritmo está sendo realizado no cérebro, no computador ou em latas amarradas por barbantes.

A idéia é que a estrutura lógica do algoritmo possui um

unesp

significado para o estado mental que deve representar, sendo totalmente irrelevante a materialização física particular desse algoritmo. Essa abordagem pode levar, em princípio, a uma forma de dualismo, uma vez que os estados mentais não seriam compostos de matéria física, podendo até existirem independentemente dela.

Neste caso, os cientistas cognitivos da IA não levam em conta a matéria física nas suas explicações. Eles acreditam que as atividades mentais podem ser simuladas apenas através da estrutura lógica de um algoritmo, associando-se assim a uma forma de dualismo, como comentam Searle (1987) e Penrose (1993). Contudo, o dualismo existente nesta linha de pesquisa da ciência cognitiva é, ao nosso ver, um dualismo de propriedades.

Já os conexionistas desenvolvem modelos que, em vez de instanciar operações seriais ou algorítmicas sobre símbolos, incorporam milhares de conexões entre centenas de unidades neurônio-símile (denominados daqui por diante *nódulos*). As redes resultantes dessas conexões entre nódulos apresentam a sinalização de excitações e inibições de uma unidade para outra.

A simulação de um dado processo cognitivo ocorrerá em conseqüência da alteração dos pesos das conexões entre seus nódulos. A tarefa de simulação dos processos mentais se conclui quando o sistema 'se acomoda' provisoriamente através da formação de padrões estáveis de conectividades, denominados *representações mentais*.

Os modelos de RNA simulam não tanto os aspectos racionais da cognição, mas a *dinâmica* que ocorre, por exemplo, na atividade perceptual. Nos primeiros anos de existência desta vertente da ciência cognitiva, seus modelos



foram elaborados para resolver problemas extraídos das diversas áreas da percepção.

Atualmente, os conexionistas se preocupam em explorar a dinâmica dos seres que se 'auto-organizam' em função dos seus estados internos e dos eventos que ocorrem no meio ambiente. Para eles o termo 'auto-organização' tem vital importância. Isto porque acredita-se, nesta tendência, que o emprego de regras abstratas ou algoritmos para a simulação da atividade cognitiva — como pretende a IA — não leva em conta um elemento importante na caracterização do conhecimento, a saber, as leis físicas.

Como diz Gonzales (1996) nos modelos conexionistas auto-organizados nenhuma regra é *explicitamente* incorporada, apesar de eles se comportarem como se estivessem seguindo regras abstratas pré-estabelecidas. Será a partir de um treinamento que este modelos 'aprendem' a encontrar os padrões representacionais corretos que os capacitam a resolver determinados problemas, sem a necessidade de um programa que explicite as regras a serem seguidas na resolução.

O termo 'auto-organização' está presente no princípio de funcionamento desses modelos. De acordo com as concepções de Ashby (1962) e Debrum (1996) temos dois sentidos de 'auto-organização': 'auto-organização primária' e 'auto-organização secundária'.

O termo 'auto-organização primária' refere-se à dinâmica de interação entre elementos independentes que interagem e criam uma relação de dependência mútua entre si, gerando novas formas de organização. A noção de 'auto-organização secundária' proposta por Debrum (1996) pressupõe que as

unesp

partes de uma totalidade já estejam relacionadas, isto é, que exista já uma certa 'organização primária'. A essa 'organização primária' acrescentam-se mecanismos de evolução, como por exemplo, o *feedback*, possibilitando o fenômeno da aprendizagem.

De maneira muito resumida, procuramos mostrar que os modelos conexionistas fundamentam-se na noção de 'auto-organização' que permite a formação de representações, ou padrões estáveis de conectividade entre nódulos, através do processo de interação das nódulos com o seu meio. Aqui está presente a idéia de padrões que se formam 'por si mesmos'. Estes, por sua vez, envolvem, de forma ainda embrionária, a noção de complexidade que será amplamente explorada pela *vertente dinâmica da ciência cognitiva* como esboçaremos a seguir.

1.3.1 A ciência cognitiva dinâmica e o paradigma da complexidade

Este tópico tem por objetivo apresentar as linhas gerais do segundo programa de pesquisa existente na ciência cognitiva batizado de *ciência cognitiva dinâmica*. Entre os autores que trabalham neste programa de pesquisa podemos citar, entre outros, Van Gelder (1998) e Thelen e Smith (1994). A proposta deste programa consiste, *grosso modo,* em elaborar um esquema que consiga explicar a complexa relação mente/mundo sem a necessidade de recorrer à noção de representação mental. Tal proposta, segundo os autores mencionados, fornece um novo conjunto de idéias para descrever o comportamento de sistemas

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

inteligentes, pois o termo sistema, neste programa, deve ser entendido como a permanente relação entre sujeito e objeto.

Na perspectiva da ciência cognitiva dinâmica, a mente e o comportamento inteligente em geral (humano ou não) deve ser explicado no seu processo de interação com o mundo. Em outros termos, mente e mundo constituem um grande sistema, que deve ser descrito em termos de equações não lineares, do tipo y = xz, onde, como diz Thagard (1998; p.170) y depende da interação dos valores x e z. Com este modelo explicativo espera-se ganhar muito na explicação dos aspectos não representacionais do conhecimento do comportamento humano. Por exemplo, acredita-se que mudanças repentinas de comportamento, do tipo: a mudança de um estado emocional de uma criança que em um momento chora e no momento seguinte sorri, não pode ser explicada em termos de representações mentais e regras. Segundo os cognitivistas do *programa dinâmico* a explicação dessa mudança de humor deve ser descrita em termos de equações que explicitam as variáveis envolvidas nesta mudança brusca.

Para os pesquisadores envolvidos na abordagem dinâmica do comportamento não devemos construir um teoria explicativa da nossa conduta apoiados apenas em um conjunto de representações físicas (como pretendem os pesquisadores da RNA) ou simbólico/abstrata (como sugere a IA), mas seguir o exemplo da física e da biologia e desenvolver padrões explicativos para descrever como a mente se transforma (adquiri e modifica hábitos) no tempo. Para que este padrão explicativo, que procura incorporar toda a complexidade que se instaura na relação mente e mundo, seja aceito, acreditamos que se faz necessário recorrer à uma outra espécie de paradigma científico. Um paradigma que incorpore a noção

de complexidade pois, segundo Morin (1996) só um paradigma como este permite uma intelecção mais completa do real.

Esta concepção de paradigma da complexidade é defendida por Morin em seu livro *Ciência com Consciência*. Segundo Morin, para se desvendar os mistérios da vida deve-se abordar os fenômenos a partir de vários planos de observação. Estes são ilustrados através das seguinte passagem:

> À primeira vista, o céu estrelado impressiona por sua desordem: um amontoado de estrelas, dispersas ao acaso. Mas, ao olhar mais atento, aparece a ordem cósmica, imperturbável — cada noite, aparentemente desde sempre e para sempre, o mesmo céu estrelado, cada estrela no seu lugar, cada planeta realizando seu ciclo impecável. Mas vem um terceiro olhar: vem pela injeção de nova e formidável desordem nessa ordem; vemos um universo em expansão, em dispersão, as estrelas nascem, explodem, morrem. Esse terceiro olhar exige que concebamos conjuntamente a ordem e a desordem; é necessária a binocularidade mental, uma vez que vemos um universo que se organiza desintegrando-se. (Morin. 1996;p.195)

Ainda não sabemos até que ponto a introdução desses 'três olhares', que incorporam a ordem e a desordem, bem como a complexidade, podem nos ajudar na explicação do real. Porém, acreditamos que alguns cientistas cognitivos já caminham rumo à um pensamento capaz de lidar com a ambivalência do real ao aceitar que o pensamento cientifico deve lidar com o principio da incompletude, da incerteza e da desordem, se quiser dar uma explicação não simplista do mental.

Para Morin, basta apenas não esquecermos que o homem é um ser biológico, social e cultural, e que os fenômenos sociais são, concomitantemente, econômicos, culturais e psicológicos, para percebermos a necessidade de aspirar à multidimensionalidade que só um pensamento complexo permite atingir.

unesp

No entanto, diz Morin (1996; p.176), a ciência tem fugido ao desafio da complexidade por dois mal-entendidos:

> O primeiro mal-entendido consiste em conceber a complexidade como receita, como resposta, em vez de considerá-la como desafio e como uma motivação para pensar. O segundo mal-entendido consiste em confundir a complexidade com a completude. Acontece que o problema da complexidade não é o da completude, mas o da incompletude do conhecimento.

Devido a esses dois mal-entendidos é que a ciência, segundo Morin, se preocupou muito mais em edificar um método simplificador, fugindo assim das armadilhas e sutilezas preparadas pela complexidade. Para o pensamento simplificador deve-se separar os diferentes aspectos da realidade, ou englobá-los por uma redução mutilante.

Acreditando fugir da redução mutilante do pensamento simplificador, Morin apresenta doze princípios que constituem, juntos, um paradigma da complexidade. Não iremos enumerar os doze princípios, apontaremos apenas alguns deles que acreditamos fundamentais para a proposta atual da ciência cognitiva de explicar a natureza da mente:

> Principio de distinção, mas não de separação, entre o objeto ou o ser e seu meio ambiente.
> Possibilidade, a partir de uma teoria da auto-produção e da auto-organização, de introduzir e de reconhecer física e biologicamente (e sobretudo antropologicamente) as categorias do ser e da existência.
> Possibilidade, a partir de uma teoria da auto-produção e da auto-organização, de reconhecer cientificamente a noção de autonomia.
> Princípio de causalidade complexa, comportando causalidade mutua inter-relacionada... (Morin 1996; p.332-33)

Estes são apenas alguns princípios que estabeleceriam, segundo Morin, um paradigma da complexidade. Acreditamos que cada vez mais os fenômenos de auto-organização, autonomia e relação entre o ser e o ambiente, vêm interessando as linhas de pesquisas existentes na ciência cognitiva. Julgamos

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

que, principalmente, os conexionistas estão envolvidos com esses princípios na construção de seus modelos que pretendem simular os processos cognitivos. Pois, como mencionamos anteriormente, os modelos conexionistas fundamentam-se na noção de 'auto-organização' que permite a formação de representações ou padrões de conectividade entre nódulos através do processo da interação das nódulos com o seu meio.

Contudo, esse paradigma da complexidade que se apresenta como o paradigma científico adequado para a explicação dos estados mentais, segundo Morin, deve ser entendido como algo que:

> ...não 'produz' nem 'determina' a inteligibilidade. Pode somente incitar a estratégia/inteligência do sujeito pesquisador a considerar a complexidade da questão estudada. (Morin 1996;p.334)

Ao nosso ver, a proposta da análise dos estados mentais através do paradigma da complexidade parece, apesar do entusiasmo de alguns de seus adeptos, ainda pouco elaborada para fornecer subsídios teóricos confiáveis aos cientistas cognitivos. Contudo, ela representa uma retomada do ideal cibernético que originou a ciência cognitiva.

Em resumo, neste capítulo traçamos um histórico do surgimento da ciência cognitiva. Esse histórico percorreu a caracterização dessa nova ciência enfocando: a) os seus pressupostos básicos, a saber, a interdisciplinaridade e o emprego de máquinas (computadores ou robôs) na elaboração e teste de modelos explicativos da atividade cognitiva, b) a formação deste novo campo de pesquisa e c) os fortes vínculos existentes entre este campo de pesquisa e a cibernética.

Iniciaremos, no próximo capítulo, uma discussão sobre uma antiga polêmica existente em torno da IA, qual seja: Teria esta vertente da ciência

unesp

cognitiva herança na filosofia cartesiana? Será o debate em torno desta questão o fio condutor do próximo capítulo.

CAPÍTULO 2

TERIA A CIÊNCIA COGNITIVA, EM SUA VERTENTE DENOMINADA IA, RAÍZES NA FILOSOFIA CARTESIANA?

CAPÍTULO 2

TERIA A CIÊNCIA COGNITIVA, EM SUA VERTENTE DENOMINADA IA, RAÍZES NA FILOSOFIA CARTESIANA?

INTRODUÇÃO

A necessidade de uma ciência da cognição realmente interdisciplinar surge, como explicitamos no capítulo anterior, da complexidade dos eventos mentais. Os cientistas cognitivos acreditam que somente uma ciência que agrupe diversas áreas do saber, apoiada em uma única metodologia (ainda que multifacetada), teria condições de abordar eficientemente os diversos aspectos da cognição. Essa 'tendência naturalizante' da ciência cognitiva, herdada da epistemologia analítico-positivista, é uma 'herança cultural' dessa jovem ciência. Uma vez que esse novo campo do saber está profundamente enraizado na mentalidade científica e tecnológica do século XX, seu vocabulário e modelos explicativos fundam-se, principalmente, no universo computacional.

Não se sabe, ainda, qual das vertentes da ciência cognitiva oferece o melhor modelo para a explicação dos processos mentais e, em conseqüência, a melhor solução para os problemas com os quais se defronta esta jovem ciência na sua pretensão de simular alguns processos cognitivos.

O objetivo deste capítulo será, portanto, iniciar as considerações à respeito das duas principais linhas de pesquisa da ciência cognitiva. Tais considerações terão início com a polêmica questão que se estabelece no âmbito da IA: teria a ciência cognitiva, em sua vertente denominada IA, raízes na filosofia cartesiana? Iniciaremos nossa discussão, na seção 2.1, explicitando a aproximação que alguns autores, dentre eles Penrose (1993) e Searle (1987) supõem existir no trabalho desenvolvido pela IA e a filosofia cartesiana. De modo a explicitar esta aproximação iremos resumir, na seção 2.2, alguns elementos da filosofia cartesiana, no que diz respeito à natureza do conhecimento, para decidirmos se é

unesp

procedente a aproximação sugerida. Cumprida essa etapa, apresentaremos, na seção 2.3, a disputa que se estabelece entre Descartes e seus críticos, acerca da natureza do conhecimento. Desse confronto entre Descartes e seus críticos, tentaremos encontrar elementos que, poderiam, hoje em dia, ser usados pelas vertentes existentes na ciência cognitiva e este é o segundo objetivo deste capítulo.

2.1 A inteligência artificial e o dualismo mente /corpo

Como mencionamos na introdução, a ciência cognitiva abrigava, até recentemente, duas vertentes de pesquisa distintas: a IA e RNA. Para nós, a diferença fundamental entre essas vertentes reside na caracterização que cada uma faz acerca do comportamento inteligente na construção de modelos que pretendem simular atividades cognitivas. Contudo, ambas enfatizam o elemento representacional do conhecimento.

Enquanto os conexionistas acreditam que uma boa caracterização do conhecimento deve levar em conta as leis físicas que atuam na micro estrutura dos sistemas físico-biológicos que processam informação, os adeptos da IA apostam, apenas, em regras abstratas que servem para estruturar o conhecimento. Essa posição dos pesquisadores da IA fica clara nessa passagem:

> ...nós e outros pesquisadores da IA sustentamos que se podem construir entidades inteligentes analisando em que consiste o comportamento inteligente, determinando as regras que governam esse comportamento e implementando tais regras em uma máquina. (Schank, Birnbaum 1996, p.78)

Podemos notar, através das palavras de Schank e Birnbaum, que

para os pesquisadores da IA um comportamento inteligente é resultado de uma série de regras que, por exemplo, uma pessoa segue quando executa uma determinada ação. Esses pesquisadores acreditam que, se pudermos explicitar muito bem as regras que determinam o comportamento inteligente e escrevê-las em termos de um algoritmo, poderemos implementá-las em um computador digital. O mais importante na explicação da inteligência, segundo os pesquisadores da IA, é o algoritmo que contém as regras que possibilitam o comportamento inteligente e não onde esse algoritmo é realizado.

Afirmamos, no capítulo anterior, que essa posição defendida pelos pesquisadores da IA, segundo a qual o que conta para uma simulação de um determinado comportamento inteligente é apenas o conjunto de regras contidas no algoritmo, e não o fato de que este está sendo realizado por um cérebro, computador eletrônico ou por um sistema de canos e barbantes foi, inicialmente, formulada pelo filósofo Putnam (1980). Tal hipótese recebeu o nome de 'instanciação múltipla' ou funcionalismo computacional.

Para autores como Searle (1987) e. Penrose (1993), entre outros, a posição defendida pela IA de uma caracterização do conhecimento em termos apenas de regras abstratas contidas no algoritmo, guarda semelhanças com uma clássica posição filosófica, a saber, o dualismo. Neste sentido, comenta Penrose (1993, p.21):

> A idéia é que simplesmente a estrutura lógica do algoritmo é que tem significado para o 'estado mental' que deve representar, sendo totalmente irrelevante a materialização física particular desse algoritmo... isso encerra na realidade uma forma de 'dualismo'.

A observação acerca do 'dualismo' que se encerra nas teses

defendidas pelos adeptos da IA é comentada, também, por Searle (1987, p.49):

> Os partidários da IA crêem que a mente é mais do que uma parte do mundo biológico natural, pensam que a mente é especificável em termos puramente formais...na realidade, toda a tese da IA se baseia numa espécie de dualismo...

Mas que dualismo é este a que se referem Penrose e Searle? Como entender a sugestão destes autores se a literatura da IA está repleta de ataques à toda postura filosófica que encerra uma forma de dualismo? De modo a entender tal controvérsia discutiremos, em seguida, o dualismo estabelecido pelo filósofo do século XVII Descartes.

2.2 A atualidade da filosofia cartesiana

Sugerimos na seção anterior que uma possível forma de conexão entre a IA e a filosofia cartesiana reside no fato de que ambas as posturas admitem uma forma de dualismo, apesar dos seguidores da IA condenarem qualquer adesão à postura dualista. Além disso, como ressalta Gardner (1995), Descartes pode ser chamado de 'antecessor filosófico prototípico' da ciência cognitiva pelo seu profundo interesse em desvendar também o funcionamento do aparelho sensorial e dos processos cerebrais:

> Escrevendo bem no começo da era moderna (a qual em parte, ajudou a definir e a lançar), ele é em alguns aspectos uma volta àquela fé confiante nas próprias intuições, aquele compromisso com a importância do pensamento matemático e àquela crença nas idéias inatas que nós associamos aos gregos. Ao mesmo tempo, em sua curiosidade a respeito do funcionamento do sistema sensorial, da natureza dos processos cerebrais e do autômato como um modelo possível da natureza, ele é virtualmente uma figura contemporânea. (Gardner 1995, p.64)

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

Esta passagem nos apresenta o cenário da filosofia cartesiana. Cenário este que revela uma certa tensão ao compactuar com idéias clássicas como imortalidade da alma, idéias inatas, distinção mente/corpo, sem, entretanto, negar a grande possibilidade que a ciência nos abre para investigar as questões relativas à natureza corpórea do homem. Por integrar em sua filosofia o universo das máquinas (ou seja, as investigações sobre as funções do corpo humano que, na sua concepção, não passa de um autômato), Descartes é agora convidado a dialogar (num experimento mental) com as mais diversas ciências, incluindo nessa lista, a ciência cognitiva.

Seja pela possibilidade de vincular a filosofia cartesiana à IA através de uma postura dualista, seja pela permanente atualidade da sua filosofia frente aos avanços das mais diversas áreas do saber, julgamos oportuno apresentar alguns elementos do sistema edificado por Descartes, sem a pretensão de examina-lo detalhadamente. A análise aqui desenvolvida apresentará um recorte das idéias de Descartes que nos parecem relevantes para a discussão do seu possível parentesco com a IA.

O primeiro ponto a ressaltar é o fato que Descartes adotou, como filósofo, uma concepção dualista acerca do ser humano. Ele afirma que todo ser humano é composto por um espírito e um corpo, que são de natureza distintas e possuem características particulares. Enquanto o espírito é imortal, imaterial e não sujeito às leis da física, o corpo é mortal, material e está sujeito às leis da física.

Segundo a ordem das razões, o espírito é o elemento central da filosofia cartesiana. Dele nascem e para ele convergem os problemas relativos à

natureza do conhecimento. O corpo humano é comandado por uma natureza superior que constitui a sua verdadeira essência, a saber, o espírito. Mas, como Descartes conclui que a verdadeira essência do ser humano reside no espírito? Será o percurso que leva a uma resposta a esta questão que trilharemos a seguir.

2.2.1 O espírito como elemento central da filosofia cartesiana

Descartes em suas *Meditações* lança um método de dúvida radical através do qual pretendia provar que: existe um fundamento de onde se pode partir para construir o edifício do conhecimento ou, alternativamente, que não existe nada de seguro que possa ser conhecido. Partindo do princípio segundo o qual o conhecimento verdadeiro resulta da experiência sensível

O primeiro passo da dúvida cartesiana é o argumento do erro dos sentidos. Este serve para pôr na berlinda as nossas percepções sensíveis pois, essas podem, muitas vezes, nos levar a erros. Segundo o argumento formulado por Descartes, para se conhecer uma verdade clara e distinta, tudo que possa nos levar a erros deve ser descartado:

> Tudo que recebi, até presentemente, como o mais verdadeiro e seguro, aprendi-o dos sentidos ou pelos sentidos: ora, experimentei algumas vezes que esses sentidos eram enganosos, e é de prudência nunca se fiar inteiramente em quem já nos enganou uma vez. (Descartes 1973, p. 93-94)

No entanto, esse primeiro passo da dúvida sistemática e radical colocava na berlinda apenas as percepções sensíveis. Torna-se necessário verificar se as idéias acerca da natureza corpórea, de fato, são claras e distintas. Para tanto,

Descartes usa o argumento dos sonhos, na tentativa de mostrar que não há evidência alguma que permita diferenciar nossas crenças na vigília ou nos sonhos:

> ...lembro-me de ter sido muitas vezes enganado, quando dormia... E, detendo-me neste pensamento, vejo tão manifestamente que não há quaisquer indícios concludentes, nem marcas assaz certas por onde se possa destinguir nitidamente a vigília do sono... (Descates 1973,p.94)

A conclusão extraída dos argumentos dos sentidos e dos sonhos descartam aquelas idéias/crenças referentes à natureza corpórea e às percepções sensíveis. Restam, para Descartes, as verdades da ordem intelectual como, por exemplo, aquelas extraídas das propriedades matemáticas. Para atacar as 'verdades intelectuais' ele lança mão do argumento do Gênio Maligno ou do Deus enganador que sempre poderia enganá-lo mesmo no domínio da matemática:

> ...pode ocorrer que Deus tenha desejado que eu me engane todas as vezes em que faço a adição de dois mais três, ou em que enumero os lados de um quadrado, ou quando julgo alguma coisa ainda mais fácil, se é que se pode imaginar algo mais fácil do que isso. (Descartes 1973, p.95)

Após ter desenvolvido esses três argumentos nada de seguro parece restar para Descartes no seu processo de busca de verdades claras e distintas. Tudo o que, até então, acreditava ser verdadeiro fica colocado em dúvida, a não ser o ato de duvidar (marca do pensar) pois, quer seja enganado, ou não, o ato de duvidar permanece. A dúvida lhe dá evidências acerca de sua existência, enquanto ser pensante:

> Eu sou, eu existo; isto é certo; mas por quanto tempo? A saber, por todo tempo em que penso... nada sou, pois, falando precisamente, se não uma coisa que pensa, isto é, um espírito, um entendimento ou uma razão... (Descartes 1973, p.102)

Será portanto a partir da constatação de sua existência, enquanto ser pensante, (sua primeira verdade clara e distinta) que Descartes começará a

construir o edifício do conhecimento. O corpo não é necessário para a caracterização da sua existência enquanto sujeito do conhecimento, pois este é assegurado por uma substância imaterial e pelos processos que nela ocorrem. Como diz Descartes:

> ...noto que não pertence necessariamente nenhuma outra coisa à minha natureza ou à minha essência, a não ser que sou uma coisa que pensa, concluo efetivamente que minha essência consiste somente em que sou uma coisa que pensa uma substância da qual toda a essência ou natureza consiste apenas em pensar. (Descartes 1973, p. 142)

Neste contexto, Descartes situa a sua análise da natureza das idéias no plano puramente abstrato. Para nós, um possível ponto de convergência entre a filosofia cartesiana e aquela subjancente à IA, reside, justamente, no enfoque abstrato que deixa de lado os elementos corpóreos possivelmente presentes na elaboração do conhecimento. Devemos lembrar que não só a IA compartilha este enfoque acerca do conhecimento mas também uma vertente em filosofia da mente fundada na década de 60 pelo filósofo Putnam (1980) denominada *funcionalismo computacional*.

Como vimos, a pesquisa desenvolvida pela vertente funcionalista computacional funda suas idéias no conceito de organização funcional. Isto é, para estudar os processos mentais devemos enfocar as estruturas funcionais que se formam em um organismo frente a um dado de entrada e uma resposta. Segundo Putnam, o advento do computador contribuiu significativamente para o desenvolvimento da idéia de organização funcional. O computador, na abordagem epistemológica funcionalista computacional, serve como modelo de um organismo que apresentaria uma certa organização funcional que é explicitada por um programa encarregado de manipular o *input* recebido e

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

devolver um *output*. A organização funcional de um organismo, portanto, pode ser descrita como um programa que, frente a um dado de entrada, produz uma resposta.

Putnam, com sua idéia funcionalista, procura explicitar uma tese, que mais tarde será adotado pelos adeptos da IA, segunda a qual o pensamento pode ocorrer não apenas em seres humanos ou entidades com estrutura cerebral semelhante à nossa mas, em sistemas gerais com capacidade de processar informações estruturadas. O que subjaz a esta tese funcionalista computacional é a hipótese que o pensamento pode ser estudado apenas a partir das suas estruturas abstratas funcionais. Esta tese afasta-se das concepções segundo as quais o pensamento só pode ocorrer em um aparato físico/biológico semelhante ao do homem.

Para Putnam, a matéria de que é constituído um sistema não importa ao projeto de simulação de determinadas atividades cognitivas. Importa, sim, a investigação sobre como se estrutura a sua organização funcional, contida no software. Desta forma, o enfoque do pensamento a partir da perspectiva funcional deixa um legado que será seguido pelos pesquisadores da IA que, em certo sentido, tem um parentesco com a abordagem cartesiana das idéias. Os modelos computacionais propostos pela IA, seguindo os pressupostos do funcionalismo, também 'desprezam', na análise da estrutura lógica do pensamento, a matéria física. A implementação feita pela IA baseia-se, principalmente, nas regras abstratas que possibilitam ao computador digital comportar-se de maneira inteligente.

A forma de explicação dos processos mentais puramente



mecanicista, proposta pelos adeptos da IA e do funcionalismo computacional, gera um paradoxo que pode ser resumido da seguinte maneira: ainda que ambas as formas de pensamento possam convergir para um dualismo, Descartes que defende, como mostramos, a tese da natureza mecânica dos corpos, jamais aceitaria, como pretendem os adeptos da IA, uma explicação puramente mecanicista da mente. Em outras palavras, o possível dualismo da IA não seria um dualismo da substância, mas sim um dualismo de propriedades.

Além disso, para Descartes, Deus concebeu a natureza humana como que dotada de livre-arbítrio. Lembramos que é o espirito, substância imaterial (não regida por leis mecânicas) que constitui a essência da natureza humana pensante. Sendo assim, os modelos computacionais gerados, principalmente, pela IA não podem, segundo os cartesianos, ser considerados legítimos herdeiros de uma filosofia que rejeita a hipótese mecanicista da mente, pois esta não deixa espaço para a livre atuação do espirito. Dessa forma, julgamos que Descartes seria pessimista em relação ao projeto da ciência cognitiva.

De modo a ressaltar a posição anti-mecanicista de Descartes no domínio da mente, apresentaremos, a seguir, aspectos de uma possível disputa entre Descartes e seus críticos, acerca da natureza do mental. Disputa essa que pode fornecer elementos importantes para a pesquisa acerca da cognição desenvolvida, principalmente, pela vertente conexionista.

2.3 Críticas ao dualismo cartesiano

Na seção anterior começamos articulando um possível elo entre

as pesquisas desenvolvidas pela IA e a filosofia cartesiana. Apontamos que, segundo autores como Penrose e Searle, a IA encerra, como a filosofia cartesiana, uma forma de dualismo. O dualismo é um ponto de vista adotado por Descartes segundo o qual existem dois tipos distintos de substâncias: a mental e a física. Sendo que a substância mental não seria composta de matéria e poderia existir independente dessa.

Em contrapartida, o dualismo que se encontra na IA (se é que ele existe) seria o chamado *dualismo de propriedades*. Para os cientistas desta linha de pesquisa da ciência cognitiva, teríamos uma única substância, a saber, a física. Esta, por sua vez, seria composta por duas propriedades distintas: uma lógica (que toma forma de algoritmo) e uma física (o *hardware* em que se processará o algoritmo). Para eles um algoritmo tem uma existência desmaterializada, sendo totalmente independente do objeto que o realiza. Teríamos, portanto, no esquema seguido pela IA, algoritmos (na caracterização da 'mente') e, distintamente, o meio físico que os realiza (matéria 'física') caracterizando o dualismo de propriedades.

Apresentamos o dualismo de substâncias que se estabelece na filosofia cartesiana, enfatizando que o espírito para Descartes constituiria a verdadeira essência humana do sujeito cognitivo enquanto puro pensamento. Tal ênfase foi necessária para ilustrar uma controvérsia atual: tanto a filosofia subjacente ao projeto da IA, como a filosofia cartesiana, apresentam uma postura dualista. Sendo a primeira uma espécie de dualismo de propriedades e a segunda um dualismo de substância.

A seguir iremos expor algumas críticas ao dualismo cartesiano a

unesp

fim de extrair dessas críticas elementos que podem impulsionar o trabalho desenvolvido na vertente da ciência cognitiva denominada RNA.

2.3.1 As hipóteses de De La Mettrie

A caracterização da mente, segundo a concepção cartesiana, entendida como substância distinta do corpo, encontrou vários adversários. Dentre eles iremos analisar, nesta seção, as hipóteses de um dos primeiros pensadores materialista da história, a saber, De La Mettrie (1982). Esse filósofo e médico francês acreditava que o cérebro deveria ser estudado pormenorizadamente, pois ele próprio poderia revelar vários mistérios que cercam a natureza humana. A substância mental eleita por Descartes como o principal atributo da inteligência pode ser relegada a um segundo plano.

Para De La Mettrie (1982, p. 96):

> ...o pensamento é tão pouco incompatível com a matéria organizada, que ele parece ser uma de suas propriedades, ao lado da eletricidade, da faculdade de movimento e da impenetrabilidade.

De La Mettrie concebe o pensamento como resultado da própria organização da matéria e não, como pensava Descartes, como uma marca do espírito. Essa é uma das razões que o leva a realizar um estudo aprofundado acerca das propriedades emergentes da matéria organizada. Estudo esse que é guiado pela experiência e observação.

Será no tratado "*O Homem Máquina*" que De La Mettrie exercitará sua observação e experiência médica para desvendar as influências que o corpo e suas necessidades exercem sobre o espírito, invertendo, assim, o vetor

de análise proposto por Descartes. Ele enfatiza que existem algumas evidências que demonstram claramente as influências do corpo e suas necessidades sobre a alma, como por exemplo, sem alimento o corpo se torna fraco e doente e, por conseguinte, a alma enfraquece e perde sua vitalidade.

O tratado de De La Mettrie torna-se mais profundo quando ele realiza um estudo de anatomia comparada entre animais e homens. Seu estudo levanta pontos interessantes, dentre eles, a seguinte tese: a diferença básica entre o homem e os animais reside no tamanho e na organização do cérebro. A partir dessa diferença básica pode-se extrair, com a ajuda da observação e comparação, três conseqüências:

> 1- Quanto mais ferozes são os animais menos cérebro têm.
> 2- Que esta víscera parece aumentar, de algum modo, em proporção com a docilidade.
> 3- Que aqui se dá uma singular conclusão imposta eternamente pela natureza, segundo a qual, quanto mais se ganha pelo lado do espírito, mais se perde pelo lado do instinto. (De La Mettrie 1982, p. 216)

De La Mettrie quer demonstrar, através das três conseqüências acima mencionadas, o quanto a organização do cérebro se reflete diretamente nos comportamentos animal e humano. Para ele, Descartes equivocou-se ao atribuir inteligência apenas ao comportamento humano que resulta das regras contidas no espírito. Descartes consentiria que o comportamento animal resulte da organização cerebral afinal, para ele, os animais são desprovidos de alma. Contudo, ele não admitiria qualquer forma de pensamento nestes animais/máquinas, pois Descartes jamais aceitaria que comportar-se inteligentemente é apenas uma questão de maior ou menor grau de organização cerebral. Um julgamento desse tipo colocaria em segundo plano (ou

desconsideraria) o papel da 'alma' e sua influência no comportamento humano.

Discordando novamente de Descartes, De La Mettrie acredita que a diferença que separa os homens dos animais não é uma diferença intransponível do tipo ter ou não ter uma 'alma'. Na sua concepção, o homem não era tão diferente do macaco antes de dominar a linguagem. Por sua vez, o domínio da linguagem, das palavras e símbolos só foi possível graças aos processos mecânicos cerebrais e à mecânica da nossa educação. Por processo mecânico de educação podemos entender um processo que consiste no ensino, a partir de regras bem especificadas, de uma determinada tarefa a um aluno que poderá, mais tarde, ensinar esta mesma tarefa a uma outra pessoa, e assim por diante, seguindo as regras mecânicas pré-estabelecidas.

Sendo assim, De La Mettrie conclui que não é um absurdo pensar que macacos, os animais que mais se assemelham ao homem em suas disposições e organização, poderiam algum dia manipular a linguagem. Foi pensando nessa hipótese que ele contratou um professor de surdo-mudo para ensinar um orangotango a falar. Porém, a sua tentativa não colheu grandes resultados na época.

Mesmo admitindo que a educação seja um processo mecânico, poderíamos levantar a seguinte questão: se a linguagem foi transmitida de um homem para o outro, quem teria sido o primeiro homem a dominar a linguagem? A esta pergunta De La Mettrie responde: "Eu ignoro. Devo crer que os melhores organizados, aqueles a que a natureza concedeu seus favores devem ter instruído os outros." (De La Mettrie 1982;p. 219-220).

Argumenta ainda De La Mettrie que o problema da origem, ou

seja, quem foi primeiro, dificilmente encontrará uma resposta objetiva. Em contraste, o processo de aquisição da linguagem pode ser caracterizado como um processo mecânico e, sendo assim, pode ser transmitido aos animais. Ele deixa como sugestão uma pista para a questão da origem da linguagem: alguns homens nasceram com o cérebro melhor organizado e, por isso, conseguiram o domínio da linguagem, elemento este que foi primordial para a espécie humana desenvolver-se.

Para De La Mettrie, a alma ou a mente é um produto da organização que recebemos da natureza. Ela é a verdadeira causa da nossa capacidade de, por exemplo, adquirir *disposições* para nos tornarmos sábios e virtuosos. Essas *disposições* adquiridas graças, principalmente, à organização do cérebro provém da natureza, porém, ela deve ser lapidada pela instrução para se tornar um bem. Neste sentido, afirma De La Mettrie: "Se a organização é um mérito e a fonte dos demais, a instrução é o segundo." (De La Mettrie 1982;p. 223).

De forma pioneira, De La Mettrie, no século XVIII, põe no centro dos debates o cérebro e ressalta o quanto esse órgão material é importante na caracterização do raciocínio inteligente, afastando a hipótese cartesiana da necessidade de um espírito distinto do corpo que possibilitaria a aquisição do conhecimento. Basta, segundo esse autor, entender/caracterizar o pensamento, a inteligência, como propriedades que emergem do cérebro bem organizado:

> ...todas as faculdades da alma dependem a tal ponto da organização do cérebro e de todo o corpo que estas se confundem visivelmente com a organização mesma. (De La Mettrie 1982;p. 235)

De La Mettrie descarta o dualismo cartesiano e aposta num

estudo pormenorizado do cérebro para compreendermos com mais profundidade a natureza da mente humana. A seguir apresentaremos uma outra crítica ao dualismo cartesiano. Tal crítica pode fornecer alguns elementos para as pesquisas desenvolvidas na ciência cognitiva sobre a natureza do comportamento inteligente.

2.3.2 Ryle e a análise dos estados mentais a partir dos estados disposicionais.

A próxima crítica que comentaremos em relação ao dualismo cartesiano foi elaborada pelo filósofo analítico Ryle. A abordagem da mente feita por Ryle, em seu trabalho *The Concept of Mind* (1949), caracteriza-se por uma tentativa de dissolver o dualismo mente/corpo, através de uma análise dos argumentos empregados por Descartes nas *Meditações*. A sua dissolução do dualismo, no entanto, não implica em uma postura eliminativista dos estados mentais.

Ryle não irá negar os fenômenos mentais, apenas, como ele avisa no prefácio de *The Concept of Mind*, determinará a exata 'geografia lógica' das proposições deste fenômenos. Em outros termos, ele pretende, através de uma severa análise da linguagem, demostrar que o problema mente/corpo é um pseudo-problema, resultado de confusões lógico-lingüísticas que Descartes e os cartesianos cometeram.

Essa tentativa de dissolver o problema mente/corpo, através de uma severa análise da linguagem, é muitas vezes conhecida como behaviorismo

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

lógico. Ryle, entretanto recusa insistentemente tal classificação para a sua análise.[2]

Ryle afirma que não pretende apresentar um novo conceito de mente, mas apenas caracterizá-lo precisamente. Para ele, o senso comum sabe muito bem manejar este conceito; as confusões conceituais ocorrem apenas no domínio filosófico da análise. Foram os filósofos que cometeram deslizes terminológicos e acabaram obscurecendo este conceito. Ciente de estar indo contra toda uma tradição filosófica, ele adverte:

> Talvez os conforte saber que as posições contra as quais mostro mais calor são posições de que eu próprio fui vítima. Só secundariamente espero ajudar outros teóricos a reconhecer a nossa doença e a beneficiar-se da minha receita. (Ryle 1949;p.9)

Sabendo que sua tentativa será recebida com muitas restrições e preconceitos, Ryle procura deixar claro o ponto contra o qual ele se volta, assumindo que, muitas vezes, esboça a teoria cartesiana com certos exageros. Tais exageros tem como finalidade chamar a atenção para os pontos confusos da doutrina cartesiana a ser combatida.

Dada a grande aceitação da filosofia cartesiana, Ryle a chama de

---

[2] Mas, o que teria em comum a análise de Ryle com a proposta dos behaviorista lógicos? Para muitos teóricos (p.ex. Fodor e Searle), Ryle, ao tentar analisar o mental em termos de disposições comportamentais estaria, mesmo de maneira sofisticada, aderindo ao projeto formulado por Carnap, a saber, analisar um certo fenômeno, por exemplo a inteligência, em termos de disposições comportamentais. Ou seja, falar de inteligência não implica em desvendar as causas ocultas e internas que passam na mente e que resultam em uma maneira de se comportar. Basta apenas, segundo os behavioristas lógicos, analisar alguém que se comporta de uma certa maneira e daí concluir se o ato foi ou não praticado com inteligência. Não cabe aqui uma defesa a Ryle. Porém gostaria de afirmar que discordamos daqueles que rotulam Ryle como behavioristas por dois motivos: a) Ryle jamais negou a existência da mente e de fenômenos mentais; b) muitos teóricos não analisam com atenção o que é uma disposição. Uma disposição não é e nem pode ser analisada em termos do comportamento imediato apresentado pelo agente. Uma disposição é o que possibilita ou não um comportamento. Sendo assim, a análise de Ryle não estaria voltada para os comportamentos que revelam se uma ação foi ou não praticada com inteligência. Mas sim, para as disposições que possibilitaram a ação. Porém tais disposições não possuem o

Doutrina Oficial e a resume na hipótese que todo ser humano possui uma mente e um corpo, ambos de naturezas distintas e características particulares, que podem ser resumidas através das seguintes proposições:

> 1- Todo corpo existe no espaço e no tempo. As mentes apenas no tempo;
> 2- O corpo é constituído por matéria, a mente por consciência ou funções da consciência;
> 3- O corpo, por ser algo material é sujeito às leis físicas, porém as mentes, sendo imateriais, não sofrem tal sujeição;
> 4- Os acontecimentos corporais são observáveis por todos, por isso carregam o estatuto de serem públicos. Os acontecimentos mentais ocorreram na própria mente e por isso são chamados privados. (Ryle 1949;p.11-12)

Esses quatros pontos sintetizam, *grosso modo*, a caracterização que a Doutrina Oficial apresenta, segundo Ryle, acerca da natureza da mente e do corpo. Tal caracterização irá legar ao ser humano a existência de 'dois mundos'. O primeiro, o mundo físico, habitado pelos corpos e do qual temos apenas um conhecimento parcial. O segundo, o mundo mental, em que a mente se situa e que, graças à consciência, autoconsciência e introspeção, pode ser passível de conhecimento.

A consciência, autoconsciência e a introspeção nos revelam, segundo a doutrina cartesiana, tudo o que ocorre no *mundo mental* ou *interior*. Porém, devemos saber que a expressão *mundo mental* diz respeito apenas ao mundo mental de cada indivíduo. A introspeção permite ao ser humano ter acesso apenas ao seu próprio mundo mental. Para os críticos a doutrina cartesiana nos leva a seguinte situação: nenhum ser humano pode *conhecer* o que se passa no interior da mente do outro.

No limite, a conseqüência do solipsismo que se instaura na

---

estatuto de uma alma cartesiana. Basta pensarmos que o vidro tem disposição para quebrar ninguém em sã consciência afirma que devido a essa disposição os vidros tem alma.

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

doutrina cartesiana leva a um resultado anti-intuitivo: nenhum comentário à respeito do comportamento de outrem é completamente seguro. Porém, afirma Ryle (1949;p.13):

> Pelo fato de que todos nós fazemos diariamente comentários ou damos palpites à respeito do que se passa na mente das pessoas é que resolvi refletir acerca da geografia lógica dos conceitos de comportamento mental, buscando, assim, demonstrar que a geografia lógica praticada oficialmente está incorreta.

O caráter anti-intuitivo de que se reveste a Doutrina Oficial leva Ryle a examinar seus fundamentos. Para ele esta se assenta sobre um falso princípio. Esse falso princípio provém de um erro de gênero especial chamado de *erro categorial*. O erro categorial consiste em considerar fatos da vida mental como pertencentes a uma determinada categoria lógica, quando na verdade pertencem à outra. A intenção de Ryle não é, como já afirmamos, negar a existência de fenômenos mentais. Mas, demostrar que a postulação de uma entidade abstrata como sede desses fatos decorre de um engano lógico.

A argumentação de Ryle será voltada à descoberta, na Doutrina Oficial, de erros categoriais para, então, retificá-los. Para ele, a origem deste erro está na tentativa que Descartes e os cartesianos fizeram para explicar a mente e os fenômenos mentais em uma linguagem extraída das ciência naturais.

Ryle sugere que Descartes talvez tivesse raciocinado da seguinte maneira: se os fenômenos físicos e fisiológicos são descritos por uma linguagem científica precisa e exclusiva da física e da fisiologia, os fenômenos mentais também deveriam ser descritos por uma linguagem que fosse igualmente precisa e científica mas que, em contrapartida, fosse adequada para explicar unicamente os fenômenos que ocorrem na substância mental.

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

Guiados por esse raciocínio os cartesianos, segundo Ryle, formularam uma linguagem para-mecânica para descrever a mente ou seja, a mente sendo algo imaterial, privado e não sujeito às leis físicas, deveria ser descrita por uma linguagem que fosse a negação da descrição dos fenômenos físicos. Por exemplo, os corpos ocupam lugar no espaço, a mente não ocupa lugar no espaço, e assim por diante.

Tal esforço, segundo Ryle, criou a seguinte situação:

> ...o corpo humano, tal como qualquer outra parte da matéria, é um campo de causas e efeitos, também a mente deve ser um outro campo de causas e efeitos ainda que (Deus seja louvado) causas e efeitos não mecânicos. (Ryle 1949;p.18-19)

Para dissolver esta situação sobre a qual se estrutura a Doutrina Oficial, Ryle procura caracterizar os estados mentais sem cair no idealismo, isto é, sem afirmar que só existem os fenômenos mentais. Porém, ele procura evitar também o materialismo, ou seja, afirmar que os estados mentais devem ser reduzidos aos cerebrais. Ambas as tentativas, materialista e idealista, segundo Ryle, são respostas a uma pergunta inadequada.

Tanto o idealismo, diz Ryle, ao negar a relevância do físico na caracterização do mental, como o materialismo, ao tentar reduzir o mental ao físico, no fundo aceitam que existe o problema da relação dualista entre mente e corpo. Na sua concepção, é o 'problema' que deve ser rejeitado na sua formulação mesma. Ele insiste que não há problema algum em afirmar que existem mentes e que existem corpos. O problema está em sustentar que a mente e o corpo são substâncias distintas que possuem formas diferentes de existência.

A análise de Ryle irá enfocar os estados mentais através,

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

principalmente, da *ação*. Sua proposta consiste em explorar uma parcela do conhecimento que a tradição filosófica relega a um segundo plano, qual seja, '*o saber como*' fazer algo. O olhar atento ao '*saber como*' pode revelar muito sobre os fenômenos mentais. Negando o postulado cartesiano segundo o qual a atuação inteligente pressupõe a observância de regras e aplicação de normas metodológicas, ele procura demonstrar que o atributo 'inteligente' deve ser dado às pessoas que aplicam critérios de relevância prática para a condução da própria ação e não para quem *teoriza* antes de agir.

Será no próprio *processo de fazer algo* que podemos notar os critérios que possibilitam a atuação inteligente. Como exemplos desses critérios podemos citar: *a crítica* que uma determinada pessoa faz ao nosso trabalho, *a atenção* na prática de alguma tarefa, *a prática* de retirar do erro uma lição para cada vez agir melhor, etc. Faz-se necessário ressaltar que Ryle não vê nesses critérios regras que precederiam a ação, mas sim atualizações de estados *disposicionais* do organismo.

A grande contribuição de Ryle ao estudo dos estados mentais parece ser a sua insistência em chamar a atenção para a importância das *disposições* ou formas de agir. Porém, foi desta insistência que surgiram as mais fortes objeções a sua teoria: se o ser humano age por disposições, podemos dizer que seu comportamento será muitas vezes determinado. Basta apenas saber como será a atualização e será possível, em princípio, prever o comportamento. Quanto a esta crítica, Ryle responde que as disposições humanas são disposições *complexas*, que possuem muitas formas de atualização. Ter uma disposição complexa não é o mesmo que afirmar uma lei. Ter uma disposição é:

unesp

> Não estar em um estado particular ou sofrer uma mudança particular, é estar sujeito ou ter a possibilidade de estar em um estado particular ou de sofrer uma mudança particular. (Ryle 1949;p.42)

A pretensão de Ryle, ao criticar o dualismo cartesiano, foi retificar o conceito de mente presente nessa doutrina e mostrar que, ao falar dessa substância, não estamos nos referindo a uma entidade fantasma que comanda nosso corpo; mas, sim, a uma disposição físico-biológica.

O livro de Ryle não só despertou interesse das correntes materialistas na filosofia da mente mas, também, a sua caracterização do 'saber como' serve como uma boa sugestão para os modelos de RNA na ciência cognitiva.

Até este momento, apresentamos o pensamento de autores que rejeitam o dualismo cartesiano, mas que não são contemporâneos desta nova ciência da mente. O que diria um autor que rejeitasse o dualismo de substância e fosse contemporâneo deste projeto científico cognitivo?

Para responder esta questão iremos, a seguir, apresentar algumas críticas elaboradas por Searle (1987) (1997) ao problema mente/corpo, bem como às propostas de simulação computacional de certas atividades mentais levada a cabo pelos pesquisadores da IA.

2.3.3 O naturalismo biológico de Searle

Searle (1987) argumenta que sua visão sobre as questões que permeiam o estudo da mente não se compatibiliza com as propostas feitas pelas correntes materialistas em filosofia da mente e nem com as oferecidas pela ciência

cognitiva. O motivo da sua discordância é que as várias correntes existentes na filosofia da mente e na ciência cognitiva, principalmente em sua vertente IA, acabam negando, ou simplesmente não levando em conta em suas pesquisas, a existência de fenômenos mentais importantes como a consciência, intencionalidade, etc.

Para Searle o responsável por esta atitude de negação frente aos estados mentais óbvios como os citados é o comportamento que toda ciência, incluindo nesse quadro a ciência cognitiva e o materialismo em filosofia da mente, têm em nossos dias. A ciência de nossos dias procura fornecer uma visão do mundo completamente objetiva embora essa visão seja, muitas vezes, incompatível com os fatos mais evidentes de nossas experiências.

O desafio que se impõe à ciência cognitiva e às correntes materialistas em filosofia da mente é, segundo Searle, a possibilidade de conciliar nossa visão comum de nós mesmos, como seres conscientes, livres e racionais, num mundo que a ciência nos diz consistir inteiramente de partículas físicas.

Como coloca Searle (1987;p.18):

> Como, por exemplo, pode ser possível que o Mundo contenha apenas partículas físicas inconscientes e, no entanto, que contenha também consciência? Como pode o Universo mecânico conter seres humanos intencionalistas – isto é, seres humanos que podem representar o Mundo para si mesmos?

Esses problemas mencionados acima por Searle levam a questões bem complexas, como por exemplo: será possível a conciliação da visão de um universo mecânico com seres humanos conscientes e intencionais? Assim, ele se questiona se o trabalho desenvolvido pela ciência cognitiva, ao procurar construir modelos computacionais para simular certas atividades cognitivas, nos

dará uma possível imagem da mente.

Como conciliar a concepção de um universo mecânico preservando elementos subjetivos da nossa vida mental? Para Searle, a conciliação da visão de um universo mecânico com aquele dos seres humanos, conscientes e intencionais, à partir da construção de modelos computacionais (que supostamente forneceriam uma possível imagem da mente), não ajuda a solucionar o problema inicial de compreender um mundo objetivo, habitado por seres humanos, dotados de consciência e intencionalidade (aspectos subjetivos). Pois, tanto o dualismo cartesiano, quanto o funcionalismo, na proposta da IA, apresentam problemas.

O problema mente/corpo, que, em sua opinião, decorre da solução cartesiana para conciliar a visão objetiva do universo com nossos estados subjetivos, é um problema simples de resolver. Só não foi resolvido até agora porque todas as propostas de solução continuaram empregando um vocabulário extraído da própria doutrina cartesiana. Isto é, continuaram usando os termos físico e mental com o mesmo sentido de que estes se revestem no sistema cartesiano.

As tentativas anteriores, segundo Searle, admitem a existência da consciência e da intencionalidade, como estados irredutíveis a fenômenos físicos, e se tornam, assim, dualistas de propriedades. Isto é, existem nelas dois planos de descrição dos fenômenos físicos: o primeiro recorrendo ao discurso da neurofisiologia, o segundo remetendo aos termos mentalistas, como estado consciente, intencional, etc. Estes dois planos de descrição estabelecem uma irredutibilidade do mental ao físico, instaurando o seguinte problema: como o físico gera o mental? Se quisermos fugir desse problema admitiremos que a



consciência e a intencionalidade existem, mas devem ser naturalizadas. Isto é, serem reduzidas a fenômenos físicos.

Ficamos, assim, na concepção de Searle, com as seguintes alternativas: podemos ser dualistas ou monistas. Se formos monistas, poderemos optar pelo dualismo de propriedade ou o reducionismo. Se optarmos pelo dualismo de substâncias, teremos que explicar como se dá a relação entre a mente e o corpo. Para Searle, todas essas categorias estão demasiadamente desgastadas. A solução considerada por ele é a seguinte:

> ...os fenômenos mentais são causados por processos neurofisiológicos no cérebro, e são, eles próprios, características do cérebro. Para distinguir esta concepção das muitas outras existente neste campo, chamamos de 'naturalismo biológico'. Os processos e fatos mentais fazem parte da nossa história natural biológica tanto quanto a digestão, a mitose, a meiose ou a secreção enzimática. (Searle 1997;p.8)

Esta solução, não tão óbvia como Searle gostaria que fosse, sugere que, a 'consciência' e a 'inteligência' resultam dos próprios processos cerebrais. O cérebro, dado o seu material de formação e constituição, produziria estados mentais. A solução de Searle, por extensão, distancia-se da tentativa da IA, mas não tanto da RNA na proposta de simular estados mentais em uma máquina.

No que diz respeito a IA , Searle foi sempre um crítico do seu projeto. Para ele, quando os cientistas cognitivos, em especial os pesquisadores da IA, afirmam que o computador adequadamente programado é realmente uma mente, no sentido em que se diz, que os computadores quando realizam um algoritmo estão entendo sua tarefa, eles estão enunciando uma falsa proposição. Seu famoso argumento do quarto chinês procura ilustrar sua descrença acerca das

pesquisas desenvolvidas pela IA.

Searle imagina que está trancado em um quarto fechado onde existe uma série de instruções, em inglês, para manipular símbolos chineses. Alguém lhe envia, através de uma fresta uma história em chinês e várias perguntas sobre essa formuladas em chinês. Nenhuma outra informação é vinda de fora. Ele, com a ajuda das instruções, manipula os símbolos chineses, e envia, pela abertura, as respostas às questões sugeridas. Como as manipulações são realizadas apenas através de um algoritmo, nesse caso o algoritmo de Schank, a sua seqüência final deve ser simplesmente o chinês para 'sim' ou 'não', dando a resposta correta para a pergunta feita em chinês, sobre uma história em chinês. Ele insiste que, segundo os preceitos da IA, o indivíduo que não sabe absolutamente nada de chinês, mas que realizasse corretamente a série de operações, que caracteriza o seu algoritmo, responderia corretamente, como um falante do chinês, as questões propostas. A moral da história que Searle nos conta é que, mesmo tendo respondido as questões, não haveria, por parte do indivíduo, um entendimento da história. Com esse argumento ele procura mostrar que os programas desenvolvidos pela IA, apesar de conter regras que permitem a manipulação e tradução dos símbolos em chinês lidam apenas com o aspectos sintático da linguagem, ignorando por completo as regras semânticas. O entendimento da história só seria possível se os programas da IA conseguissem combinar e manipular as regras sintáticas e semânticas da linguagem.

Searle aponta, assim, para uma dupla limitação do projeto da IA de simular em máquinas o comportamento inteligente. A primeira limitação é que os programas desenvolvidos pela IA lidam apenas com as regras sintáticas da

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

linguagem. A segunda limitação sustenta que 'cérebros de silício' não possuem a constituição necessária para causar fenômenos mentais. Como mencionamos anteriormente, os fenômenos mentais resultam, segundo ele, dos processos neurofisiológicos que se realizam no cérebro (biológico). Esta limitação permite estabelecer o distanciamento entre a concepção de Searle e a dos conexionistas sobre a natureza dos estados mentais.

Em resumo, neste capítulo indicamos como se estabelece o dualismo cartesiano e tentamos demostrar que espécie de dualismo poderia estar presente nas teses da IA. Apresentamos, ainda, algumas críticas ao dualismo de substância que aparece no sistema cartesiano. Sabemos que essa postura hoje em dia não é assumida por muitas vertentes engajadas nos estudos da mente. O motivo que leva a essa aversão generalizada à qualquer forma de dualismo parece ser a incompatibilidade desta com o saber científico atual.

Se quisermos desenvolver um enfoque científico da mente temos que, entre outras coisas, descrever, em linguagem científica, os fenômenos estudados. O dualismo aceita a existência da 'consciência', 'intencionalidade', etc, mesmo admitindo que estes possam escapar de uma descrição puramente científica.

Por essa razão, entre outras, que o filósofo Dennett (1997;p.29) afirma:

> O dualismo (o ponto de vista de que as mentes são composta de uma matéria não física e absolutamente misteriosa) e o vitalismo (o ponto de vista de que as coisas vivas contêm algum tipo de substância física mas igualmente misteriosa – o elâ vital) foram relegados ao monte de lixo da história, junto com a alquimia e a astrologia

Com essa forte crítica de Dennett ao dualismo, encerramos esta

primeira parte da dissertação para, na segunda parte, apresentar uma analise dos modelos computacionais elaborados pela ciência cognitiva, bem como algumas críticas a estes modelos.

PARTE II

ALCANCE E LIMITES DO MECANICISMO NA EXPLICAÇÃO DOS PROCESSOS MENTAIS

PARTE II

ALCANCE E LIMITES DO MECANICISMO NA EXPLICAÇÃO DOS PROCESSOS MENTAIS

CAPÍTULO 3

PERSPECTIVAS MECANICISTAS DO PENSAMENTO: da máquina de Turing à inteligência artificial

INTRODUÇÃO

No capítulo anterior analisamos argumentos estratégicos para responder à questão, que constituiu o fio condutor da discussão passada, a saber: teria a ciência cognitiva raízes na filosofia cartesiana? Inicialmente apresentamos um possível elo de ligação, apontado por autores como Searle (1984) e Penrose (1993), existente entre o programa da Inteligência Artificial e a filosofia cartesiana, qual seja, ambos encerram uma forma de dualismo: a IA um dualismo de propriedades e a filosofia cartesiana um dualismo de substância.

Na perspectiva dos adeptos da IA, conforme mencionado na apresentação do primeiro capítulo, o conhecimento é caracterizado em termos de regras abstratas que estruturam as representações mentais segundo algum tipo de algoritmo. Tal perspectiva pode ser caracterizada como dualista por admitir a independência entre o plano físico (hardware da máquina) e o computacional (o algoritmo). Paralelamente, no caso da filosofia cartesiana, o dualismo se instaura pela distinção entre mente e corpo. Mencionamos, porém, que esta aparente filiação da IA ao cartesianismo é extremamente polêmica, uma vez que a literatura da IA está repleta de ataques a toda postura filosófica que encerre uma forma de dualismo. Além disso, os cartesianos condenam, por outro lado, qualquer forma de explicação dos processos mentais a partir da hipótese mecanicista da mente.

A polêmica questão, teria a ciência cognitiva, em sua vertente denominada IA, raízes na filosofia cartesiana?, propiciou-nos a oportunidade de apresentar alguns elementos da doutrina edificada por Descartes, bem como, algumas críticas a essa doutrina.

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

Neste capítulo iremos primeiramente descrever o modelo da máquina de Turing na seção 3.1. Deste modelo extrairemos a proposta central dos modelos da IA e o fio condutor deste capítulo, qual seja, a caracterização do pensamento enquanto atividade *mecânica*. A partir da hipótese mecanicista da mente proposta por Turing, iremos analisar o jogo de imitação e vincular as conclusões por ele extraídas desse jogo ao programa da IA. Vinculada à tese mecanicista da mente, proposta por Turing e generalizada pela IA, enfocaremos alguns modelos computacionais da IA na seção 3.2. Encerraremos este capítulo expondo as críticas de Searle e Button, Coulter, Lee e Sharrock ao programa da IA.

3.1 A DESCRIÇÃO DA MÁQUINA DE TURING

Iniciaremos esta seção descrevendo o modelo de máquina de Turing. Tal máquina foi concebida pelo matemático inglês Turing na década de 30. Esta foi, inicialmente, idealizada com o intuito de resolver um problema lógico proposto por Hilbert, a saber, o problema da decisão. *Grosso modo*, este problema pode ser caracterizado como: existe um *procedimento efetivo* que permite determinar se uma fórmula do cálculo dos predicados é demonstrável ou não?

Esta questão revestiu-se de grande importância quando Gödel estabeleceu o conhecido teorema da incompletude. Tal teorema é definido por Dupuy da seguinte maneira:

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

> ...para todo sistema formal no sentido da lógica (o que implica ser dada uma linguagem formal, fórmulas dessa linguagem tidas como axiomas e regras de inferência) suficientemente rico para conter a aritmética, temos a seguinte propriedade: ou o sistema é inconsistente (leva a teoremas contraditórios), ou existe pelo menos uma proposição verdadeira que não é demonstrável nesse sistema. A sentença que lhe corresponde, portanto, não é um teorema, como tampouco, evidentemente, a sua negação, à qual corresponde a uma proposição falsa. (Dupuy 1996;p.28-9).

O teorema da incompletude colocado por Gödel, deixa em suspenso uma resposta para a questão inicial, qual seja: existe um procedimento efetivo que permita determinar se uma fórmula do cálculo dos predicados é demostrável ou não? Pois, a decisão sobre a inexistência de tal procedimento, que responderia a questão de Hilbert, depende inteiramente do que se entende por procedimento efetivo ou algoritmo. Intuitivamente a noção de *procedimento efetivo* remete às seguintes hipóteses:

- um procedimento efetivo deve ser uma espécie de conjunto de instruções de tamanho finito[3];
- deve existir um agente que reage às instruções executando-as;
- tal sistema deve apresentar facilidade para armazenar, recuperar e executar os passos das instruções[4];
- O agente (L) reage as instruções (P) passo a passo e de forma discreta, isto é, não contínua[5];
- (L) reage a (P) de maneira determinística, isto é, sua ação segue um único caminho bem definido.

Apesar do conceito de procedimento efetivo remeter a essas cinco hipóteses intuitivas ainda falta, como comenta Dupuy "...uma definição

---

[3] Instruções infinitas não são executáveis pois, não há possibilidade de cumpri-las.
[4] Exigência de uma memória de armazenagem.
[5] Não há como determinar o passo seguinte em um processo contínuo pois, neste tipo de processo, existe um infinito entre o passo anterior e o próximo passo.

rigorosa, matemática, do algoritmo. Ora, sem uma tal definição, não há resposta definida ao problema de Hilbert." (Dupuy 1996, p.29)

Uma definição rigorosa, matemática, da noção intuitiva de procedimento efetivo é proposta por Turing (1936). Ele propõe uma formalização matemática da noção abstrata de máquina que, por sua vez, definiria, de maneira rigorosa, o procedimento efetivo. Esta máquina idealizada para caracterizar a idéia intuitiva de procedimento efetivo tornou-se conhecida como *máquina de Turing*. Esta idealização apresentaria as seguintes características:

a) Ela possui uma fita longa, dividida em células discretas com um caractere escrito em cada uma delas (em geral os caracteres escritos são 0 e 1).

b) Existe um 'Scanner' (leitor) capaz de 'ler' o que está escrito em uma célula da fita.

c) Quando a máquina está 'lendo' um caractere em uma célula ela pode deixar o símbolo inalterado, apagá-lo ou imprimir outro símbolo.

Turing (1936) sugere que sua máquina, apenas com essas características, seria capaz de realizar qualquer processo ou programa que pudesse ser expresso em um código binário. Bastaria apenas apresentar o programa para o leitor da máquina, para este executar as seguintes funções em cada ciclo operacional:

(i) Ler o conteúdo da célula;
(ii) Escrever um símbolo na célula;
(iii) Mover um quadrado para direita;
(iv) Mover um quadrado para esquerda;
(v) Parar. (Turing, 1936)



O *input* para a máquina, num certo instante (t), é o símbolo que está registrado na célula onde se encontra o leitor. O *output* pode ser de dois tipos: escrever um símbolo na célula ou mover-se para a célula adjacente.

Turing acreditou que com sua definição de máquina atingiu-se a essência do procedimento efetivo. Ele e outros dois proeminentes matemáticos, Church e Markov, propuseram que a classe de funções definidas pelos sistemas formais por eles construídos, coincide com a classe de funções intuitivamente computáveis. Essa proposta foi denominada 'tese de Church-Turing"[6]. (Hartley; 1988).

Como afirma Minsky (1967;p.108) a formulação feita por Turing da tese Church-Turing diz que: "Qualquer processo que naturalmente possa ser chamado de procedimento efetivo pode ser executado por uma máquina de Turing."

Esta tese de Turing, assim como a de Church, não pode, como salienta Dupuy, ser demonstrada, pois elas relacionam os planos intuitivo e formal. A calculabilidade mecânica, proposta por Church, é uma noção intuitiva, ao passo que a máquina de Turing é definida formalmente. Dessa maneira, a tese Church-Turing tem o estatuto de uma crença que dever ser sustentada através de evidências, sendo que várias delas têm sido fornecidas a seu favor.

Dentre essas evidências podemos mencionar que diversas tentativas independentes de formulação precisa da idéia intuitiva de procedimento

---

[6]A tese de Church diz que toda função calculável (no sentido intuitivo) é uma função recursiva (no sentido matemático).



efetivo produziram uma mesma classe de funções chamadas de classe C[7] (Hartley, 1988). Outra evidência é obtida quando se demonstra que uma vasta coleção de funções intuitivamente computáveis pertencem à classe C. Tal classe pode ser ampliada *ad infinitum* utilizando regras de substituição, recursão e minimalização. Uma outra evidência, que sustenta a tese de Church-Turing, advém do fato de que não se apresentou um função intuitivamente computável que não pertencesse à classe C.

Enfatizamos a tese de Church-Turin ,pois a existência de Redes Neurais Artificiais sugere uma possível violação desta tese. Esta violação ocorreria com a proposta de uma abordagem não seqüencial e subsimbólica de sistemas mecânicos de processamento paralelo de informação. Em outras palavras, segundo alguns conexionistas (McClelland e Rumelhart, por exemplo) os modelos de RNA não seriam redutíveis às máquinas de Turing.

Retomando as idéias de Turing, este com sua caracterização de procedimento efetivo obteve uma fórmula capaz de resolver o problema da decisão de Hilbert, já mencionado anteriormente, qual seja: existe um procedimento efetivo que permita determinar se uma fórmula do cálculo dos predicados é demonstrável ou não? Turing propõe reformular tal questão da seguinte maneira: existe uma máquina de Turing capaz de decidir se uma fórmula do cálculo dos predicados e demonstrável ou não?

A proposta de Turing é trocar o problema da decisão de Hilbert por uma formulação equivalente, porém de natureza mecanicista, a saber, a

---

[7] A classe C das funções parciais recursivas podem ser construídas a partir de funções básicas C, por exemplo, função constante zero, função sucessora e função projeção, por um número finito de operações de substituição, recursão e minimalização.



indecidibilidade do problema da parada de uma máquina. Como descreve Dupuy, sua proposta poderia ser formulada da seguinte maneira:

> ...existe um procedimento mecânico – portanto, como a 'tese de Turing' exige, uma outra máquina de Turing – capaz de determinar se a primeira máquina vai parar ou, pelo contrário, entrar num processo de cálculo sem fim? (Dupuy 1996;p.34)

Turing demonstra que tal máquina não existe. Em outras palavras, não existe um método geral efetivo para se saber se uma máquina para ou não na computação de um input qualquer.

No mesmo artigo de 1936, Turing alude à existência de uma máquina de Turing universal. Tal máquina universal poderia simular, em diferentes tipos de dispositivos físicos, o comportamento de uma outra máquina de Turing, caso essa máquina universal dispusesse de tempo suficiente e de uma descrição da máquina que se quer simular. A máquina universal teria as seguintes características apresentadas neste artigo:

- Embora tal máquina seja uma idealização, uma vez que ela tem um potencial infinito, esta pode ser concebida tanto como um mecanismo abstrato de computação como um sistema físico. De forma que sempre que uma máquina de Turing estiver em um estado particular, ela também se encontrará em um estado computacional. No entanto, os processos computacionais da máquina de Turing são independentes de qualquer estado físico particular. As mesmas computações (os mesmos programas) podem ser, em geral, executadas (rodados) em máquinas fisicamente diferentes. Chamaremos esta característica

peculiar da máquina de Turing de *independência entre os estados físicos e computacionais da máquina*.

- As máquinas de Turing operam exclusivamente *com símbolos e regras pré-estabelecidas*, de acordo com a sua tabela de instruções.

- As máquinas de Turing operam *serialmente*, de acordo com os seus *estados discretos*, sendo o resultado de suas operações registrados em um endereço específico na memória.

Tendo a máquina de Turing universal a capacidade de imitar uma máquina de Turing específica, esta primeira constitui o modelo dos modelos computacionais. Dessa forma, a questão que nos parece relevante aqui é: pode esse modelo dos modelos simular a mente humana, que modela nossas representações do mundo? Teria a máquina de Turing universal a capacidade de reproduzir o desempenho inteligente da mente humana? Para responder a essas questões, Turing propõe, no artigo de 1973, o jogo da imitação.

Turing em um artigo intitulado "*Computadores e Inteligência*" (1973) apresenta uma forma de verificar a possibilidade de uma máquina pensar. A maneira de verificação da questão: pode a máquina pensar? é colocada na forma de um jogo. Esse jogo recebeu o nome de *jogo da imitação*, porém, tornou-se mais conhecido como *teste de Turing*.

O teste de Turing ou jogo da imitação sugere a seguinte situação: um homem e uma mulher ficarão ocultos a um interrogador. O interrogador terá que saber, por meio das respostas dadas, tanto pelo homem como pela mulher, quem é quem. Isto é, quem é o homem e quem é a mulher. A

comunicação entre o interrogador e os interrogados se dá, digamos, via uma passagem qualquer, por exemplo, o terminal de um computador. O importante é que a caligrafia não sirva de indício para ajudar a decisão do interrogador. As respostas serão escritas, neste caso, por meio mecânico. A mulher irá responder as perguntas procurando convencer o interrogador que ela é a mulher. O homem, por sua vez, mentirá dizendo que ele é a mulher.

Nessa primeira etapa do teste o interrogador tem como missão distinguir quem é a mulher e quem é o homem sendo que o último, a cada resposta, tenta se passar pela primeira. A partir desse esquema inicial, Turing sugere a seguinte questão: o que acontecerá se substituirmos o homem por uma máquina? Mantêm-se o esquema inicial do teste, a saber, por meio de respostas digitadas procura-se saber quem é a mulher e quem é a máquina. A máquina, que ocupará o lugar do homem no teste, terá que simular seu comportamento procurando convencer o interrogador de que ele é o ser humano. Se após várias perguntas e respostas o interrogador não for capaz de identificar o verdadeiro ser humano, então a máquina terá passado no teste de Turing e, por conseqüência, pode se inferir, segundo Turing, que ela pensa inteligentemente.

A idéia central nesse teste é mostrar que um artefato artificial, como por exemplo, uma máquina de Turing, que executa operações simples do tipo ir para direita, imprimir e apagar um símbolo e parar, pode simular o mesmo desempenho da mente humana. Turing, brilhantemente, transformou o problema complexo inicial de saber se uma máquina pode pensar inteligentemente e, consequentemente, possuir capacidades mentais como entender, conhecer, etc, na seguinte questão: pode uma máquina que segue operações simples, como ler uma

unesp

fita, move-se para direita, etc, ter o mesmo desempenho que um ser humano inteligente? O teste de Turing procura mostrar que é possível dar uma resposta afirmativa a esta questão.

A partir desta ênfase mecânicista da inteligência, proposta por Turing, o pensamento, a atividade por excelência do homem, torna-se nada mais que um procedimento mecânico. Tão mecânico que até uma máquina, seguindo instruções simples, poderia simulá-lo a partir da idealização da máquina universal e do teste de Turing. Estão lançadas assim as primeiras sementes de uma nova ciência da mente.

A conclusão extraída por Turing, de seu teste, qual seja, que o pensamento pode ser visto como uma espécie de procedimento efetivo arrebata adeptos[8] que irão apresentar teses mais fortes (como se a do próprio Turing não fosse suficientemente forte) do tipo, a mente humana pode ser vista como um programa de computador.

Outros teóricos procuram aproveitar a máquina de Turing para esboçar uma teoria da mente. Tal teoria tornou-se conhecida, na filosofia da mente, como *funcionalismo computacional* e foi desenvolvida por Putnam.

Putnam, *grosso modo*, argumentou que o pensamento inteligente é comparável à máquina de Turing e que seus estados mentais podem ser entendidos como estados de uma máquina de Turing, ou pelo menos disjunções de estados desta máquina. Como comentam os autores do livro *Computadores, Mentes e Conduta*:

---

[8] Estamos nos referindo, nesta passagem, a Newell e Simon fundadores, junto com McCarty, da inteligência artificial.

unesp

> A teoria funcionalista sustentava que os mecanismos da mente eram definíveis em termos de suas inter-relações formais e funcionais, por um lado, e de sua posição mediadora entre inputs ambientais e comportamentos reais. (Button et al,1998;p.33)

Além de definir os estados mentais em termos das inter-relações formais e funcionais, Putnam (1980) argumentava ainda que tal definição dos estados mentais permitiria uma instanciação em qualquer sistema com capacidade de processar informação estruturada.

Por alguns anos a teoria funcional de Putnam apresentou-se como uma via intermediária entre o dualismo de substâncias (do tipo cartesiano) e o reducionismo behaviorista, dado o seu liberalismo em relação à questão de quem possuiria estados mentais. Neste contexto, Abrantes (1994) ressalta que:

> Na perspectiva funcionalista, não há razão para se supor que somente o cérebro humano é capaz de 'instanciar' processos cognitivos, apontando-se para a possibilidade de que estruturas não biológicas – como computadores possuindo uma particular arquitetura – tenham 'mentes' de determinados tipos, ou seja, exibam a funcionalidade especificada. (Abrantes 1994;p.11)

Putnam em seu artigo '*The Nature of Mental States*' sustenta que a riqueza da sua posição está na hipótese de que os estados funcionais não são puramente físicos e nem abstratos. Essa perspectiva funcionalista não é, portanto, incompatível com o dualismo, apesar de ser uma abordagem mecanicista, e nem com o comportamentalismo. Ele acreditava que sua perspectiva ao descrever um estado, por exemplo de dor, seria mais rico que as alternativas rivais, pois:

> ...parece ser mais provável que o estado funcional que nós descrevemos correlacionando a dor, independe de que haja um estado químico-físico do cérebro (um organismo necessita ter um cérebro para sentir dor? Talvez alguns ganglios o fizessem) ou uma disposição comportamental a ele correlacionada. (Putnam 1980;p.230)

unesp

Com o passar dos anos, Putnam (1995) rejeitou suas idéias funcionalistas, bem como toda aspiração a unidade das ciências pelo viés da física. Uma dificuldade deste empreendimento, segundo ele, diz respeito ao fato de que os fenômenos mentais humanos não se ajustariam aos moldes explicativos das ciências naturais.

Renunciando à visão mecanicista da mente Putnam (1995; p.2) afirma:

> ...era o materialismo científico que dominava nas décadas de cinqüenta e sessenta. Eu acreditava que tudo o que existe pode ser explicado e descrito por uma única teoria...Eu acreditava que a melhor metafísica era aquela que os positivistas chamavam de 'ciência unificada' ...pela aplicação de leis fundamentais da física.

Se, em um primeiro momento, a máquina de Turing empolgou Putnam (1975) (1988), hoje este é mais cético em relação à questão segundo a qual a máquina de Turing seria o modelo correto para a organização funcional da mente/cérebro. Para ele existem poucas evidências de que o modelo de máquina de Turing seja o modelo correto para tal organização. Porém, a sua teoria funcionalista ainda não foi descartada por aqueles que trabalham na ciência cognitiva e aqueles que estão particularmente interessados nas questões acerca dos estados mentais e sua relação com o cérebro. Lycan (1987;p.37), por exemplo, é um filósofo que se enquadra na última categoria: "...ela (a teoria funcionalista) é a única doutrina positiva em toda a filosofia pela qual estou disposto (embora não tenha permissão para isso) a me matar."

Veremos, a seguir, que este entusiasmo pela teoria funcionalista ajudou a edificar um programa de pesquisa na ciência cognitiva que reinou soberano até o início dos anos 80, a saber, a IA.



3.2 Os Modelos da Inteligência Artificial

Nesta seção, apresentaremos as linhas gerais dos modelos da IA articulando as idéias desse programa de pesquisa com as noções presentes na máquina de Turing e com a perspectiva funcionalista vigente em grande parte da atual filosofia da mente. Foi a partir das idéias desenvolvidas por Turing, a saber, que o pensamento pode ser concebido como a expressão de um procedimento mecânico, que se originou a concepção dominante na IA, segundo a qual é possível simular certas atividades mentais em uma máquina de Turing. Para que ocorra tal simulação basta apenas fornecer os algoritmos apropriados.

Um algoritmo é um conjunto de instruções que explicita para cada passo do processamento de informação de um sistema o que deve ser feito. Duas exigências devem ser cumpridas para a execução de tais instruções. Primeiro, que exista um lugar para armazenar, recuperar e executar os passos das instruções, ou seja, que exista uma memória. Segundo, as instruções devem estar bem definidas e serem finitas.

O importante em um procedimento efetivo não é quem executa as instruções, mas sim a forma como estas são executadas. Basta apenas que sejam executados os passos descritos por Turing: imprimir, ir para a direita ou esquerda, apagar e parar. A forma de proceder determina um procedimento algorítmico, independente de se considerar se este é executado ou ordenado por um agente humano, animal, etc.



Turing idealizou a máquina de Turing como exemplo formal do que seria um procedimento algorítmico. Tal máquina teria as seguintes características:

- seus estados físicos seriam independentes dos seus estados computacionais;
- suas ações seriam operações sobre símbolos e regras pré-estabelecidas;
- a máquina operaria *sempre* serialmente e de forma discreta.

A partir dessa caracterização de máquina sugerida por Turing, os pesquisadores da IA extraem a seguinte hipótese: tudo aquilo que a mente humana pode realizar e que seja passível de uma descrição precisa pode ser executável por um computador devidamente programado. A IA toma a hipótese de Turing sobre a natureza mecânica do pensamento e a generaliza através da seguinte tese: a mente, nas suas operações inteligentes, pode ser entendida como uma espécie de programa de computador.

Além desse vínculo com as idéias de Turing, a IA compartilha algumas das teses do funcionalismo, entre elas, a tese de instanciação múltipla, segundo a qual não há razão para se supor que somente o cérebro humano é capaz de instanciar processos cognitivos. Existe a possibilidade de que estruturas não-biológicas tenham mentes de determinados tipos.

Ao nosso ver, a tese da instanciação múltipla já estava presente na idealização da máquina de Turing universal, em especial no pressuposto da independência entre os estados físicos e computacionais da máquina. Ou seja, os estados computacionais, dada sua independência em relação aos estados físicos, podem ser executados em qualquer sistema que processe informação. Além da independência entre os estados físicos e computacionais da máquina, devemos

lembrar que a característica fundamental de um algoritmo não reside em *onde* ele é executado mas sim, em *como* ele é executado.

A tese da independência entre os estados físicos e computacionais da máquina gerou a conhecida metáfora do cérebro, entendido como o hardware, e a mente como exemplo do *software*.

Johnson-Laird (1988;p.39) ressalta a tese da instanciação múltipla presente na teoria da IA. Ele diz que uma máquina:

> ...poderia ser feita de dentes de engrenagem e alavancas, como a calculadora antiquada, poderia ser feita de um sistema hidráulico através do qual a água circula; poderia ser feita de transistores impressos em um chip de silício através do qual a corrente elétrica flui, poderia ser executada até pelo cérebro.

Ao colocar esta citação, procuramos ressaltar que, na perspectiva da IA, os cérebros estão aí, mas não são essenciais. O essencial é que os estados mentais, embora implementados no cérebro, poderiam ser executados em um número indefinido de *hardwares*. O essencial é que os diferentes *hardwares* possam desempenhar o mesmo papel que o cérebro. Se eles assim o fizerem, afirmam os adeptos da IA, poderemos concluir que os processos mentais são como processos computacionais. Isto é, são como algoritmos que podem ser elaborados a partir de regras e executados por qualquer sistema de processamento seqüencial de informação que receba um *input* e produza um *output*.

Antes de prosseguirmos com a análise das linhas gerais dos modelos da IA gostaríamos de aproveitar a discussão da noção de instanciação múltipla e mencionar a revolução que a ciência cognitiva admite ter introduzido no estudo da mente inteligente, qual seja, a revolução copernicana da inteligência.

unesp

A ciência cognitiva, com os trabalhos desenvolvidos inicialmente pelos cientistas da IA e mais tarde pelos conexionistas, fortalece os estudos dos processos mentais realizados sem um sujeito cognitivo absoluto. Isto é, sem aquele sujeito clássico da filosofia do século XVII.

Esta nova ciência da mente, ao levantar a possibilidade de sistemas artificiais, construídos a partir de *chips* e silício, possuírem certas atividades cognitivas, retira da esfera do humano capacidades até então exclusivas dela, como pensamento, entendimento, conhecimento, etc. De acordo com os idealizadores da revolução copernicana da inteligência, teríamos tipos de mentes que produziriam diferentes tipos de inteligência.

Os cientistas cognitivos que trabalham na IA e na RNA acreditam que termos como *inteligência e entendimento* podem ser empregados em qualquer sistema que se comporte como o cérebro humano, processador de informação. O filósofo Dennett (1996), por exemplo, acredita que esta postura de interpretar o comportamento de uma entidade, seja ela pessoa, animal, artefato, como se fosse um agente racional que governa suas ações e escolhas, pode ser proveitosa para os estudos da mente. Ele chama tal postura de intencional e julga que:

> ...se adotada com cuidado, a postura intencional não é apenas uma boa idéia mas também é a chave para desvelar os mistérios da mente – de todos os tipos de mente. É um método que explora as similaridades para descobrir as diferenças que se acumularam entre as mentes de nossos ancestrais e as nossas mentes e aquelas de nossos companheiros habitantes do planeta. (Dennett 1996;p.32)

Dennett sugere que a postura intencional, manejada com cuidado, pode nos fornecer uma perspectiva sólida para estudar os fenômenos

mentais nas mais diversas áreas do conhecimento. A estratégia adotada em tal postura consiste em abordar a entidade estudada como um agente e, com isso, predizer, e em certo sentido explicar, suas ações.

Dupuy (1996) refere-se à Dennett como sendo um desconstrutor da consciência e simpatizante da concepção de processos sem sujeito. Este trecho extraído do livro *Consciousness Explained* (1991) parece comprovar tal observação a seu respeito:

> Da mesma maneira que as aranhas não têm que pensar na maneira como tecem sua teia, assim também nós, por nós mesmos, os seres humanos, não temos de determinar consciente e voluntariamente quais narrativas temos de contar, e como contá-las, a fim de construir nosso eu: nossas narrativas são, sem dúvida, tecidas (são texturas, textos), mas, quanto ao essencial, não somos nós que as tecemos. A consciência humana, ou seja, o si como narrativa, é o produto desses textos, não é a fonte deles. (Dennett 1991;p.418)

A questão da revolução copernicana realizada pela ciência cognitiva sobre a natureza da inteligência seria por si só tema de uma tese. Encerraremos esta discussão por aqui e retomaremos a caracterização dos modelos da IA. Devemos lembrar as duas teses aqui mencionadas, presentes na teoria destes modelos: a proposta da IA é entender os processos mentais a partir de uma perspectiva conhecida como *plano representacional de análise,* que é realizado independente dos aspectos físicos e biológicos. Postulando um plano representacional de análise, os cientistas da IA acreditam ser possível tratar a atividade mental como uma abstração que pode ser transcrita em termos de símbolos e regras.

Dessa forma, a caracterização do conhecimento presente nos modelos da IA baseia-se em computações compostas por regras abstratas sobre representações simbólicas. Podemos descrever tais modelos como sendo

constituído por: a) um programa ou um conjunto de algoritmos; b) circuitos lógicos que 'reagem' às instruções transcritas nos algoritmos; c) uma memória de armazenagem; d) mecanismos determinísticos que seguem regras pré-estabelecidas.

Algumas variações podem ser acrescentadas a este esquema geral Minsky (1976), por exemplo, sugere que o programa de jogo de dama desenvolvido por Newell, envolve uma heurística. Isto é, o programa avalia qual é a melhor jogada a ser feita de maneira não determinística. A esse respeito ele comenta:

> ...o que torna tal programa interessante e significativo não é o fato de poder usar tentativa e erro para resolver problemas. O que induz ao comportamento inteligente é a coleção de métodos e técnicas que selecionam o que deve ser tentado em seguida, que medem a situação e escolhem uma jogada plausível (nem sempre boa) e usam a informação obtida em tentativas anteriores a fim de dirigir as análise subsequentes para melhores direções. (Minsky 1976;p.153).

Essa capacidade de planejar seu passo futuro a partir das lições extraídas nos passos anteriores, isto é, esse *feedback* possibilitado pelo programa, faz com que Minsky afirme que a idéia de que um computador só resolve um dado problema quando cada etapa da solução estiver claramente especificada pelo programador é, ao pé-da-letra, enganadora. Para ele, programas que seguem o modelo proposto por Simon e Newell não podem ser chamados de determinísticos.

Em favor de Minsky, Papert (1994) argumenta que ninguém está mais convencido de que a máquina só realiza aquilo para a qual foi programada. Para ele os especialistas em programação estão construindo programas cada vez mais complexos ao ponto de serem:

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

> ...incapazes de antecipar todas as situações, todos os problemas que se apresentarão ao computador, eles tentam portanto captar 'princípios gerais' e 'implementá-los' no computador, de modo que este se sirva nos casos não previstos, com resultados muitas vezes surpreendentes para o próprio programador. (Papert 1994;p.247)

No mesmo contexto, Simon e Newell apresentam uma outra variação do esquema geral dos modelos propostos anteriormente. Esse novo programa chama-se SOAR (State, Operator and Result). Escolhemos este programa para análise, pois ele possui a capacidade de se estruturar a partir das informações colhidas no meio.

O SOAR[9] é uma espécie de arquitetura computacional capaz de executar programas. Um dos programas implementados no SOAR era incorporado a um robô que movimenta blocos. O robô responde corretamente aos comandos do tipo: pegue este bloco em forma de círculo e mova três espaços à direita. Ele está equipado com sensores óticos e braços mecânicos para executar esta operação. Acoplado ao robô, temos um conjunto de instruções para manipulações de símbolos formais que, quando devidamente conectados, recebem *inputs* dos sensores óticos e enviam *outputs* aos mecanismos motores.

Nos anos 60 os pesquisadores da IA voltaram a atenção para a capacidade humana de conversação. A idéia predominante entre estes pesquisadores era elaborar modelos que poderiam ter sucesso no jogo da imitação (teste de Turing). Das muitas tentativas de elaborar programas que simulassem a capacidade humana de conversação duas, em especial, tornaram-se famosas: o programa de Colby (1975) que modelava o diálogo entre uma personalidade paranóica com uma certa obsessão.

---

[9] Para mais detalhes sobre este programa ver Waldrop (1988).

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

Este programa foi elaborado com o intuito de simular o desempenho lingüistico de uma pessoa paranóica. Vejamos um pequeno trecho de tal programa:

> Por que você está no hospital?
> ***Eu não deveria estar aqui.***
> Quem o trouxe para cá?
> ***A polícia.***
> Que problema você teve com a polícia?
> ***Os policiais não cumprem seu dever.***
> E o que é que eles fazem então?
> ***Os policiais prendem as pessoas erradas.*** (Colby,1975)

Além deste programa, elaborado para o estudo do desempenho lingüistico de pessoas com distúrbios psiquiátricos, um segundo programa, elaborado por Weizenbaum (1966) tornou-se muito famoso pois, para alguns pesquisadores da IA este representava um bom exemplo do sucesso desta nova linha de pesquisa. O programa de Weizenbaum tentava reproduzir a conduta de um terapeuta que adota a postura *não diretiva* com seu paciente. Vejamos um pequeno trecho deste programa:

> *Os homens são todos iguais.*
> ***De que maneira?***
> *Eles estão sempre nos irritando de uma forma ou de outra.*
> ***Você pode dar um exemplo específico?***
> *Bem, meu namorado me fez vir aqui.*
> ***Seu namorado a fez vir aqui.***
> *Ele acha que eu estou deprimida.*
> ***Eu sinto muito em saber que você esta deprimida.*** (Boden 1977;p.106-107)



Os dois programas foram elaborados de forma que pudessem enganar, por alguns momentos, alguém que não conhecesse a forma de amarração do programa. Porém, ambos os construtores admitem que qualquer um que possuísse um certo conhecimento de programação poderia, como diz Gardner, desmascarar o programa.

Em resumo, procuramos, com tais exemplos, ilustrar as principais características aqui mencionadas dos modelos da IA. Estes são construídos à partir de algoritmos que especificam, passo a passo, a conduta a ser seguida pelo sistema. Os programas são escritos em símbolos geralmente representados pelos algarismos 0 e 1.

Devemos lembrar ainda de que os pesquisadores da IA sustentam a tese da independência dos planos biológicos e/ou físicos no estudo da mente. Para a simulação das atividades cognitivas, estes pesquisadores afirmam que o importante é o programa e não onde ele é instanciado. Levando às últimas conseqüências a tese mecanicista de Turing, os cientistas da IA sugerem que a mente humana pode ser entendida como um programa de computador. Na próxima seção iniciaremos as críticas a este programa de pesquisa existente na ciência cognitiva.

3.3 ALCANCE E LIMITES DOS PROGRAMAS DA INTELIGÊNCIA ARTIFICIAL

3.3.1 O PROBLEMA SEMÂNTICO

O tom geral desta seção parecerá, numa primeira leitura, extremamente pessimista com a relação ao projeto da IA. Porém, devemos informar que nossa intenção, ao apontar os problemas desta vertente da ciência cognitiva, possui um aspecto positivo, qual seja: delineando os limites da IA indicamos o alcance da abordagem lógico/abstrata na explicação da atividade mental.

Foi graças às tentativas dos pioneiros da IA que sabemos hoje da possibilidade de explicar, satisfatoriamente, em termos de algoritmos, as estratégias para a resolução, por exemplo, de um problema lógico. Essas tentativas obtiveram muito êxito em programas que jogavam dama e xadrez. Também, graças a estes pesquisadores é que conhecemos as limitações da abordagem representacional na explicação, por exemplo, das repentinas mudanças de humor ou controle motor de uma pessoa. Tal comportamento não pode ser inscrito em um padrão explicativo construído a partir de representações lógicas.

Conhecer as vantagens e desvantagens de um modelo explicativo sobre o comportamento inteligente acrescenta muito à toda comunidade interessada na investigação deste fenômeno. Devemos entender o papel do erro na construção do conhecimento. Este pode ser visto como fonte, não

de ironias ou críticas, mas sim de uma, permanente, atualização das nossa antigas crenças.

Começaremos esta seção examinando uma crítica que contesta a tese central subjacente aos modelos de IA, a saber, que a mente é uma espécie de programa de computador. Usaremos para combater esta tese a réplica elaborada por Searle (1984). Tal réplica está contida no argumento do quarto chinês. Como já descrevemos este argumento no capítulo anterior, iremos, inicialmente, apresentar algumas conclusões que podem ser extraídas deste argumento e que apontam os limites dos modelos da IA.

Searle elabora sua crítica a partir das teses centrais da IA radical anteriormente mencionadas: a) que a mente é uma espécie de programa de computador e b) que qualquer sistema físico que possua um programa correto com *inputs* e *outputs* tem uma mente. O cérebro humano nessa concepção é mais um entre os vários sistemas físicos que executam programas bem definidos. O ponto crucial que será criticado por Searle é, como menciona Humphrey (1994;p.8), a concepção segundo a qual:

> ...o cérebro pode ser caracterizado como um dispositivo que recebe "informação" e a processa para produzir mais "informação" (ele certamente assim o faz, se desejarmos dar-lhe essa descrição) e poderíamos dizer que o importante é a relação matemática entre o insumo e o produto. Nesse caso, os sentidos subjetivos específicos seriam idênticos não aos processos cerebrais físicos específicos, e sim às operações lógicas específicas realizadas.

De acordo com essa perspectiva, a ênfase da abordagem da IA reside, não tanto no material que executa os programas específicos, mas nas operações lógicas que possibilitam tal programa. Mencionamos que esta abordagem deriva, em parte, do funcionalismo de Putnam. Esta concepção sugere

que os estados mentais podem ser entendidos em termos de suas inter-relações formais e funcionais. Para refutar esta posição, Searle elabora um argumento que, ao seu ver, não está ligado a qualquer estágio tecnológico dos computadores, mas repousa sobre o princípio básico da IA. Sendo assim, ele diz que, pode se aperfeiçoar tanto os programas como os computadores que seu argumento, ainda sim, será válido.

Searle parte para construir seu argumento da caracterização de computador. A caracterização de computador por ele mencionada remete àquela proposta por Turing, qual seja, um sistema que tem suas operações especificadas em termos de símbolos e regras que determinam o próximo passo a ser dado. Será essa caracterização em termos formais que ele criticará, pois os pesquisadores da IA admitem que os processos mentais são idênticos aos processos computacionais.

Afirma Searle que sendo os programas da IA apenas manipuladores de símbolos, estes possuem apenas uma sintaxe enquanto que a mente tem por definição conteúdos. Ou seja, na sua concepção, os estados mentais são sempre acerca de alguma coisa. Ele admite que mesmo se o pensamento ocorresse em uma série de símbolos, como sugerem Turing e os pesquisadores da IA, deve haver algo mais no pensamento do que essas séries abstratas, pois as séries por si só não têm qualquer significado. Isto é, zeros e uns impressos na fita não possuem significados, ou seja, não são acerca de coisa alguma. Argumenta Searle (1984;p.39):

> A razão porque nenhum programa de computador pode ser uma mente é simplesmente porque um programa de computador é apenas sintático, e

> as mentes são mais do que sintáticas, no sentido de que possuem mais do que uma estrutura formal, têm um conteúdo.

Para ilustrar esse ponto, a saber, que os programas não possuem semânticas, Searle elabora o já conhecido argumento do quarto chinês. Em linhas gerais, pois já apresentemos este argumento no capítulo anterior, o argumento sugere o seguinte experimento mental: uma pessoa que não sabe chinês recebe um conjunto de instruções (em português) que lhe permite responder questões em chinês, sem realmente adquirir qualquer entendimento de tal língua. O ponto crucial da experiência é que, o fato de que o *output* do indivíduo não possa ser distinguido de alguém que fala chinês, não demonstra, por si só, que o indivíduo que produziu estes *outputs* entende chinês.

Essa experiência de pensamento guarda estreitas semelhanças com o teste de Turing. Lembremos que no jogo da imitação o computador teria que simular, por meio de respostas, o comportamento de um homem. Tal computador opera apenas segundo manipulações formais. Searle inverte o jogo. Será o próprio ser humano que fará a manipulação formal e produzirá os *outputs*. Essa inversão no jogo parte da seguinte tese da IA: tanto os seres humanos quanto às máquinas, apesar da diferença dos seus elementos físicos, podem apresentar o mesmo esquema lógico computacional nas suas operações mentais.

Ora, mas qual é a conclusão a que chegamos no caso do argumento do quarto chinês? A conclusão a que chegamos é que, apesar do indivíduo manipular as regras e produzir *outputs* em chinês, ele não compreende tal língua. A partir dessa conclusão, Searle aponta duas dificuldades presentes no projeto da IA. A primeira é que, a despeito dos computadores manipularem os símbolos e produzirem *outputs* que deixam o interrogador indeciso sobre quem é a

mulher e quem é o computador, no jogo da imitação a máquina não compreende nada do que está executando.

A segunda dificuldade, apontada por Searle, diz respeito a tese da IA que afirma os processos mentais são como processos de um programa computacional. Então aquilo que os seres humanos não conseguem realizar, nenhum programa, por mais sofisticado e complexo que seja, conseguirá. Isto é, se meu programa não foi suficiente para compreender chinês, os programas computacionais, criados pelos adeptos da IA, que são idênticos ao meu, também não compreenderão. Conclui Searle (1984;p.41):

> Se não compreendemos o chinês, então nenhum outro computador, pode compreender o chinês, porque nenhum computador digital, em virtude da simples execução de um programa, tem algo que nós não temos.

O que Searle busca enfatizar com seu argumento do quarto chinês é que a mera manipulação de símbolos não gera os aspectos semânticos da linguagem pois, assim como a consciência, estão ligados ao aparelho biológico humano. Veremos, a seguir, uma segunda limitação das hipóteses subjacentes aos modelos da IA.

3.3.2 OS ABUSOS FILOSÓFICOS DA LINGUAGEM ORDINÁRIA

Esta limitação da hipótese mecanicista subjacente aos modelos da IA diz respeito ao uso dos termos da linguagem ordinária. Poderíamos dizer que tal crítica, como a anterior, refuta pontos centrais da teoria da IA. Ela diz

respeito ao uso (ou abuso) dos termos como 'entendimento', 'pensamento', 'inteligente', etc.

Os adeptos do projeto da IA não se cansaram de proferir sentenças bombásticas do tipo: "*A longo prazo, os computadores serão capazes de fazer tudo.*" (Simon 1994;p.249). Ou ainda, "A máquina também tem um pouco de criatividade, fazendo coisas surpreendentes." (Papert 1994;p.249). Sem mencionar a célebre resposta de McCarty a Searle sobre quais crenças teria o termostato: "O meu termostato tem três crenças – está demasiado frio aqui está demasiado quente, e está bem aqui." (Searle, 1984, p.38)

Os autores do livro *Computadores, Mentes e Conduta* (1998) sentiram-se instigados a combater o que eles chamam de "filosofia que dá origem ao feitichismo do computador." A estratégia usada para esse combate é uma vigorosa análise da linguagem. No melhor estilo da filosofia analítica, seguindo principalmente Ryle e Wittgenstein, os autores do livro procuram discutir a validade do emprego de termos como 'inteligente', 'pensamento', 'crença', etc., para os *outputs* produzidos por um computador.

Para localizar a discussão, os autores remetem ao, já discutido, teste de Turing. Segundo eles, a questão crucial que deve ser formulada sobre este é: se existisse um máquina que passasse pelo teste de Turing, deveríamos dizer que essa máquina é tão inteligente, a ponto de serem seus *outputs* indestingüíveis dos outputs humanos? Esta não é um questão que requeira apenas a discussão acerca da construção dos programas. Requer, antes de tudo, uma discussão sobre o uso das palavras 'inteligente', 'pensamento', etc.



A discussão sobre o correto uso das palavras remete à análise da linguagem ordinária pois:

> ...o próprio teste de Turing invoca o julgamento dos usuários da linguagem ordinária, pois são eles que são convidados a dizer se há ou não uma diferença entre o input do humano e o da máquina. (Button et al.1998;p.33)

Ou seja, os autores argumentam que a questão: existe uma máquina que passe no teste de Turing e, por conseqüência, possa ser chamada de inteligente, merece uma minuciosa análise, pois está no âmbito da filosofia da linguagem ordinária. Esta se preocupa com as respostas daqueles que como nós, falam a linguagem ordinária e de onde as palavras 'pensar', 'entender', 'inteligente', usadas no teste de Turing, são tomadas.

Sendo assim, a questão é colocada não no âmbito científico ou computacional, mas sim no âmbito dos *usuários da linguagem*. A questão, colocada neste âmbito, ganha a seguinte forma: deveríamos nós, em qualquer sentido comum das palavras 'pensar', 'entender', 'inteligente', dizer que uma máquina que passe no teste de Turing deve possuir tais capacidades?

A recolocação da questão procura explicitar que o que está em jogo não é uma disputa entre termos técnicos, provenientes da ciência, e ordinários. Mas, sim, entre usos ordinários e abusos filosóficos. Um vez que os autores do livro em questão partem do seguinte pressuposto: muitos problemas filosóficos originam-se do uso incorreto das palavras ordinárias, a atribuição de pensamento e entendimento a uma máquina, para esse autores, é um exemplo típico do uso incorreto.

A problemática em questão diz respeito ao emprego dos termos 'pensamentos' e 'entendimento' na dimensão das máquinas. Segundo os autores, não há problema algum no fato de que os engenheiros chamem seus computadores de máquinas pensantes, desde que a expressão máquina pensante seja um termo técnico, próprio dessa área. Porém, o problema é que os engenheiros e cientistas cometem abusos filosóficos e insistem que o termo pensamento, por exemplo, por eles usado, tem o mesmo significado que o nosso uso ordinário desse termo.

A insistência em dizer que o uso dos termos 'pensamento', 'inteligente', entre outros, é usado no mesmo sentido que os usuários da linguagem ordinária os fazem tem o objetivo de:

> ...provocar, desafiar e enfrentar o resto de nós. Seu objetivo é contradizer algo em que o resto de nós acreditamos. Aquilo em que o "resto de nós" acreditamos é simplesmente o seguinte: pensar e entender são algo distintivo dos seres humanos ou, talvez, também de outras formas superiores de vida, e essas capacidades nos apartam do meramente mecânico. (Button et al.1998;p.40)

Mas, porque pensar e entender, para esses autores, seria uma capacidade peculiar de algumas formas superiores de vida e dos seres humanos? O que impede um sistema meramente mecânico de adquirir estas capacidades?

A concepção seguida por Button, Coulter, Lee e Sharrock, compartilhada por Ryle e Wittgenstein, sugere que a linguagem que falamos é uma construção social. Não existe, segundo eles, regras inatas que podem ser transformadas em algoritmos e implementadas em um computador. A linguagem que falamos resulta, dizem eles, do fenômeno da vida e das interações sociais, mais do que da neurofisiologia. A linguagem que resulta do fenômeno da vida e das interações sociais, desenvolveu-se ao longo do tempo. Muitas coisas que

dizemos nascem das transformações da vida em sociedade e tem sentido, único e exclusivo, naquele contexto social.

Retomando Ryle (1949), tais autores contestam a concepção que trata a linguagem como um produto da mente que atua separadamente da atividade humana.

Ryle, ao criticar o dualismo cartesiano, tentava retificar os erros de uma doutrina que separa a mente do corpo. A sua retificação buscava explicitar que os estados mentais se constituem na plena interação do homem (dotado de um corpo) com o mundo. Conforme ressaltamos no capítulo anterior, a sua teoria conceberia a mente em termos de disposições físico-biológicas do organismo, que se atualizam na ação.

Esta concepção de mente deve ser preservada se quisermos entender a linguagem como um produto social. Pois, afinal, a linguagem enquanto fenômeno produzido na atividade social não pode, como querem os pesquisadores da IA, ser captada a partir de regras determinísticas. Sendo assim, podemos dizer que para Button, Coulter, Lee e Sharrock, a linguagem ordinária, dada a sua característica dinâmica, não pode ser traduzida para sistemas mecânicos que manipulam símbolos. Além disso, tais sistemas manipuladores de símbolos existem isoladamente, eles não possuem qualquer atividade coletiva que se assemelhe à nossa vida social e, portanto, não são capazes de produzir a linguagem ordinária.

Após estas críticas aos modelos da IA iremos, no capítulo seguinte, apresentar a descrição e alguns modelos de RNA que foram elaborados com a esperança de superar as limitações encontradas pelos modelos algorítmicos.



CAPÍTULO 4

MODELOS DE REDES NEURAIS

INTRODUÇÃO

No capítulo anterior descrevemos a máquina de Turing. Tal descrição foi necessária, pois dela se extrai uma das teses centrais do programa da IA, a saber, que o pensamento pode ser entendido como um procedimento algorítmico. Esta hipótese mecanicista do pensamento, proposta inicialmente por Turing, remeteu-nos a seguinte questão: podem as máquinas pensar? Para responder tal questão descrevemos o teste idealizado por Turing nomeado jogo da imitação. Este jogo propõe uma resposta afirmativa para a questão sobre a possibilidade de existência de pensamentos em máquinas.

A partir da resposta sugerida por Turing, a tese mecanicista do pensamento ganha entusiastas. Dentre estes citamos, em especial, Putnam que aproveita o modelo de máquina de Turing para esboçar uma teoria da mente conhecida, na filosofia da mente, como funcionalismo. A teoria funcionalista proposta por Putnam, como já salientamos, ajudou a edificar o programa de pesquisa existente dentro da ciência cognitiva denominado IA. Dentre as teses defendidas pelos pesquisadores da IA uma, em especial, remete à proposta de Turing, qual seja, que a mente pode ser concebida como uma espécie de programa de computador. Os modelos da IA, construídos para simular determinadas atividades cognitivas, estão assentados nesta tese. Estes modelos de simulação da atividade cognitiva funcionam a partir de regras e manipulações de proposições, que, conforme indicamos, possuem sérias limitações. No início da década de 80, várias tentativas para superar as limitações da IA foram realizadas na ciência cognitiva, surgindo aí o movimento conexionista ou redes neurais artificiais.



Com intuito de analisar tais tentativas, analisaremos, neste último capítulo, na seção 4.1, os modelos de neurônios formais de McCulloch e Pitts (1998). Em seguida, na seção 4.2, apresentamos uma breve descrição das Redes Neurais Artificiais (RNA). Feita a descrição geral do modelo conexionista, descreveremos, na seção 4.3, a rede denominada *Perceptron* e a rede de *Kohonen*. Por fim, na seção 4.4, apresentaremos algumas críticas a estes modelo.

4.1 OS NEURÔNIOS FORMAIS DE McCULLOCH E PITTS

As RNA foram, inicialmente, definidas por McCulloch e Pitts no artigo "A Logical Calculus of the Ideas Immanent in Nervous Activity" (1998). Tal artigo descrevia redes de neurônios formais que representariam a lógica subjacente aos processos discretos de manipulação de sinais no cérebro ou no computador. O neurônio formal, por eles descrito, é extremamente simples se comparado ao neurônio biológico. Uma ilustração simplificada deste neurônio pode ser verificada na figura 1 abaixo:

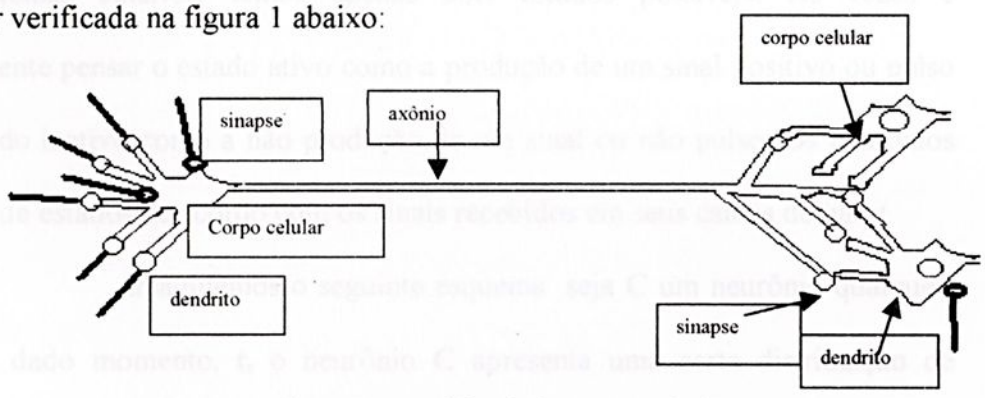


Figura 1 – descrição simplificada do neurônio biológico

O neurônio formal de McCulloch e Pitts, como descreve Kovács (1996;p.28-9), apresenta as seguintes propriedades:

> ...a sua saída poderia ser um pulso ou não pulso, e suas várias entradas tinham ganho arbitrário e poderiam ser excitatórias ou inibitórias. Para determinar a saída do neurônio, calculava-se a soma ponderada das entradas com os respectivos ganhos como fatores de ponderação, positivos nos casos excitatórios e negativos nos casos inibitórios. Se este resultado fosse maior ou igual a um certo limiar então a saída do neurônio era pulso, e caso contrário era não pulso.

Abaixo veremos, na figura 2, uma representação localizada do neurônio formal:

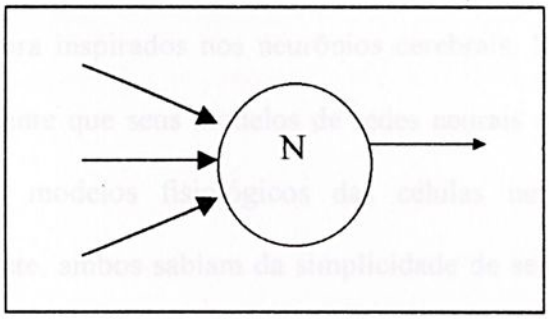


Figura 2- Representação de um neurônio de McCulloch-Pitts. As setas à esquerda do círculo representam os dados de entrada do neurônio e a seta à direita representa a sua resposta.

Em cada instante, um neurônio formal se encontra em dois estados possíveis: ativo ou inativo. Para cada um desses estados existe um sinal de *output* associado, transmitido ao longo da ramificações do canal de *output*. Como nos sistemas binários, temos apenas dois estados possíveis. As vezes é conveniente pensar o estado ativo como a produção de um sinal positivo ou pulso e o estado inativo como a não produção de um sinal ou não pulso. Os neurônios mudam de estados de acordo com os sinais recebidos em seus canais de *input*.

Imaginemos o seguinte esquema: seja **C** um neurônio qualquer. Em um dado momento, **t,** o neurônio **C** apresenta uma certa distribuição de atividade em suas conexões de *input*. Se um ou mais canais de *input* inibitórios estiverem transmitindo sinais então, no instante **t+1**, o neurônio estará no estado inativo, não transmitindo sinal algum pelo seu canal de *output*. Por outro lado,

quando nenhum canal de *input* inibitório estiver transmitindo sinais, somamos o número de canais excitatórios que estão transmitindo sinais e, se este número for igual ou superior ao limiar do neurônio **C**, ele estará, no instante **t+1**, no estado ativo disparando um sinal pelo seu canal de *output*. Em outras palavras, um dado neurônio **C** estará ativo, no instante **t+1**, se e somente se no instante **t** o número de *inputs* excitatórios for igual ou superior ao seu valor limiar de disparo.

Embora inspirados nos neurônios cerebrais, McCulloch e Pitts compreendiam claramente que seus modelos de redes neurais não podiam servir adequadamente como modelos fisiológicos das células nervosas. Como já afirmamos anteriormente, ambos sabiam da simplicidade de seu neurônio formal frente ao comportamento elétrico da célula nervosa.

Podemos dizer que até o momento de criação da rede do tipo McCulloch e Pitts, a abordagem dos estados mentais era apenas matemática. Isto é, a rede neural de McCulloch e Pitts não era concebida como um modelo físico do cérebro capaz de simular certas atividades cognitivas. A partir dos anos 80 é que os modelos conexionistas passam a ser encarados como modelos para explicar, segundo Rumelhart, (1991;p.8) "...a nossa visão da natureza da mente e da sua relação com o cérebro." Porém, como procuraremos argumentar na seção sequinte, as RNA's, modernamente definidas, podem ser consideradas, em certo sentido, como generalizações do modelo inicial proposto no artigo de 1943 (em nossa referência 1998).



4.2 DESCRIÇÃO DAS REDES NEURAIS ARTIFICIAIS

Nesta seção iremos descrever os princípios fundamentais das redes neurais modernamente definidas. Segundo Caudill & Butler (1992) elas apresentam as seguintes características:

·As RNA são *sistemas dinâmicos* de processamento de informação distribuída em paralelo;

·As unidades básicas de processamento de informação das RNA's são *unidades físicas* (não simbólicas) conhecidas como *neurônio símile*. (daqui por diante *nódulos*). Estes constituem redes ou camadas que se intercomunicam;

·O *valor de ativação*, **ai**, das unidades atua como parâmetro de uma função *output* **F (ai)** produzindo o valor de output **Oi**;

·Os *nódulos* possuem, a cada momento, estados de ativação aos quais correspondem valores de ativação, **ai**, que podem ser contínuos ou discretos;

·A conexão entre as unidades ativadas dá lugar a *emergência de padrões de conectividade* que estocam informação à respeito do estado geral do sistema;

·Os padrões de conectividade possuem valores de ativação que são propagados através do sistema produzindo *redes de retroalimentação*. Estas permitem a atualização do estado de ativação geral do sistema à medida que ele interage com o meio ambiente;



·A informação disponível na rede, em dado momento, é distribuída através dos padrões de conectividade que se estabilizam no sistema. Eles constituirão a memória dinâmica (não localizada pontualmente) na RNA.

Uma ilustração geral das RNA's é apresentada na Figura 3 :

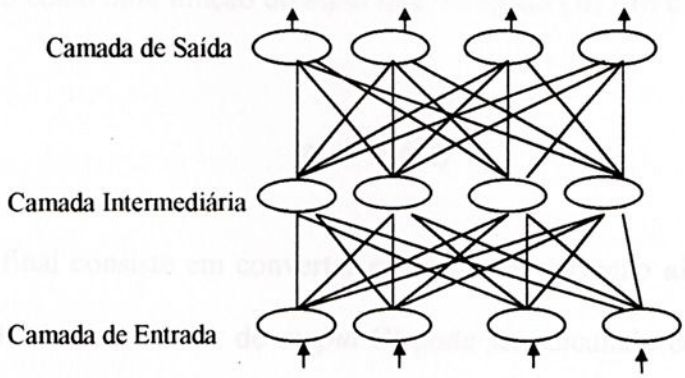


Figura 3 – Esquema geral de uma RNA

De maneira resumida podemos reunir o funcionamento de uma RNA em três fases de funcionamento, ainda segundo Caudill & Butler (1992):

1) Cada nódulo, ou neurônio artificial, recebe *input* dos neurônios vizinhos cujo valor total dos *inputs* recebidos pode ser calculado pela seguinte expressão

$$Ii = \sum_{J=1}^{n} WijXj$$

onde: **Ii** é a função soma dos input recebidos pelo neurônio **i** dos n neurônios da rede;

Xj é o sinal de input transmitido pelo neurônio **j**;

Wij é o valor que representa o peso na conexão que liga o neurônio **i** ao neurônio **j**.

2) O *input* total recebido pelo neurônio **i** determina um valor de ativação (**ai**), equivalente ao estado de excitação de um neurônio biológico, que pode ser calculado como uma função do *input* total recebido (**Ii**) isto é:

$$\mathbf{ai = F\ (Ii)}$$

3) A etapa final consiste em converter este valor de ativação **ai** em um sinal de *output* **Oi**. Assim, o valor do *output* **Oi** pode ser calculado como uma função do valor de ativação **ai**.

$$\mathbf{Oi = F(ai)}$$

Podemos estabelecer, a partir desta descrição de uma RNA, algumas diferenças fundamentais entre os modelos computacionais conexionistas e os modelos computacionais da IA. Nos modelos computacionais da IA temos o chamado procedimento algorítmico para a resolução de problemas. Isto é, o modelo terá, de antemão, um conjunto finito de instruções que serão executados de forma discreta e determinística. Nesta perspectiva, os modelos da IA são construídos sob a pressuposição de que os estados mentais são abstrações que podem ser estudadas em termos de símbolos e regras que compõe os algoritmos. Esse plano representacional de análise permite o tratamento da atividade mental como sendo uma seqüência bem definida de operações denominadas de *procedimento efetivo*.



A operação de uma RNA, afirmam os conexionistas, como Rumelhart e McClelland (1986) e Caudill e Butler (1992), entre outros, não é algorítmica. As razões para essa afirmação podem ser encontradas na generalidade da escolha das funções, **F**, e em seus processos de auto-organização presentes nos modelos conexionistas. A simulação de um dado processo cognitivo, em uma RNA, ocorrerá, não a partir da execução determinada pelo algoritmo, como nos modelos da IA, mas, sim, pela alteração dos pesos das conexões entre os nódulos. Uma dada tarefa, por exemplo, de reconhecimento de padrões, será concluída em uma rede quando ela se acomodar em um conjunto imprevisível de valores de conectividade. Diferentemente, um modelo da IA concluirá uma dada tarefa quando se percorrer todos os passos descritos pelo algoritmo.

Os conexionistas afirmam que uma RNA não segue um conjunto de instruções para resolver um problema. Eles alegam que seus modelos computacionais aprendem a resolver um problema utilizando treinamento e leis de aprendizagem formuladas através de funções não algorítmicas. Como afirmam Caudill & Butler (1992;p.7): "...redes neurais aprendem a resolver um problema; elas não são programadas para resolvê-lo."

Será pela modificações nos pesos das conexões que uma RNA pode aprender a resolver problemas. Essas modificações nos pesos dão à rede um caráter dinâmico, que aparentemente não se encaixa nos padrões dos modelos da IA. Os padrões de conectividade de uma rede não são fixos no tempo, de tal forma que os pesos das conexões podem sofrer modificações com a experiência. Treinamento e aprendizagem são características fundamentais de uma RNA. A aprendizagem é alcançada, em geral, pela modificação dos pesos das conexões



entre os nódulos que compõem a rede. A lei de Hebb (1949), como sugerida pela versão de Gonzales (1996;p.279), fornece o esquema de funcionamento dos mecanismos de aprendizagem de uma RNA. Segundo essa lei:

> ...pode-se dizer que sempre que uma unidade neuronal estimular, repetidamente, outra unidade, e ambas estiverem altamente ativadas, então a conexão sináptica que se estabelece entre elas será fortalecida. Através desta lei pode-se modificar os pesos das conexões existentes fazendo com que determinadas unidades estimulem (ou inibam) umas as outras através de treinamento.

Existem três modos distintos, segundo Caudill e Butler, de treinamento para uma RNA modificar seus pesos: (1) Treinamento Supervisionado; (2) Treinamento por Reforço; (3) Treinamento não Supervisionado ou por Auto-Organização.

(1) No *treinamento supervisionado* é fornecido à rede, simultaneamente, um conjunto de *inputs* e o padrão de output que se deseja obter. A lei de aprendizagem para tal rede consiste, em geral, em mecanismos associativos que direcionam o ajuste dos pesos das conexões, de tal forma que o resultado obtido seja o mais próximo possível do desejado. Cada erro cometido pode ser aproveitado na determinação dos novos valores para os pesos das conexões.

(2) O *treinamento por reforço* consiste em fornecer um conjunto de *inputs* e (ao invés de fornecer os *outputs* desejados) também uma medida do grau de sucesso. Isto é, informa-se a rede quão bem ela está computando na realização de uma tarefa. O que pode ser feito através, por exemplo, de esquemas do tipo *teve sucesso ou não teve sucesso* ou ainda de esquemas mais precisos que informam o grau de acerto ou erro.



(3) O último modo de treinamento é denominado *treinamento não supervisionado* (ou por auto-organização). Este consiste em fornecer apenas um padrão de *inputs* sem qualquer informação sobre a resposta desejada. Cabe a RNA classificar, criar ou detectar um padrão, por meio do reconhecimento das regularidades na informação disponível no meio ambiente, sem a ajuda de um supervisor. Segundo Gonzales (1996;p.280):

> Um dos mecanismos mais simples de ajuste de pesos utilizados no treinamento não supervisionado, ou por auto-organização, é o conhecido princípio de aprendizagem por competição (ver Kohonen, 1984 para inúmeras ilustrações deste mecanismo) Segundo este princípio, uma UNS, quando ativada, pode influenciar o comportamento de suas UNS's vizinhas, inibindo ou ativando-as, desencadeando assim uma reação global que determinará o estado final da rede.

Este princípio de aprendizagem, segundo Gonzales, possibilitará a formação de um certo padrão informacional. Pois, à medida que as chamadas conexões excitatórias existentes entre os nódulos se espalharem pela rede, elas ativarão outras unidades que possuírem elementos em comum. Ao mesmo tempo, inibirão aquelas que não tiverem nenhuma afinidade com os nódulos ativados inicialmente.

Nesta dinâmica entra em cena a lei de Hebb, ou seja, sempre que um nódulo estimular repetidamente um outro, e ambos estiverem altamente ativados, reforça-se a conexão entre eles. O reforço de algumas conexões entre os nódulos ativados produz a inibição de outros tornando-os cada vez menos ativados. Esta oposição entre padrões de nódulos faz com que a rede adquira uma forma de padrão informacional. Isto é, uma dada estrutura se formará na rede.

Contudo, conforme ressaltamos, para os conexionistas, seus modelos não seguem regras pré-estabelecidas – como seguem os modelos da IA –



apesar de se comportarem como se estivessem seguindo algum tipo de regra. Para eles, os mecanismos de aprendizagem, do tipo lei de Hebb, não são considerados como regras abstratas ou programas, mas sim descrições de processos físicos que, supostamente, ocorrem no organismo. Devido a essa concepção à respeito dos mecanismos físicos de aprendizagem, os conexionistas afirmam que as RNA's processam informação no plano sub-símbólico, onde apenas impulsos elétricos são propagados. A seguir veremos dois exemplos de modelos conexionistas.

4.3 OS MODELOS DE REDES NEURAIS ARTIFICIAIS

4.3.1 A REDE DE PERCEPTRON

Após dez anos da publicação do artigo de McCulloch e Pitts ressurge, através dos trabalhos de Rosenblatt (1962), o interesse em prosseguir as pesquisas acerca das redes de neurônios artificiais. Rosenblatt procurou mostrar como as redes de neurônios de McCulloch e Pitts podem, com conexões modificáveis, ser treinadas para realizar certas funções. Tal rede recebeu o nome de *perceptrons*.

Podemos descrever o perceptron como uma rede que possui uma camada de unidades sensoriais ou, de unidades **E**, uma camada de associação ou unidades **I** e apenas uma camada de resposta ou unidade **R**. As unidades **E**, de entrada, recebem o *input* do meio ambiente. A unidade **R**, de resposta, está conectada com o meio no outro extremo do sistema e emite um sinal de saída. As



unidades **I** intermediárias, como o próprio nome diz, ficam entre o canal de *input* e de *output*, podendo estar conectadas diretamente com a unidade **R**. As unidades **R** e **I** possuem uma função limiar e uma função que calcula a soma ponderada de seus *inputs*. A atividade das unidades será determinada pelos pesos associados às conexões. A figura 4 ilustra um tipo elementar de perceptron:

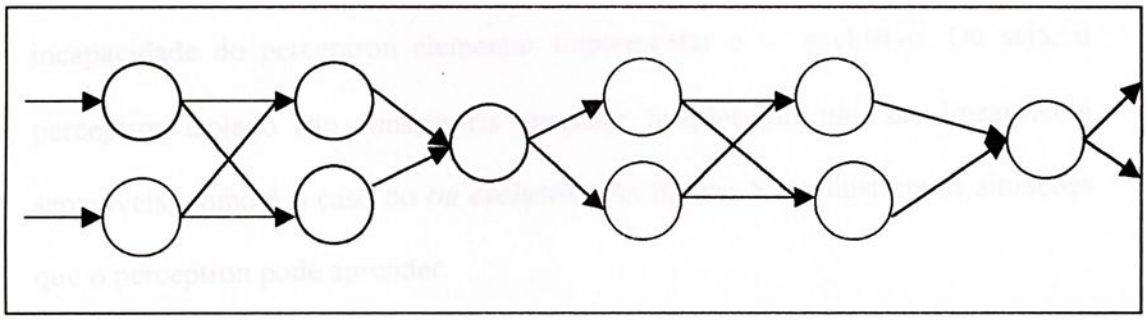


Figura 4- modelo de perceptron elementar

Rosenblatt pretendia simular com seu perceptron a capacidade dos sistemas biológicos de aprenderem um padrão. Mas qual o método usado por ele para que a rede aprenda a reconhecer um padrão? Rosenblatt imaginou uma forma de ensinar a rede artificial até que ela aprendesse o padrão desejado, através da familiarização e classificação de exemplares. O ensinamento usado foi o mais intuitivo possível: o treinamento a partir de exemplos:

> Por este método são apresentados exemplos de comportamento à rede, isto é, para um estímulo x^dz a saída deve ser y^dz, onde o índice d refere-se ao resultado desejado. Os exemplos são repassados até que a rede aprenda o comportamento correto... (Kovács,1996,p.41)

Como podemos notar, a rede de perceptrons segue o modo de treinamento denominado *treinamento supervisionado*. Será por esse modo de treinamento que a rede poderá aprender a encontrar o peso correto das conexões sinápticas que lhe permite resolver um determinado problema. O treinamento da

rede de perceptrons, através de exemplos, elimina a necessidade de um programa que especifique, passo a passo, o caminho para a solução de problemas.

As pesquisas com redes do tipo perceptron perderam força quando o trabalho de Rosenblatt foi duramente criticado por dois influentes pesquisadores da IA, a saber, Minsky e Papert no livro *Perceptrons* (1969). A objeção levantada por eles é que tornou-se muito conhecida, diz respeito à incapacidade do perceptron elementar implementar o ∨' exclusivo. Ou seja, o perceptron isolado não conseguiria aprender funções que não são linearmente separáveis, como é o caso do *ou exclusivo*. As figuras 5 e 6 ilustram as situações que o perceptron pode aprender:

Nesta figura 5 temos o perceptron reconhecendo situações que se ajustam à conjugação lógica "Λ". Apresentamos a seguinte tarefa para a rede: separar, por uma reta, cubos grandes e pesados de círculos pequenos e leves.

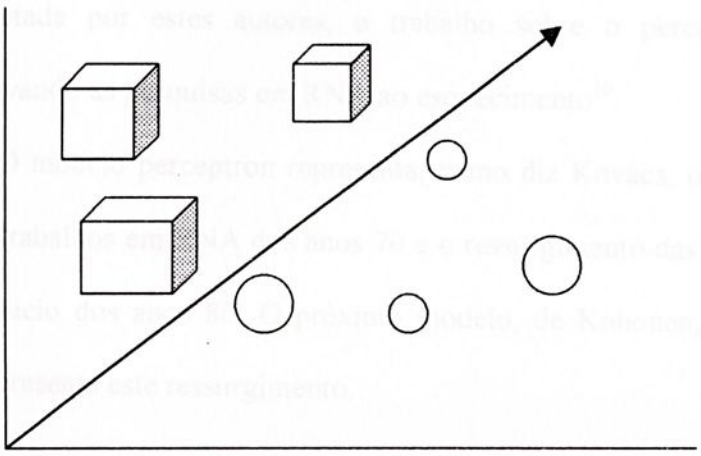


Figura 5 - O percepton, neste caso, consegue traçar a linha para separar as figuras.

Na próxima figura veremos a rede de perceptron atuando para representar a função lógica '∨'. A tarefa apresentada a esta rede será: separar, por uma reta, figuras grandes das pequenas.

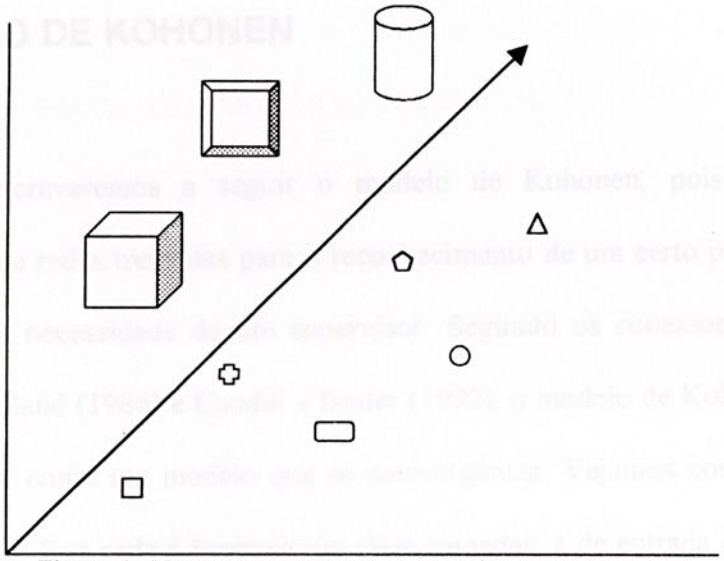


Figura 6 - Neste caso o perceptron, também, consegue traçar uma linha que separa as figuras grandes das pequenas.

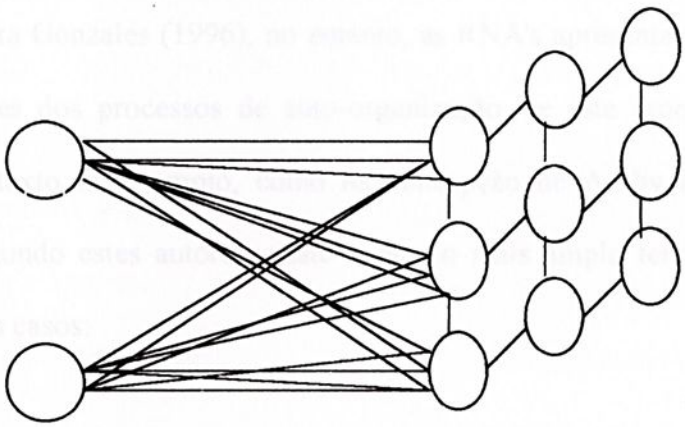
Minsky e Papert ressaltam, nas suas críticas ao perceptron, que, apesar do sucesso na classificação das tarefas ilustradas acima, esta rede apresenta dificuldades na implementação da função lógica '$\vee$' exclusivo. A partir desta dificuldade, levantada por estes autores, o trabalho sobre o perceptron foi perdendo força, levando as pesquisas em RNA ao esquecimento[10].

O modelo perceptron representa, como diz Kovács, um divisor de águas entre os trabalhos em RNA dos anos 70 e o ressurgimento das pesquisas neste campo no inicio dos anos 80. O próximo modelo, de Kohonen, que será descrito abaixo representa este ressurgimento.

[10] O modelo de perceptron multi-camadas, com a função limiar, uma grande novidade nestes modelos, resolve o problema do '$\vee$' exclusivo.

4.3.2 O MODELO DE KOHONEN

Descreveremos a seguir o modelo de Kohonen, pois este representa um caso de redes treinadas para o reconhecimento de um certo padrão informacional sem a necessidade de um supervisor. Segundo os conexionistas, (Rumelhart e McClelland (1986) e Caudill e Butler (1992), o modelo de Kohonen pode ser classificado como um modelo que se auto-organiza. Vejamos como se estrutura esse modelo. Esta rede é formada por duas camadas: a de entrada e a de Kohonen, sendo a última camada a própria de saída, conforme ilustrada na figura 7:



Camada de entrada *Camada de Kohonen*

Figura 8 - Ilustração da rede de Kohonen (1984)

O modelo de Kohonen atua segundo o princípio de *aprendizagem por competição*. Esta forma de aprendizagem, como já mencionamos, atua nos nódulos fazendo com que aqueles ativados "influenciem" o estado dos nódulos vizinhos, inibindo-os ou ativando-os gerando uma reação no estado final da rede. Como este modelo realiza o treinamento sem supervisor, nenhum padrão especifico é oferecido na entrada da rede. Ou seja, os valores iniciais dos pesos das conexões entre os nódulos serão arbitrariamente

estabelecidos, deixando que a rede, a partir de vários ciclos de interações entre os nódulos e o meio ambiente, se acomode e forneça um certo padrão.

Espera-se que as redes treinadas, sem um supervisor, aprendam a encontrar o peso correto das conexões sinápticas que lhes permita resolver uma tarefa determinada, qual seja, representar algum padrão informacional disponível no meio ambiente. Os pesos das conexões serão determinados pela própria dinâmica da rede. Devido a essa ausência de programa que determine a operação da rede, os conexionistas, como Caudill e Butler, sugerem que este tipo de rede se auto-organiza.

Para Gonzales (1996), no entanto, as RNA's apresentam apenas algumas propriedades dos processos de auto-organização, se este processo for entendido num contexto mais amplo, como na concepção de Ashby (1962) e Debrun (1996). Segundo estes autores, neste contexto mais amplo temos *auto-organização* em dois casos:

A) Quando elementos independentes interagem criando-se uma dependência mútua entre eles de modo a formar uma estrutura ou sistema. Dá-se a este processo o nome de *auto-organização primária.*

B) No segundo caso, quando um sistema *já formado* evolui a partir de seus próprios mecanismos. Este processo recebe o nome de *auto-organização secundária*.

Segundo Gonzales (1996;p.284) as RNA's apresentam o primeiro caso de auto-organização:

> Um exemplo de auto-organização primária pode ser concebido em uma rede neural cuja UNS's encontram-se, inicialmente, conectadas de forma aleatória, sem qualquer relação de dependência existente entre elas. A medida que a rede interage com as invariantes do meio ambiente, valores de ativação serão estabelecidos nas suas unidades vizinhas. Com o treino não supervisionado, a rede eventualmente convergirá para um estado tal em que os pesos das conexões entre as suas UNS's se manterão estáveis, dando lugar a um padrão de conectividade, que será entendido pelos conexionistas como um modelo de representação mental do objeto ou invariante. ... podemos considerar que a rede se auto-organizou primariamente, uma vez que uma estrutura foi criada a partir de elementos inicialmente independentes, sem a intervenção de um centro organizador.

Caberia aqui questionar o segundo caso de auto-organização: como ele se dá em uma RNA? Poderiam as RNA's criarem, por si mesmas, mecanismos de evolução? Uma situação que os conexionistas consideram como a possibilidade de evolução de uma RNA é a auto-correção através do *feedback*. A RNA que apresentar uma resposta inadequada a um problema determinado pode, a partir de reiteradas exposição ao meio ambiente e vários ajustes de peso, fornecer o padrão correto. Neste caso teria a rede evoluído? As repetidas exposições ao meio, bem como os diversos ajustes dos pesos parecem levar a rede a criar apenas hábitos e não um conhecimento que lhe possibilite resolver problemas.

O hábito, como sabemos, não traz, por si só, evolução pois, quando praticamos uma certa ação por mero hábito não conseguimos extrair destas lições elementos para corrigirmos os erros que cometemos para atuarmos com mais segurança. Sendo assim, começamos a esboçar um problema central nos modelos conexionistas, qual seja, não possuir a capacidade de generalização forte ou universal. Os padrões representados pela RNA são padrões locais, disponíveis em um dado *input*. Tomemos como exemplo a seguinte tarefa: reconhecer a letra 'A'



de forma. Nesta tarefa a RNA oferecerá como padrão de saída vários tipos de 'A's porém, ela não reconhecerá como sendo tal letra, este tipo 'a'.

As RNA's não possuem, portanto, uma capacidade fundamental para a aquisição de conhecimento, a saber, adquirir conceitos universais. Podemos dizer que uma RNA possui, apenas, a capacidade de adquirir hábitos. No que se segue, continuaremos apresentando algumas outras críticas a estes modelos.

4.4 CRÍTICAS AOS MODELOS CONEXIONISTAS

Nesta seção apresentaremos duas críticas aos modelos de RNA. A primeira foi elaborada por Dreyfus (1975) (1993), que sustenta que o corpo ten um papel importante no comportamento inteligente e que a ausência deste elemento fisíco/biológico impossibilita o sucesso na atividade de reconhecer padrões. A segunda crítica, elaborada pelos autores do livro *Computadores, Mentes e Conduta* (1998), diz respeito à noção de *reconhecimento de padrões*, empregada nos modelos de RNA.

4.4.1 O PAPEL DO CORPO HUMANO NO RECONHECIMENTO DE PADRÕES

As RNA's são amplamentes empregadas em tarefas que exigem o reconhecimento de padrões. Sabemos que tais sistemas são desprovidos de corpo biológico, e este é, segundo Dreyfus (1975), o ponto frágil destes modelos

que pretendem simular a atividade de reconhecimento de padrões efetuada pelos humanos. Para Dreyfus, é no tocante ao reconhecimento de padrões que o corpo é imprescíndivel. Tomando emprestados conceitos da fenomenologia, ele procura mostrar que o reconhecimento de um padrão requer uma "espécie de antecipação indeterminada e global". Para nós humanos, dotados de corpo, o reconhecimento de certos objetos se dá, não pelas suas características especiais e isoladas, mas, sim, pela percepção de todo o conjunto no qual os objetos se incluem. Percepção essa que é garantida pela forma que nosso corpo experimenta. Diz Dreyfus (1975;p.212):

> ...ao reconhecermos uma melodia, as notas ganham seus valores ao serem percebidas como parte da melodia, não é a melodia que é reconhecida em função de notas identificadas isoladamente. De igual maneira, na percepção dos objetos não existem traços neutros. A mesma camada nevoenta que eu veria como poeira se julgasse estar diante de uma maçã de cera poderia me parecer orvalho se pensasse estar vendo uma maçã de verdade. O significado dos detalhes e, de fato, a sua própria aparência, é determinado pela percepção que eu tenho do conjunto.

O grande problema apontado por Dreyfus é a construção de um dispositivo tão versátil que simule nossa percepção corporal. Graças ao nosso corpo é que podemos, diz ele, captar os ritmos sonoros. Nosso conjunto corpóreo, constituído por nervos e músculos, parece ser apropriado para responder aos estímulos sonoros. Não existe uma regra capaz de descrever este processo.

Mesmo nos casos onde atuamos com o auxílio de regras, como no caso de dirigir um veículo, dançar ou mesmo falar uma língua estrangeira, nós, aos poucos, vamos transferindo o controle para o corpo. Ninguém ao dirigir um veículo fica, a todo instante, repassando as regras que permitem manter o carro em movimento. Nós simplesmente guiamos. Na percepção, diz Dreyfus, ocorre o

unesp

mesmo. Aprendemos a sentir uma dada textura, deslizando as mãos ou pés sobre ela e, a partir disso, construímos certas expectativas. Sem essa habilidade de experimentar, proporcionada pelo corpo, recolhemos apenas sensações confusas.

Para adquirir essa habilidade corporificada, um modelo computacional, além de um corpo, teria que:

> ...ser programado para fazer uma relação específica das características de um objeto visualmente analisado, e comparar essa lista com uma lista explícita de características registradas, movimentando receptores táteis sobre o objeto. (Dreyfus 1975;p.226)

O corpo humano dispensa toda essa análise formal. Sua constituição é tal que podemos experimentar uma habilidade de muitas maneiras distintas. Quando percebemos que adquirimos uma certa habilidade, temos a capacidade de responder, com um determinado tipo de solução, a situações de uma determinada forma geral. Isto é, conseguimos generalizar nossa habilidade para diversas situações.

Como ficariam, neste tipo de análise, os modelos de RNA's que são desprovidos de corpo biológico e não possuem nenhuma regra explícita para dirigir sua análise? Dreyfus alega que nestes modelos podem surgir problemas ao se procurar treinar as redes para reconhecer um certo padrão no meio ambiente. Para ele, estes sistemas não terão sucesso se algumas hipóteses não forem pré-especificadas. Em seu livro *What Computers Still Can't Do* (1993), ele descreve um teste onde redes neurais estavam sendo treinadas para reconhecer tanques numa floresta. Mostraram fotos de uma floresta , primeiramente com tanques e em seguida sem estes. O resultado foi satisfatório. Para confirmar que a rede aprendeu a reconhecer florestas com e sem tanques tiraram mais um lote de fotos e as

apresentaram, novamente, à rede. Dessa vez a rede fracassou. Mas por que a rede fracassou? Segundo Dreyfus (1993;p.xxxvi):

> ...o mistério só foi resolvido quando se notou que as primeiras fotos da floresta sem tanques tinham sido tiradas num dia nublado, e as com tanques num dia de sol. A rede aparentemente aprenderá a reconhecer e generalizar a diferença entre uma floresta com e sem sombras.

Qual seria, segundo Dreyfus, a solução para resolver este problema de reconhecimento de padrão? Seria necessário, diz ele, restringir a rede a uma classe de respostas apropriadas. Ou seja, apresentar um certo contexto, para que a rede possa reconhecer, de forma correta, o padrão desejado. Porém, tal reconhecimento ficaria restrito ao contexto determinado de antemão. Aceitariam os conexionistas esta sugestão de pré - especificar o contexto? Procurariam eles adaptar em seus modelos um corpo biológico? Se eles escolherem a segunda opção o caminho torna-se difícil, diz Dreyfus, pois esta habilidade de processar informação do todo para as partes parece ser uma habilidade mágica do ponto de vista da ciência. Neste sentido ele afirma:

> O que é importante a respeito das habilidades é que, embora a ciência exija que o desempenho eficiente seja descrito de acordo com regras, essa regras não estão absolutamente implícitas na produção do desempenho. (Dreyfus 1975;p.231)

Ou seja, os seres humanos são capazes de reconhecer e classificar certos padrões a partir de esquemas indeterminados. Talvez o atributo *inteligente* nos seja imputado devido a essa habilidade de comportar-nos, sem regras explícitas, em situações novas, aproveitando ao máximo nossas experiências passadas.

4.4.2 O PROBLEMA DO RECONHECIMENTO DE PADRÕES, A PARTIR DO TREINO, EM REDES NEURAIS ARTIFICIAIS

As RNA, como já mencionamos, são treinadas para reconhecer um determinado padrão. Algumas redes são treinadas para aprender uma certa linguagem Como procederia o treinamento destas redes? Em geral, o seu treinamento segue o modelo de aprendizado por ostensão. Isto é, são apresentados vários exemplos à rede, a fim de que ela encontre os pesos adequados das conexões que lhe permitiriam associar o padrão de entrada ao padrão de saída estabelecido.

Button, Coulter, Lee e Sharrock (1998), apoiados na teoria da linguagem de Wittgenstein, argumentam que os conceitos assumem diversas formas, exigindo assim, uma pluralidade de práticas de aquisição. Não basta, segundo eles, uniformizar o treinamento de uma rede para aprender conceitos. Apenas o treinamento, a partir de exemplos, não consegue captar a riqueza dos conceitos. Expor uma rede a muitos casos do mesmo tipo, até que este possa ser reproduzido com fidelidade, não ajuda tal sistema no aprendizado de uma língua. O que estes autores afirmam é que:

> ...devemos lembrar que o que está sendo treinado não é um cérebro, mas sim um ser humano; o que um ser humano esta sendo treinado para fazer sob a rubrica de 'aprender o que uma palavra significa' (mesmo uma palavra referencialmente utilizável) é muito mais do que repeti-la corretamente na presença de um caso do fenômeno nomeado... (Button et al.1998;p.172)

Estes autores querem explicitar um erro nas hipóteses conexionista, a saber, restringir o treinamento para a aquisição de conceitos à atividade de associação entre neurônios. Para eles, não são os neurônios que devem ser treinados para adquirir um conceito ou mesmo aprender a fazer algo. Aliás os neurônios não podem fazer ou aprender coisa alguma. Quem pode ser treinado para aprender são as pessoas, pois estas podem aprender com o meio e interferir nele para modificá-lo. Além de modificar o seu meio, elas também podem ensinar as outras pessoas a fazê-lo e assim por diante.

Button, Coulter, Lee e Sharrock sugerem que *a expressão treinamento de uma rede a partir de exemplos* deve ser diferenciada do uso que fazemos quando falamos em treinar determinadas pessoas para fazer algo. O aprendizado humano envolve fatores do tipo relação social, envolvimento corporal, etc., que não fazem parte do treinamento de uma RNA. Sendo assim, eles concluem, nenhuma tese psicológica sobre o aprendizado pode ser extraída do treinamento das RNA's, pois o uso do termo *treinamento* nas redes é completamente diferente do uso que nós fazemos deste termo.

Porém, Button, Coulter, Lee e Sharrock argumentam que a maior confusão da teoria conexionista está presente na noção de *reconhecimento*. Para eles, novamente, não podemos dizer que uma RNA reconhece um certo padrão como nós reconhecemos. O funcionamento de uma RNA nada explica acerca da capacidade humana de reconhecer um certo padrão.

Existem várias propriedades presentes no processo de reconhecimento humano de padrões que não se ajustam à definição de reconhecimento presente na teoria conexionista:



> ...reconhecer algo significa freqüentemente, mas nem sempre, estar correto sobre o que ele é....Pode-se também reconhecer algo específico sem sequer tê-lo visto antes...,e pode-se reconhecer algo como tendo sido visto antes, sem se saber exatamente o que seja. Além disso, pode-se dizer sinceramente que se reconhece, realmente pensar que se reconheceu algo ou alguém e estar enganado...(Button et al.1998;p.174)

Reconhecer, nestes casos, não é o mesmo que ser guiado pelos ajustes de pesos e acomodação até um certo padrão. Reconhecer, segundo estes autores, é uma realização que assume vários aspectos. Dentre estes aspectos podemos citar: saber quais são as características fundamentais do objeto que se quer reconhecer, em que contexto esta disposto tal objeto, retirar do erro do reconhecimento lições para que, em uma próxima ocasião, se reconheça corretamente (reconhecer o erro como um aspecto do reconhecimento), entre outros.

Devido a estes vários aspectos é que para Button, Coulter, Lee e Sharrock :

> Os próprios cérebros nada reconhecem, as pessoas sim. Um sistema computacional, portanto, no sentido humano de 'reconhecer' 'nada reconhece': nós é que reconhecemos o que queremos que o computador 'reconheça' no sentido especificamente técnico de produzir um output que desejamos dele, e que nós podemos reconhecer como o 'output' correto dadas as circunstâncias (Button et al.1998;p.175)

Segundo os autores, nós reconhecemos algo mesmo sem ter certeza, mesmo errando. Não basta uma descrição puramente neurofisilógica para explicar como reconhecemos. Pois quem reconhece são as pessoas e não seus cérebros.

Em resumo, neste capítulo apresentamos alguns a descrição geral de RNA. Em seguida, descrevemos os modelos de perceptron e Kohonen. Para finalizar, esboçamos algumas críticas a estes modelos. Na próxima seção apresentarmos nossas considerações finais.

unesp

CONSIDERAÇÕES FINAIS

Procuramos mostrar neste trabalho que a ciência cognitiva é uma área nova de pesquisa onde muito, ainda, está por se fazer. Nossa dissertação procurou refletir, principalmente, como está se formando este novo campo de pesquisa, quais são os pressupostos filosóficos das suas mais estabelecidas linhas de pesquisa, IA e RNA e qual é o alcance e os limites dos modelos destas linhas de pesquisa

Após este percurso sobre a ciência cognitiva uma questão deve ser respondida: o que podemos aprender sobre a natureza da cognição, a partir das pesquisas desenvolvidas neste campo de investigação?

Podemos constatar que a tarefa que a ciência cognitiva reivindica para si é árdua. Como diz Thagard (1998), parece ser um grande, ou talvez o grande, desafio da humanidade juntar as diversas peças do maior quebra-cabeças que os seres humanos já tentaram montar; ou seja, o quebra-cabeça que nos forneça uma explicação bem fundamentada sobre a natureza e estrutura da cognição.

Mas este desafio esbarra, ao nosso ver, em sérios problemas. Entre estes podemos citar aqueles relativos à natureza das emoções e da consciência. Parece que, apesar das varias abordagens presentes na ciência cognitiva, tais aspectos escapam do tratamento naturalista. O debate sobre a naturalização da consciência vem, há alguns anos, dominando às discussões na filosofia da mente. Ao nosso ver, os instrumentais metodológicos/conceituais de que se dispõe atualmente na ciência cognitiva não permitem um tratamento



adequado deste fenômeno. Isto porque, a consciência, na sua característica mais fundamental, implica na construção do 'distinto de si'. Conforme vimos, os modelos de IA e RNA são sistemas preparados para resolver problemas locais, dispensando a noção de sujeito. A medida que a ciência cognitiva retira de cena tal noção, não se pode atingir a característica mais fundamental deste fenômeno. Em outras palavras, os modelos da ciência cognitiva, não edificam 'o outro'.

Outro fenômeno de difícil tratamento pela abordagem naturalista que vigora na ciência cognitiva é a linguagem. Apesar do crescente avanço nesta área conquistado pela abordagem cognitivista, um fator de suma importância fica sempre à margem, a saber, o social. A ciência cognitiva devido a sua herança analítico-positivista privilegia, em suas investigações, o contexto físico e biológico em detrimento do social. Como desprezar, no caso da linguagem, a importância fundamental das interações sociais? Não pretendemos afirmar que a linguagem não seja um fenômeno que pode ser explicado unicamente através de regras. Apenas queremos chamar a atenção para o fato de que ela é, também, um fenômeno social.

Por meio das interações sociais é que construímos uma parte substancial da nossa linguagem e, em certos casos, o significado de nosso discurso. Interações estas que são facilitadas pela nossa estrutura biológica. Sendo assim, o problema que deixamos para os cientistas cognitivos é: como simular em um modelo desprovido de um corpo biológico a intensa dinâmica da relações que possibilitam a construção da linguagem?

A linguagem e a consciência são aspectos da cognição que parecem, ainda, apresentar sérias dificuldades ao empreendimento científico

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

cognitivista. Apesar destas dificuldades o que podemos extrair de positivo da pesquisa cognitivista?

A resposta a tal questão envolve, entre outras coisas, uma reflexão sobre a formação deste novo campo interdisciplinar de pesquisa. Assinalamos que esta nova ciência da mente está se configurando a partir de dois programas de pesquisa: um representacionista e outro não representacionista. Tal divisão leva ao seguinte impasse: qual é o paradigma dominante nesta ciência? O representacionista, com suas vertentes IA e RNA, ou o não representacionista com sua proposta dinâmica de estudo da cognição?

Nossa resposta, a esta questão, é: os dois. Para nós, apesar das profundas diferenças entre os dois programas, estes podem ser aproximados a partir de duas teses em comum, a saber, a tese mecanicista da natureza da cognição e a intedisciplinariedade como metodologia apropriada para estudar tal natureza. Tal aproximação amenizaria a possível incomensurabilidade que se estabeleceria entre os diferentes programas da ciência cognitiva e se formaria um único paradigma de pesquisa.

Além de amenizar a incomensurabilidade entre os programas, tal aproximação seria extremamente enriquecedora uma vez que, estabelecido o paradigma, se poderia abordar os vários aspectos da cognição a partir de perspectivas distintas. Quando o fenômeno investigado fosse, por exemplo a capacidade de resolver problemas, se escolheria as técnicas lógico/abstrata da IA. Caso contrário, se uma dada tarefa envolvesse representações, mas não fosse passível de abordagem algorítmica, por exemplo o reconhecimento de padrões, se escolheria as redes neurais artificiais. Por fim, se o objeto de análise escapasse de

uma abordagem representacionista, como parece ser o caso da atividade motora, se optaria por uma técnica fundada em atratores e equações não lineares. Isto é, o fenômeno seria enfocado a partir de uma abordagem dinâmica.

Para vislumbrar nossa proposta de unificação paradigmática desta nova ciência, podemos adaptar e empregar a metáfora dos três olhares usada por Morin: num primeiro momento, a cognição aparece como algo caótico e intocável a partir dos padrões científicos vigentes. Em um segundo momento, as abordagens sobre a cognição se concentram em modelos representacionais fundados sobre regras ou padrões de conectividade. Nesta etapa de investigação o enfoque representacional apresenta-se como uma ferramenta útil para estudar a cognição humana. Um terceiro olhar nos apresenta a cognição como um sistema dinâmico que contém um grande número de variáveis. Neste momento, conceitos como os de *não linearidade*, *atratores e caos* entram em cena. O desafio que se impõe a esta nova ciência da mente é, conjugando estes três olhares, fornecer um explicação mais acabada da natureza dos processos cognitivos.

Um fator positivo que pode ser ressaltado no trabalho científico cognitivista, unificado a partir desta proposta, diz respeito à analogia entre a mente e as máquinas. Para muitos críticos tal analogia pode, no limite, igualar a mente a um programa de computador. Porém, acreditamos que, se manejada com cuidado, isto é, se concebermos a tese mecanicista apenas como uma ferramenta útil para explorar as semelhanças e diferenças entre objetos distintos, como a mente e a máquina, esta poderá ser de grande valia para os estudos sobre a cognição. Mesmo os críticos não podem negar que, a partir desta analogia, se conseguiu uma gama de hipóteses à respeito das estruturas e processos mentais.

unesp

De certa forma, esta analogia, mente/máquina, através da delimitação crítica de seus limites, representou um avanço frente as abordagens behaviorista e dualista sobre a cognição.

Outro fator positivo advindo desta abordagem pluralista que, ao nosso ver, deve ser ressaltado são as integrações advindas desta nova *interdisciplinariedade*. Seria importante que diversas áreas do saber fossem unificadas em termos de conceitos compartilhados. Esta unificação conceitual pode fornecer boas teorias explicativas. Tomemos, por exemplo, a linguagem, citada anteriormente como um aspecto problemático para a explicações científico-cognitivista. Talvez a solução para explicar satisfatoriamente um fenômeno tão complexo como este seja combinar os dados obtidos em diversos campos, como a neurologia, a psicologia, antropologia e sociologia. O ganho com esta nova interdisciplinariedade seria duplo: a) o objeto de estudo seria abarcado em sua complexidade, pois teríamos uma pluralidade de abordagens; b) estariam ampliando-se os contextos de investigação, privilegiaria-se não só o contexto físico e biológico, mas também o contexto social e cultural.

Esta proposta de unificação paradigmática da ciência cognitiva possui a vantagem de responder àqueles críticos que julgavam que, nem a proposta da IA ou da RNA, por serem representacionistas, não seriam adequadas para o estudo de determinados aspectos da cognição. A estes críticos seria possível responder que, se as propostas destas linhas de pesquisa não são adequadas, podemos optar por uma abordagem não representacionista, presente no programa dinâmico da ciência cognitiva

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

Tal proposta se apresenta, ainda, pouco elaborada. Porém, acreditamos que, se bem refletida e explorada, pode indicar um bom caminho a ser trilhado rumo à uma explicação satisfatória da natureza e estrutura da cognição.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRANTES, P. (Org.) *Epistemologia e cognição*. Brasília: Editora da Universidade de Brasília, 1994.

ASHBY, R. Principles of the self organizing system. In: FOERSTER, H., ZOPF JUNIOR, G.W. (Orgs.) *Principles of self-organization*. Oxford: Pergamon Press, 1962. p.255-78.

BODEN, M. *Artificial intelligence and natural men*. Nova York: Basic Books,1977.

BUTTON, G., COULTER, J., LEE, J., SHARROCK, W. *Computadores, Mentes e Conduta*. Trad. R. L. Ferreira. São Paulo: Fundação Ed. Unesp, 1998.

BRETON, P. *História da informática*. Trad. E. Fernandes. São Paulo: Ed. Unesp, 1991.

CAUDILL, M., BUTLER, C. *Understanding neural networks*. Cambridge: MIT, 1992. v. 1

COLBI, M. *Artificial paranoia*. Nova York: Pergamon Press, 1975.

DE LA METTRIE, J. *Hombre máquina*. In: Obra Filosófica. *Madrid*: Editora Nacional, 1983.

DEBRUN, M. A idéia de auto-organização. In: DEBRUN, M., GONZALES, M. E. Q., PESSOA JUNIOR, O. (Orgs.) *Auto-Organização:* estudos interdisciplinares. Campinas: Centro de Lógica, Epistemologia e História da Ciência. UNICAMP, 1996. p.3-23.

DENNET, D. *Consciousness explained*. Little: Brown and Compagny,1991.

_____. *Kinds of minds*. Cambridge: MIT, 1996.

DESCARTES, R. *Meditações*. Trad. J. Guinsburg e B. Prado Júnior. São Paulo: Abril Cultural, 1973.

DUPUY, J.P. *Nas origens das ciências cognitivas*. Trad. R. L. Ferreira. São Paulo: Ed. Unesp, 1996.

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

_____. *What computers still can't do*. Cambridge: MIT, 1993.

GARDNER, H. *A nova ciência da mente*. Trad. C. Malburguer. São Paulo: Ed. da Universidade de São Paulo, 1995.

GONZALES, M. E. Q. LUNGARZO, C., A., MILIDONI, C., B., PEREIRA JÚNIOR, A., WRIGLEY, M., B. (Orgs) *Encontro com as ciências cognitivas*. Unesp: Marília Publicações, 1996.

_____. Ação, causalidade e ruídos nas redes neurais auto-organizadas. In: DEBRUN, M., GONZALES, M. E. Q., PESSOA JUNIOR., O. (Orgs.) *Auto-Organização:* estudos interdisciplinares. Campinas: UNICAMP, Centro de Lógica, Epistemologia e História da Ciência. UNICAMP, 1996. p.273-97.

HARTLEY, R. *Theory of recursive functions and effective computability*. Cambridge: MIT, 1988.

HEBB, D. Organization of behavior. In: ANDERSON, J. A., ROSENFELD, E. (Orgs.) *Neurocomputing*. Cambridge : MIT, 1998. p.45-46.

HINTON, G. Learning and relearning in boltzmann machines. In: RUMELHART, D., McCLELLAND, J. (Orgs.) Parallel distributed processing: explorations in the microstructure of cognition. Cambridge: MIT, 1986. v. 1

KOHONEN, T. *Self-organazing and associative memory.* Berlim: Springer, 1984.

KOVÁCS, Z. *Redes neurais artificiais:* fundamentos e aplicações. São Paulo: Edição Acadêmica, 1996.

KUHN, T. *As estruturas das revoluções cientificas*. Trad. B. Boeira e N. Boeira. São Paulo: Perspectiva, 1997.

HUMPHREY, N. *Uma história da mente.* Trad. W. Dutra. Rio de Janeiro: Campus, 1994.

JOHNSON-LAIRD, P. *The computer and the mind.* Cambridge: Harvard University Press, 1988.

LYCAN, W. *Consciousness*. Cambridge: MIT, 1987.

MINSKY, M. Inteligência artificial In: *Computadores e Computação*. São Paulo: Editora Perspectiva, 1976. p. 151-160.

_____. *Computation*: finite and infinite machines. Prentice Hall. Englewood Cliffs,1967.

MINSKY, M., PAPERT, S. *Perceptrons*. Cambridge: MIT, 1969.

McCULLOCH, W., PITTS, W. A logical calculus of the ideas immanent in nervous activity. In: ANDERSON, J. A., ROSENFELD, E. (Orgs.) *Neurocomputing*. Cambridge: MIT, 1998. p.18-27.

MORIN, E. *Ciência com consciência*. Trad. M. Alexandre e M. A. S. Doria. Rio de Janeiro: Bertrand Brasil, 1996.

OLIVEIRA, M. B. de. O que todo cientista cognitivo deve saber sobre a lógica. In: GONZALES M. E. Q. et al. (Org.) *Encontro com as ciências cognitivas*. Marília: Faculdade de Filosofia e Ciências, 1996. p.15-27.

_____. A ciência cognitiva e a robotização do homem. (versão preliminar) In: *Da ciência cognitiva à dialética*. Tese de Livre Docência. 1998. P.1-15

PAPERT, S. Seymour Papert, ou a Tartaruga Logo. In: PESSSIS-PASTERNAK, G. (Org) *Do Caos à Inteligência Artificial*. Trad. L. P. Rounet. São Paulo: Ed. Unesp. 1993. p. 243-253.

PENROSE, R. *A mente nova do rei*. Trad. W. Dutra. Rio de Janeiro: Campus, 1993.

PUTNAM, H. Mind, language and reality. Cambridge: MIT, 1975.

_____. The nature of mental states In: BLOCK, N. (Org) *Readings in the Philosophy of Psychology*. 1980, p.223-31.v.1

_____. Representation and reality. Cambridge: MIT, 1988.

_____. Renewing philosophy. Cambridge: MIT, 1995.

ROSENBLATT, F. *Principles of Neurodynamics*: perceptrons and theory of brain mechanisms. Spartan Books, 1962.

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

RYLE, G. *The concept of mind.* Londres: Barnes & Nobles, 1949.

RUMELHART, D., McCLELLAND, J. (Orgs) *Parallel distributed processing*: explorations in the microstructure of cognition. Cambridge: MIT, 1986. v. 1.

_____. *Philosophy and connectionist theory*. Hillsdale: Lawrence Erlbaum, 1991.

SCHANK, R, BIRNBAUM, L. Aumentando a inteligência. In: KHALFA, J. (Org.) *Natureza Da Inteligência*. Trad. L. P. Rouanet. São Paulo: Ed. Unesp, 1996.

SEARLE, J. *Mente, cérebro e ciência*. Trad. A. Morão. Lisboa: Edições 70, 1987.

_____. *A redescoberta da mente*. Trad. E. Pereira e Ferreira. São Paulo: Martins Fontes, 1997.

SIMON, H. Herbert A. Simon, o computador-rei. In: PESSIS-PASTERNAK, G. (Org) *Do Caos à Inteligência Artificial.* Trad. Luiz Paulo Rounet. São Paulo: Ed. Unesp, 1994.

THAGARD, P. *Mente:* uma introdução à ciência cognitiva. Trad. M. R. Hofmeister. Porto Alegre: ArteMed, 1998.

THELEN, E., SMITH, L., B. *A dynamic systems approach to the development of cognition and action*. Cambridge: MIT, 1994.

TURING, A. On computable numbers with application to the entschudungsproblem. *Proceedings of the London Mathematical Society,* v. 42. p. 230-235, 1936.



_____. Computadores e inteligência. In: EPSTEIN, I.(Org) *Cibernética e comunicação.* São Paulo: Cultrix/Edusp, 1973. p. 45-82.

VAN GELDER, T. *The dynamical hypothesis in cognitive science*. Behavioral and Brain Sciences. Cambridge: MIT, 1998.

VON ECKARDT, B. *What is cognitive science?* Cambridge: MIT, 1995.

VON NEUMANN, J. *The computer and the brain*. New Haven: Yale University Press, 1958.

WALDROP, M. SOAR: A Unified Theory of Cognition. *Science*, v. 241. p. 296-298, 1988.

WIENER, N. *Cibernética e sociedade:* o uso humano de seres humanos. Trad. J. P. Paes. São Paulo: Cultrix, 1990.


ILEUS
ENCUADERNADORA
Faxe (014) 433 66 86 -Marília-SP