

Marcos Antonio Alves

MECANICISMO E INTELIGÊNCIA

Um Estudo Sobre O Conceito De Inteligência
Na Ciência Cognitiva

Marília
1999



MARCOS ANTONIO ALVES

MECANICISMO E INTELIGÊNCIA

**UM ESTUDO SOBRE O CONCEITO DE INTELIGÊNCIA NA
CIÊNCIA COGNITIVA**

1410000996



+ 996

MECANICISMO E INTELIGÊNCIA

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).

Marília
1999



Nº cham 153.4 FFC
A 474m

Nº Tombo 996-lex
Aquisição 5
Preço R\$5,00
Data 17/08/00
Proced. autor

A474m Alves, Marcos Antonio.

Mecanicismo e inteligência: um estudo sobre o conceito de inteligência na ciência cognitiva / Marcos Antonio Alves. – 1999.

193 p. ; 30 cm.

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

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

CDD 20 – 153.4

MARCOS ANTONIO ALVES

MECANICISMO E INTELIGÊNCIA

**UM ESTUDO SOBRE O CONCEITO DE INTELIGÊNCIA NA
CIÊNCIA COGNITIVA**

COMISSÃO JULGADORA

DISSERTAÇÃO PARA OBTENÇÃO DO GRAU DE MESTRE

Presidente e Orientador: Prof. Dr. Cosme Damião Bastos Massi,
Departamento de Filosofia – FFC -UNESP

2º Examinador: Profa. Dra. Plácida Leopoldina Amorim da Costa
Santos, Departamento de Bibioleconomia - FFC – UNESP.

3º Examinador: Profa. Dra. Maria Eunice Quilici Gonzales,
Departamento de Filosofia – FFC -UNESP

4º Examinador (Suplente): Profa. Dra. Carmen Beatriz Milidoni,
Departamento de Filosofia – FFC -UNESP

5º Examinador (suplente): Profa. Dra. Mariana Cláudia Broens,
Departamento de Filosofia – FFC -UNESP

DADOS CURRICULARES

MARCOS ANTONIO ALVES

Nascimento: 20.5.1975 – Santa Rosa, R.S. (BR)

Filiação: Clecio Alves

Jandir Alves

1993-1995: Curso de Filosofia na Universidade Católica de Pelotas
– UCPel – Pelotas, R.S.

1996-1999: Curso de Pós-Graduação em Filosofia, nível de
mestrado, na Faculdade de Filosofia e Ciências, Unesp,
Área de Concentração: Ciência Cognitiva e Filosofia da
Mente.

*À minha família
À Marisa*

Cujo apoio e incentivo me
proporcionaram chegar até aqui.

Agradecimentos

A realização deste trabalho só foi possível graças à colaboração direta ou indireta de muitas pessoas e instituições. Manifesto minha gratidão a todas elas e de forma particular:

à UNESP e à FAPESP pelo apoio institucional e pelas bolsas de estudo concedidas;

aos examinadores participantes da banca, cujas contribuições fizeram com que esta dissertação contivesse menos erros do que já contém;

ao meu orientador, Prof. Dr. Cosme Damião Bastos Massi que me acompanhou desde o início deste mestrado. Suas leituras críticas e atentas deste trabalho, função necessária de um bom orientador, contribuíram na sua melhor apresentação;

aos professores deste mestrado. O ambiente acolhedor e amistoso deste programa auxiliam no progresso de seus componentes. Meu agradecimento especial às professoras Mariana Cláudia Broens e principalmente Maria Eunice Quilici Gonzales, pelas diversas leituras e releituras no decorrer da preparação desta dissertação. Sou grato ainda a meus colegas de curso, pela recepção, convivência e diálogos enriquecedores;

à Lilian Guibbina, cujas conversas me ajudaram a terminar este trabalho.

Expresso, por fim, meu agradecimento especial àqueles que, há muito tempo, vem me ajudando ou me ajudaram a chegar até aqui:

ao meu professor e orientador de graduação, Abel Lassalle Casanave;

à minha família, principalmente aos meus pais, que fizeram o possível e o impossível para me auxiliar sempre que necessitasse;

à Marisa, meu suporte nos momentos mais difíceis desta etapa de minha vida. Agradeço-lhe por ter sabido ser amiga, irmã e mulher nas horas mais próprias e necessárias.

*"...Tinhã conversado sobre robôs. Jorunn considerava o
cêrebro humano um computador complicado. Sofia não
estava bem certa se concordava com isto. O ser humano não
seria algo mais do que uma máquina?"
(Jostein Gaarder)*

*"Meu objetivo neste trabalho é saber como sabemos..."
(Warren McCulloch)*

*"Qualquer livro intitulado Como a mente funciona deveria
começar com uma nota de humildade;...
não entendemos como a mente funciona..."
(Steven Pinker)*

*"...Não combatemos com vista a uma vitória definitiva para um
futuro indefinido. A maior vitória possível é continuar a ser e
ter sido... só podemos permanecer onde estamos, se
correremos o mais rápido possível... A tragédia está aqui, mas
também a glória."
(Norbert Wiener)*

Índice

<i>Resumo</i>	3
<i>Abstract</i>	4
<i>Introdução</i>	5

Parte I

<i>Cap. 1 - Teorias da mente</i>	10
1 Apresentação	10
2 Algumas teorias dualistas da mente	11
3 Algumas teorias monista-materialistas da mente	20
4 A mente na Ciência Cognitiva	40
<i>Cap. 2 - Modelos algorítmicos da mente</i>	49
1 Apresentação	49
2 A origem da Máquina de Turing	51
3 A definição da Máquina de Turing	59
3 Discussão sobre o conceito de Máquina	76
<i>Cap. 3 - Modelos conexionistas da mente</i>	88
1 Apresentação	88
2 A rede de McCulloch-Pitts	89
3 Caracterização das redes neurais artificiais modernas	103
4 O perceptron e a rede de kohonen	110

Parte II

Cap. 4 - <u>O conceito de Inteligência</u>	124
1 Apresentação	124
2 O conceito de inteligência em geral	125
3 O conceito de Inteligência na Psicologia Cognitiva	132
Cap. 5 - <u>A inteligência na Ciência Cognitiva</u>	136
1 Apresentação	136
2 O conceito de inteligência na Inteligência Artificial	137
3 O conceito de inteligência no Conexionismo	151
Cap. 6 - <u>Da possibilidade de uma Inteligência Artificial</u>	159
1 Apresentação	159
2 O Teste de Turing	160
3 O Argumento do Quarto Chinês	172
 <u>Conclusão</u>	 185
<u>Referências bibliográficas</u>	190



Resumo

Neste trabalho, realizamos um estudo sobre o conceito de inteligência nas duas principais vertentes da Ciência Cognitiva - Inteligência Artificial e Conexionismo. Visamos, com isto, extrair as características do comportamento inteligente em cada corrente. Uma das hipóteses fundamentais da Ciência Cognitiva é que, em princípio, a inteligência e o comportamento humano inteligente podem ser simulados e explicados por modelos computacionais da mente (do tipo redes neurais artificiais e máquinas de Turing). Para alguns adeptos desta área de pesquisa estas máquinas são inteligentes, pensam, possuem estados mentais. Por isso, após uma exposição de algumas teorias sobre a mente, apresentamos três destas máquinas. Comentamos seu modo de funcionamento, características, capacidades e limitações. Numa segunda parte, analisamos o conceito de inteligência num sentido geral e, em especial, do ponto de vista dos dois grupos de pesquisa da Ciência Cognitiva acima citadas. Por fim, expomos argumentos a favor e contra a possibilidade das máquinas serem consideradas inteligentes, indicando o alcance e os possíveis limites da visão mecanicista da mente.



Abstract

In this work, we accomplish a study about the concept of Artificial Intelligence considering the two main tendencies of cognitive science. We intend to extract the characteristics of intelligent behavior from each current. One of the fundamental hypotheses of cognitive science is, at first, that human intelligence and behavior can be simulated by computational models of mind (like Artificial Neural Network and Turing Machine). According to some researchers of this science, these are intelligent machines, they think and have mental states. So, after explanation of some theories of mind, we present three kinds of the mentioned machines. We explain their operation, their characteristics, their possibilities, and their limitations. In the second part, we analyze the concept of intelligence in a general sense and the point of view of two researcher's groups of cognitive science. Finally, we display on behalf of unfavourable the possibility of machines be considered intelligent, taking into consideration characterization of the concept of intelligence given by the arguments' authors.



Introdução

O que é a mente? Qual a natureza dos estados e faculdades mentais? Qual a natureza do pensamento, da consciência, da inteligência, do conhecimento, da crença? Como sabemos que outros seres humanos têm mente? Será que apenas os seres humanos a possuem? Qual a relação da mente com o corpo, em especial com o cérebro?

As questões acima são apenas algumas das indagações que afligem filósofos há muitos séculos. A tentativa de entender como funciona a mente humana e como ela é caracterizada é muito antiga. Porém, devido à própria natureza da investigação filosófica, o método adotado pelos filósofos no estudo da mente é na maioria das vezes discursivo. Neste século, surge uma série de novas áreas ou partes de áreas de pesquisa já existentes, com o objetivo de dar um outro enfoque a tal estudo. Os cientistas destas áreas, porém, não pretendiam adotar um método simplesmente discursivo. Visavam, de um modo ou de outro, poder observar e manipular estados e faculdades mentais. Muitos desejavam adotar um método empírico, algumas vezes semelhante ao da Física em suas pesquisas. Dentre as diversas áreas envolvidas neste trabalho podemos citar a Biologia, Psicologia, Física, Linguística, Neurociências, Filosofia, Cibernética e Computação. Tais ciências foram denominadas *Ciências Cognitivas*.

Há uma distinção entre o que denominamos *Ciências Cognitivas* e *Ciência Cognitiva*. As ciências pertencentes ao grupo das primeiras caracterizam-se pelo fato de pesquisar estados e faculdades mentais e, mais especificamente, os processos cognitivos humanos, cada uma a seu modo. Existe a preocupação, por parte de alguns pesquisadores em fazer com que a interação entre cientistas de áreas diferentes crie um campo de pesquisa unificado e um vocabulário comum para a explicação dos processos cognitivos humanos. Na visão idealizada destes pesquisadores, todos os cientistas das ciências cognitivas deveriam adequar-se a um mesmo método e vocabulário. Ainda que seja esse o objetivo de alguns pesquisadores, deve-se notar que nem por isso as ciências cognitivas perderiam sua identidade. Seus integrantes continuariam mantendo um diálogo com estudiosos de outras áreas sem deixar de ter seu modo particular de pesquisa.

Uma das Ciências Cognitivas é a *Ciência Cognitiva*. Sua característica principal é a ênfase dada à função exercida pelos computadores no estudo da mente, qual seja, o de fornecer estudos explicativos e preditivos de estados e faculdades mentais. Ainda que busque uma interdisciplinaridade, esta ciência possui, ou pretende possuir, método e vocabulário próprios no que tange ao estudo da mente. É sobre a Ciência Cognitiva que direcionamos a presente dissertação.

Neste trabalho, nos deteremos em apenas uma pequena parte da pesquisa sobre a mente. Nosso objeto de estudo é o conceito de inteligência adotado pelos pesquisadores da Ciência Cognitiva, mais especificamente em duas de suas principais correntes: a *Inteligência Artificial* (IA) e *Conexionismo* ou *Redes neurais Artificiais* (RNA). Dividimos a dissertação em duas partes. A primeira compreende os três primeiros capítulos, nos quais são fornecidos uma noção geral de diversas teorias sobre a mente e os sistemas construídos na Ciência

Cognitiva para simular aspectos cognitivos humanos. Na segunda parte, que engloba os três últimos capítulos, apresentamos o estudo sobre o conceito de inteligência na Ciência Cognitiva.

No primeiro capítulo apresentamos algumas das principais teorias da mente. Dentre elas, expomos a teoria segundo a qual os cientistas da IA e RNA apoiam-se para a construção de seus modelos computacionais. Tal teoria é denominada *funcionalismo*. Acreditamos que uma apresentação prévia desta teoria nos proporciona um entendimento mais claro do que as correntes aqui apresentadas pretendem dizer a respeito da mente. Apresentamos juntamente outras teorias da mente e as dividimos em monistas e dualistas. Expomos, por fim, uma teoria cujo objetivo não é nem materializar a mente nem torná-la algo imaterial, inacessível ao domínio científico, denominada Naturalismo Biológico. Terminamos o capítulo apresentando uma caracterização da Ciência Cognitiva e de suas duas correntes analisadas nesta dissertação.

Uma das hipóteses fundamentais de muitos pesquisadores da IA como Newell, Shaw & Simon (1958) e Minsky (1967) é que estados e faculdades mentais podem ser simulados e explicados através de modelos computacionais. Para pesquisadores como Putnam (1967), por exemplo, tais modelos exibem estados e faculdades mentais. Desse modo, tais características mentais poderiam ser simuladas e explicadas por certas máquinas. Por outro lado, integrantes da IA mais fraca e conexionistas como Kohonem (1987), McClelland & Rumelhart (1988) acreditam que seus modelos apresentam apenas algumas características semelhantes ao funcionamento da mente. Porém, não chegam a afirmar que tais modelos realmente possuem estados e faculdades mentais. Os modelos construídos pelos cientistas cognitivos constituem um tipo de *Máquina de Turing* ou de *Rede Neural Artificial*. O



primeiro grupo de máquinas, de grande aplicação na Inteligência Artificial, é apresentado no segundo capítulo deste trabalho. O segundo grupo, de grande emprego no Conexionismo, é exposto no terceiro capítulo. Procuramos apresentar sua definição, modo de funcionamento, características capacidades e limitações.

A idéia fundamental dos pesquisadores da IA é que a mente é um sistema de processamento de informações. Ela segue regras e manipula símbolos como uma máquina de Turing. Do mesmo modo que na IA, a mente é também entendida como um sistema processador de informações no Conexionismo. Porém, ao estudá-la, existem algumas características físicas e biológicas fundamentais que devem ser levadas em consideração. Os pesquisadores de ambas as correntes entendem que seus modelos mecanicistas da mente podem auxiliar no estudo sobre a natureza de estados e faculdades mentais.

Uma vez conhecidas as teorias da mente e as máquinas usadas para simular estados, processos e faculdades mentais, podemos examinar o conceito que nos interessa nesta dissertação, a saber, o de inteligência. A análise do conceito de inteligência é feita na parte dois do trabalho, que compreende os capítulos quatro a seis.

No quarto capítulo, investigamos o conceito de inteligência a partir de uma perspectiva do senso comum. Apresentamos algumas contradições e contrariedades encontradas entre as diversas culturas que procuram caracterizá-lo. Em seguida, passamos à uma análise científica desse conceito. Mostramos a existência de discordâncias e desentendimentos sobre a caracterização desse conceito no âmbito científico. Terminamos o capítulo expondo uma análise feita pela Psicologia cognitiva sobre a inteligência.

No quinto capítulo, analisamos como as duas vertentes da Ciência Cognitiva caracterizam a inteligência. Para a IA, a inteligência está fundamentalmente ligada à



resolução de problemas, em especial àqueles do tipo lógico-matemático. Pesquisadores dessa corrente como Newell & Simon (1972), acreditam que os processos pelos quais o ser humano passa ao resolver problemas podem ser descritos e simulados através de programas computacionais. O comportamento inteligente, deste modo, caracteriza-se pelo agrupamento correto de regras lógicas. Por outro lado, conexionistas como Hopfield (1982), Kohonen (1987), McClelland e Rumelhart (1988) relacionam a inteligência com a capacidade de reconhecimento de padrões. A rede aprende a reconhecer padrões relacionando suas partes básicas sem a necessidade de regras explícitas.

No capítulo 6, apresentamos alguns argumentos a favor e contra a possibilidade das máquinas serem consideradas inteligentes. Dentre os argumentos a favor, o principal é o argumento proposto por Turing (1950), que envolve o famoso teste de Turing. Dentre os argumentos contra, destacamos o argumento do quarto chinês. Neste argumento, Searle (1984) afirma que o computador digital jamais poderá ser inteligente. Isto porque funciona apenas sintaticamente. Além disto, faltam características básicas a esta máquina para ser inteligente. Dentre estas encontramos a consciência, por exemplo.

Concluimos esta dissertação esboçando alguns caminhos segundo os quais consideramos que o estudo sobre o conceito de inteligência deva tomar. Dentre eles estão a importância de alguns aspectos sócio-culturais, que devem ser levados em consideração ao analisá-lo. Por outro lado, ressaltamos a necessidade de um esforço interdisciplinar para reunir as diversas caracterizações desse conceito.



Parte I

Cap. 1

Teorias da mente

1 Apresentação

Nosso objetivo central nessa pesquisa é analisar o conceito de inteligência desde um ponto de vista do mecanicismo. Para tanto, pensamos ser fundamental a compreensão da teoria funcionalista da mente, uma vez que, como veremos, o projeto mecanicista é baseado no funcionalismo. Porém, acreditamos que a compreensão do funcionalismo fica mais clara se entendermos algumas outras das principais teorias da mente. Por isso, este capítulo é dedicado a elas e termina com uma análise da mente segundo a ciência cognitiva.

As teorias explicativas da mente são geralmente classificadas em dois grandes grupos: *dualistas* e *monistas*. As primeiras possuem igualmente uma divisão: são de *substância* ou de *propriedades*. Para os dualistas dos dois tipos, existe algo além da pura matéria ao retratarmos-nos à mente. Por outro lado, os monistas (aqui entendidos como materialistas) afirmam nada existir além de sistemas físicos empiricamente observáveis.



Analisaremos, neste capítulo, algumas teorias dualistas e materialistas, procurando apresentar como a mente é caracterizada em cada uma delas. Ao fazer esta exposição, acreditamos poder entender melhor a postura adotada pela Ciência Cognitiva no tocante ao estudo da mente. Em outras palavras, teremos mais indícios para compreender e distinguir a posição defendida pelos pesquisadores desta “ciência” a respeito da mente em contraposição às posições adotadas por outras áreas de pesquisa no tocante ao mesmo tema. Apresentaremos, por fim, uma teoria que, segundo Searle, seu criador, não é nem monista materialista nem dualista. Com isso, Searle pretende eliminar a idéia de que somos necessariamente ou dualistas ou materialistas. Na última seção deste capítulo analisamos a caracterização da mente feita na Ciência Cognitiva, especialmente na IA e RNA.

2 Algumas teorias dualistas da mente

2.1 Dualismo de Substâncias.

Ao falar em dualismo de substâncias, o nome diretamente associado a ele é o de Descartes (1996a;1996b). O dualismo cartesiano, tal como apresentado principalmente nas *Meditações*, caracteriza-se pelo fato de pressupor a existência de duas substâncias distintas quando falamos de mente e corpo. Segundo Descartes, embora nossa alma, esteja ligada a um corpo físico, ela é substancialmente distinta dele.

De acordo com Descartes, o corpo não pode existir sem a alma, embora esta possa existir sem aquele. Para este filósofo, a propriedade intrínseca da alma é a de ser pensante. Por outro lado, a propriedade intrínseca do corpo é a de ser extenso. À alma pertencem



faculdades como a de duvidar, negar, conceber, afirmar, imaginar e sentir. Tais faculdades encerram alguma espécie de inteligência e como elas pertencem à alma (substância imaterial), Descartes (1996b, p. 326) a denomina substância inteligente. Ao corpo pertencem certas faculdades como a de locomover-se e características biológicas como a digestão, por exemplo. De acordo com Descartes (1996b, p. 250-1), podemos notar a diferença entre o corpo e a alma

... pelo fato de não concebermos qualquer corpo senão como divisível, ao passo que o espírito ou a alma do homem não se pode conceber senão como indivisível: pois, com efeito, não podemos conceber a metade de alma alguma, como podemos fazer com o menor de todos os corpos; de sorte que suas naturezas não são somente reconhecidas com diversas, porém mesmo, de alguma maneira, como contrárias... o corpo humano... não é formado e composto senão de certa configuração e membros e outros acidentes semelhantes; e a alma humana, ao contrário, não é assim composta de quaisquer acidentes, mas é uma pura substância. Pois, ainda que todos os seus acidentes se modifiquem, por exemplo, que ela conceba certas coisas, que ela queira outras, que ela sinta outras etc., é, no entanto, sempre a mesma alma; ao passo que o corpo humano não mais é o mesmo pelo simples fato de se encontrar mudada a figura de alguma de suas partes. Donde se segue que o corpo humano pode facilmente perecer, mas que o espírito ou a alma do homem (o que eu absolutamente não distingo) é imortal por sua natureza.

É difícil encontrar atualmente defensores do dualismo. Seus adeptos estão hoje em dia resumidos em sua grande maioria entre teólogos e religiosos. Na comunidade científica, os que são dualistas geralmente não o confessam explicitamente. Uma das exceções é Penrose (1993), sobre o qual falaremos mais no último capítulo.

Ryle (1949) tentou descrever o dualismo cartesiano, definindo a alma como a negação do corpo. Desse ponto de vista, a mente é não-extensa, não-física, não-regida por leis físicas e assim por diante. Porém, de acordo com o que dissemos acima, não existe uma negação entre corpo e alma. Ao invés de contraditórias, elas são contrárias.

De acordo com Ryle (1949, p.13), o dualismo cartesiano pode ser descrito do seguinte modo:

Com a duvidosa exceção dos idiotas e bebês, todo ser humano tem um corpo e uma mente... Este corpo e esta mente estão ordinariamente arnesados juntos, mas depois da morte do corpo sua mente pode continuar a existir e funcionar.



Os corpos humanos estão no espaço e estão sujeitos às leis mecânicas que governam todos os outros corpos no espaço. Processos e estados corporais podem ser inspecionados por observadores externos...Mas as mentes não estão no espaço nem estão suas operações sujeitas a leis mecânicas. O funcionamento de uma mente não é testemunhável por outros observadores; seu ofício é privado. Só eu posso ter conhecimento direto dos estados e processos de minha própria mente.

Embora esta apresentação do dualismo de substâncias seja um tanto caricatural, como admite o próprio Ryle, ele acredita que é um bom esquema para a compreensão desta teoria.

Como dissemos, Descartes acredita que estamos ligados a um corpo físico. Porém, de que modo acontece tal ligação? Segundo Descartes (1996b, p. 329), há uma união de fato entre a alma e o corpo. Embora sejam essencialmente distintos, corpo e alma estão tão unidos de tal forma que quase formam um todo único.

Para Descartes (1996b), corpo e alma não estão ligados assim como um piloto está ligado a um navio. Embora esses estejam ligados, são duas coisas distintas. O piloto reconhece problemas no navio através de sua percepção, por exemplo. Porém, um problema no navio não implica diretamente um problema no piloto e vice-versa. Desse modo, se quebrar uma tábua do navio, por exemplo, isso pode não trazer nenhuma consequência direta ao seu piloto. Este, por outro lado, nota o estrago no navio quando consegue observá-lo ou notá-lo por meio de alguma consequência provada por aquele estrago.

Se a alma e o corpo estivessem ligados assim como o piloto e o navio, ela sentiria uma dor provocada, por exemplo, por um ferimento no corpo, apenas quando seu entendimento fosse capaz de perceber tal ferimento. Porém, não é assim que acontece. Mesmo sem pensar no ferimento, a alma sente a dor provocada por ele. Segundo Descartes (1996b, p. 328-9), isso mostra que

... não somente estou alojado em meu corpo, como um piloto em seu navio, mas que, além disso, lhe estou conjugado muito estreitamente e de tal modo confundido e misturado, que componho com ele um único todo. Pois, se assim não fosse, quando meu corpo é ferido não sentiria por isso dor alguma, que eu não sou senão uma coisa pensante, e apenas perceberia esse ferimento pelo

entendimento, como o piloto percebe pela vista se algo se rompe em seu navio; e quando meu corpo tem necessidade de beber ou de comer, simplesmente perceberia isto mesmo, sem disso ser advertido por sentimentos confusos de fome de sede. Pois, com efeito, todos esses sentimentos de fome, de sede, de dor etc., nada são exceto maneiras confusas de pensar que provêm e dependem da união e como que dá a mistura entre o espírito e o corpo.

Embora a alma componha com o corpo um todo único, ambos possuem uma distinção real, ainda segundo Descartes (1996b, p. 326). Cada um possui suas características e funções específicas. Assim, habilidades como conceber, negar, duvidar, bem como outros sentimentos e emoções são atribuídas à alma e não ao corpo.

Ainda que determinados estados e faculdades pertençam à alma e não ao corpo, existem alguns deles que são causados, quer direta ou indiretamente pelo aparato físico. De fato, corpo e alma estão em constante relação. Um exemplo em que a notamos é o de que, quando batemos o dedo de tal forma que o machucamos, sentimos dor. Bater o dedo é uma ação corporal e sentir dor é um estado mental. Desse modo, o corpo, de certa forma, afeta a alma. Além disso, é comum tomarmos como causa de nossos comportamentos alguns desejos, emoções, sentimentos. Se assim é, como podemos explicar esta relação entre duas substâncias contrárias? Qual é o ponto onde o corpo afeta a alma ou vice-versa?

Descartes tentou resolver o problema acima, denominado *problema da relação mente-corpo*, afirmando que há um lugar em especial onde a mente e o corpo mantém sua comunicação. Segundo Descartes (1996a, p. 150), "... a glândula pineal é onde a alma exerce suas funções mais particularmente do que nas outras partes. Esta glândula é a principal sede da alma." As percepções, imaginações e funções corporais chegam à alma através da glândula pineal. Porém, com isso, Descartes não resolveu o problema, uma vez que a glândula pineal faz parte do corpo. A explicação da ligação entre esse objeto físico



com a alma não foi não determinada. Descartes realmente apenas modificou o problema, que continua em aberto até hoje (do ponto de vista dualista), sem resolvê-lo.

O fato de não resolver ou não explicar a relação entre a mente e o corpo é apenas uma das muitas críticas direcionadas ao dualismo de substâncias. Outra delas refere-se ao fato das hipóteses de que a mente é imaterial, imortal, independente de leis físicas, não passarem de meras conjecturas. Até hoje não se conseguiu mostrar, cientificamente, nada sobre tais características da mente. Segundo Churchland (1984, p. 17), "... o dualismo de substâncias envolve, ao contrário, muita crença, porém, pouco de um estudo realmente científico a respeito da mente." Além disso, continua tal pesquisador, o dualismo não é uma boa teoria da mente por causa de sua pouca força explicativa sobre as propriedades mentais. Outras áreas de pesquisa sobre a mente, como as neurociências, por exemplo, já conseguiram descobrir inúmeros problemas cerebrais diretamente relacionados com alguns estados mentais ou problemas como inabilidade de falar, ler, reconhecer objetos. Churchland (1984, p. 19) propõe que comparemos

... o que o neurocientista pode nos dizer sobre o cérebro, e o que ele pode fazer com este conhecimento, com o que o dualista pode nos dizer sobre a substância espiritual, e o que ele pode fazer com estas suposições. Pode o dualista nos dizer qualquer coisa sobre a constituição interna da 'matéria-prima' da mente? Dos elementos que a constituem? Sobre leis que governam seu comportamento? Sobre as conexões estruturais da mente com o corpo? Sobre o modo de suas operações? Pode ele explicar as capacidades e patologias humanas em termos de suas estruturas e seus defeitos? O fato é que o dualista não pode fazer nada disto, porque nenhuma teoria detalhada da mente foi formulada [por ele].

Pesquisadores como Rorty (1965) e o próprio Churchland (1984) afirmam que algumas teorias materialistas, como as que tentam identificar a mente com o cérebro, estão muito mais avançadas desde um ponto de vistas explicativo da mente do que o dualismo. Tomam como exemplo casos em que uma lesão cerebral provocou fortes influências sobre estados e faculdades mentais.

Os dualistas, de certa forma, concordam que alguns estados e faculdades mentais devem ter alguma dependência com o corpo e, mais especificamente, com o cérebro. Porém, ao contrário do que afirmam seus críticos, acreditam que o aparato físico não tem controle sobre aqueles estados e faculdades genuinamente mentais. Dentre estes, estão as capacidades centrais da mente, como o pensamento, raciocínio, algumas emoções e a consciência. Para rebater tal afirmação, materialistas como Rorty (1965) e Stich (1983) novamente apelam às experiências nas quais nota-se que distúrbios ou alterações cerebrais modificam a capacidade de raciocínio e pensamento do indivíduo. Algumas vezes, afirmam eles, podem torná-lo consciente ou inconsciente de alguns fatos. Para eles, isso mostra que tais estados mentais possuem uma dependência física.

Um dos exemplos mais citados para mostrar que estados e faculdades mentais não dependem simplesmente de uma mente imaterial, mas que estão ligados à matéria, é o de que já existem máquinas que pensam, raciocinam, e podem até ter emoções. É claro que os dualistas acreditam que as máquinas, de fato não possuem nada disso. Isso porque não possuem uma alma como a humana.

Diante de todos os argumentos contrários ao dualismo de substâncias - como o de não resolver o problema da relação mente-corpo, pressupor a existência de algo imaterial além do domínio da ciência empírica e exigir uma porcentagem de crença nas afirmações acerca da mente -, um grupo de pesquisadores cria um dualismo menos forte do que o de substâncias, denominado *dualismo de propriedades*.



2.2 Dualismo de propriedades.

Nesta forma de dualismo, a mente e o corpo não são duas substâncias distintas, como na forma anterior. Ao contrário, a única substância existente é a do cérebro, que possui certas propriedades físicas (materiais) e outras não-físicas. Estas últimas fogem ao domínio da observação e manipulação empírica.

O dualismo de propriedades é conhecido também por *monismo anômalo*. É monista porque pressupõe uma só substância quando falamos de mente-cérebro. É anômalo por admitir que esta mesma substância possui dois tipos de propriedades de natureza distintas. Suas propriedades não-materiais são exatamente os estados mentais. Eles não podem ser reduzidos nem explicados em termos puramente físicos. Para explicá-los, é preciso uma “ciência totalmente nova e autônoma”, a ciência do fenômeno mental (Churchland, 1984, p. 10). O dualismo de propriedades possui diversas variações, das quais citaremos apenas duas. A primeira é conhecida como *epifenomenalismo* e a segunda como *interacionismo*.

A idéia básica do epifenomenalismo é que os estados mentais emergem (aparecem) do cérebro, quando este alcança um certo grau de complexidade. Porém, uma vez emergidos, os estados mentais não podem ser explicados puramente apenas observando o próprio cérebro. Eles não são uma parte do fenômeno físico no cérebro (que determina nossas ações e comportamentos em geral), mas estão “acima” dele.

Esta teoria apresenta uma característica no mínimo estranha: estados físicos causam estados mentais, porém, o inverso não acontece.¹ Os estados mentais são impotentes frente ao comportamento humano, explicado em termos puramente físicos. Desse modo, nossas

ações, comportamentos não são causados por desejos, emoções, crenças. Eles são, de fato, causados pelo cérebro. Sendo assim, parece que os estados mentais não têm nenhuma função no sistema. Para que, então, postular sua existência?

A segunda versão do dualismo de propriedades a que nos referimos, muito semelhante à primeira, é a interacionista. Assim como na epifenomenalista, nesta teoria os estados mentais emergem do cérebro quando este alcança um certo nível de complexidade. Porém, uma vez originados, são irreduzíveis aos mesmos fenômenos físicos dos quais surgiram.

O interacionismo, porém, tem uma diferença fundamental em relação ao epifenomenalismo: estados mentais têm efeitos causais sobre o cérebro e, portanto, sobre o comportamento (uma vez que o cérebro é o causador direto do comportamento). Nossos desejos, crenças, sentimentos têm alguma influência causal sobre nosso modo de agir.

Existem algumas falhas neste tipo de dualismo. Por exemplo, é aparentemente muito contra-intuitivo afirmar que estados mentais emergem do cérebro e, uma vez originados, não podem ser mais explicados nem reduzidos a ele. Como explicar esta irreducibilidade? Se os estados mentais são acrescidos de algumas propriedades especiais quando emergidos, quais são, como são e de que modo são acrescidas? Tais explicações geralmente não são proporcionadas nesta forma de dualismo. Além disso, não é bem explicado qual o grau de complexidade que o cérebro deve atingir para que se forme um estado mental e que tipo de complexidade é esta. Sabe-se que tal complexidade origina-se através da relação entre neurônios. Porém, como deve ser tal relação e quais neurônios estão envolvidos? Essas e outras explicações necessárias para o esclarecimento da emergência dos estados mentais de estados físicos não são fornecidas.

¹ Esta característica só é válida para a noção de causalidade, segundo a qual a causa de algo físico deve ser

Uma vez apresentados os dois modos de dualismo, podemos nos perguntar por que surgiram ou o que leva os cientistas a adotar tais teorias. Citamos aqui apenas três motivos. Em primeiro lugar, principalmente no que diz respeito ao dualismo de substâncias, a adoção do dualismo tem uma origem religiosa. A crença na existência de uma alma imortal e independente do corpo é um dos motivos fundamentais para a criação de uma teoria dualista. Porém, perguntam-se muitos pesquisadores como Churchland (1984), que razões teríamos para aceitar crenças religiosas? A história já não nos mostrou variados casos em que crenças religiosas impediram o progresso científico, inclusive com perseguições e torturas aos cientistas?

Um segundo motivo para a adoção de um dualismo é a introspecção. Ao que tudo indica, há uma subjetividade no que diz respeito aos estados e faculdades mentais. Algo que é interno e que só o próprio indivíduo consegue notar. Não pode ser objetivado nem verificado empiricamente. Como no caso da dor, por exemplo, conseguimos senti-la e sabemos quando e de que modo a temos internamente, sem precisar de observações externas ou físicas e nem tornar este sentimento objetivo. É um sentimento meu e apenas meu. Como isso é subjetivo e não pode ser objetivado, é preciso manter os estados mentais em um lugar de realce, não sendo possível seu entendimento em terceira pessoa.

Em terceiro lugar, está a idéia de destaque do ser humano em relação ao resto das criaturas do universo. A maioria dos dualistas acredita que apenas algumas espécies animais podem ter estados mentais (porque possuem uma alma perceptiva, como afirmava Santo Tomás de Aquino). Além disso, julgam que apenas o ser humano pode pensar, raciocinar, e possuir outras faculdades pertencentes à razão (porque apenas ele possui uma alma

também algo físico.

racional). Nesse sentido, se estados mentais pudessem ser reduzidos a estados físicos, outras entidades poderiam tê-los, privando o homem deste lugar especial. Porém, afirmam materialistas como Minsky (1964) e Lewis (1972), entidades puramente físicas e “sem alma”, como os computadores, podem realizar variadas atividades inteligentes que eram em outros tempos de exclusividade humana. É uma questão de tempo até que eles possam ter também estados mentais.

Como dissemos, o dualismo tem muitas críticas e críticos. As hipóteses desta teoria são pouco aceitáveis pela comunidade científica atual. Além disso, revela uma incapacidade de resolver os problemas cuja solução se propõe a alcançar. Por esses e outros motivos, tal teoria não tem muitos adeptos, pelo menos entre os cientistas. De fato, acredita-se que o dualismo não tem lugar na visão científica contemporânea. Enquanto a teoria da evolução e a teoria molecular (atômica) predominarem, o dualismo terá um lugar marginal na ciência. As teorias da mente que estão mais de acordo com tal visão são as monistas materialistas. Estas serão apresentadas na próxima seção.

3 Algumas teorias monista-materialistas da mente

A teoria contrária ao dualismo é a *teoria monista da mente*. Para os monistas, mente e corpo fundem-se num elemento apenas e, se ele é material, temos o que chamamos *monismo materialista*. Ao fundir mente e corpo, o problema da relação entre eles não é solucionado, mas sim, dissolvido. Como não existem dois objetos de substâncias distintas, mas apenas um elemento, não há relação a ser explicada entre objetos.



Muitas são as teorias materialistas da mente. Elas começaram a surgir com grande expressão no início do século XX, embaladas principalmente pela Psicologia cognitiva. Um dos objetivos fundamentais desta “ciência” era poder entender os processos pelos quais o homem passa quando pensa, raciocina, aprende. A caracterização dos estados e faculdades mentais deveria ser feita de uma maneira empírica, observável, positiva. A teoria de maior expressão nesse cenário foi a *behaviorista*.

3.1 O Behaviorismo

Um dos objetivos principais do behaviorismo é o de tornar o método do estudo da mente empírico e descrever suas características de modo preciso. Seus adeptos pretendiam, assim, eliminar a obscuridade existente no dualismo. O behaviorismo teve seu auge logo após o término da Segunda Guerra Mundial. Em linhas gerais, podemos dizer que tem como objetivo tratar estados mentais em termos comportamentais. Para saber se um indivíduo tem um determinado estado mental, analisamos seu comportamento. Behavioristas como Watson (1913) pretendiam analisar estados e faculdades mentais de modo objetivo. Isso seria feito através da análise comportamental do indivíduo.

Os behavioristas não buscam entender as constituições internas da mente. Esta era para eles algo obscuro e desconhecido. Como afirma Gardner (1998, p. 174)

... a Psicologia comportamentalista investigava os comportamentos observáveis: estímulos e respostas ('E-R'). Ela não tentava compreender as engrenagens internas da mente, que os comportamentalistas tendiam a ver como 'uma caixa preta'.

O behaviorismo é dividido em dois ramos principais: o *metodológico* e o *lógico*. O behaviorismo metodológico consiste em encontrar as relações existentes entre dados de



estímulos e respostas comportamentais e caracteriza-se como uma ciência empírica. Os estados mentais são analisados pura e simplesmente por meio de estímulo-resposta. Seus adeptos, como Skinner, por exemplo, não levam em consideração, nesta análise, qualquer característica subjetiva ou introspectiva. A mente é analisada via comportamento, relacionado com os *inputs* que o ocasionaram. Segundo Watson (1925, p. 78),

o behaviorismo é uma estratégia de pesquisa em Psicologia, no sentido de que uma ciência da Psicologia deva consistir em descobrir as correlações entre *input* de estímulos e *outputs* comportamentais.

Dois exemplos de pesquisa da análise de estados mentais através de comportamento são os experimentos realizados por Skinner e Pavlov. As experiências desses cientistas eram realizadas com animais como ratos e cães. Eles tentaram treinar organismos para dar uma resposta específica (R), mensurável a um estímulo específico (E) numa dada situação. Segundo Gardner (1998, pg. 66), "... na teoria 'E-R', os comportamentos não eram o resultado de pensamentos, planos, vontades ou algo parecido; eram reações a eventos do meio ambiente."

Outro cientista a trabalhar nesta área foi Thorndike (1913). Seu trabalho consistia em fazer experiências com animais sobre a aprendizagem. Descobriu que quanto mais os animais treinavam determinados experimentos, mais rapidamente resolviam certos problemas. Thorndike acreditava que suas pesquisas com esses animais poderiam revelar algo sobre os estados mentais humanos.

O behaviorismo lógico procura definir estados mentais em termos de afirmações, geralmente hipotéticas, sobre o comportamento. A diferença entre esse tipo de behaviorismo e o anterior é que, enquanto o primeiro procura caracterizar estados mentais empiricamente



via relação estímulo-resposta, o último busca tal caracterização via disposições de comportamento ditadas por sentenças lógicas.

Segundo Fodor (1968, p. 52),

Para ser qualificado como behaviorista,... alguém precisa simplesmente acreditar que a seguinte proposição expressa uma verdade necessária: para cada predicado mental que possa ser empregado numa explicação psicológica deve haver pelo menos uma descrição do comportamento com a qual ele mantém uma conexão lógica.

A sentença será algo do tipo: x é o estado mental y se e somente se tem z comportamentos (z é determinado por cada estado mental).

Como afirma a sentença anterior, um estado mental pode ser relacionado com muitos comportamentos. Isso não significa porém que, para possuir um estado mental, um indivíduo precisa apresentar todos os comportamentos associados ao estado em questão. De fato, um estado mental não implica necessariamente um comportamento, mas uma disposição para ele. Se o sujeito tem um determinado estado mental, tem disposição para agir de certos modos específicos. Assim, se tiver dor, por exemplo, tende a gritar, a chorar, etc, sem necessariamente ter de apresentá-los todos cada vez que tiver dor.

O conceito de disposição representa um dos grandes pontos de debate no behaviorismo lógico. Esse termo é apresentado por Ryle (1949) que, por isso, foi erroneamente caracterizado behaviorista. Uma diferença fundamental entre este pesquisador e os behavioristas é que, enquanto estes não se preocupam em entender as constituições internas da mente, Ryle tinha tal preocupação. Para ele, ao querer saber sobre o estado mental de uma pessoa, não é suficiente analisar apenas seu comportamento. Deve-se, além disso, levar em consideração, dentre outros aspectos, a história pessoal do indivíduo e sua relação com o meio.



Para explicar o conceito de disposição, recorreremos ao exemplo da solubilidade do açúcar. Ao dizer que o açúcar é solúvel, pretendemos afirmar que, ao colocá-lo na água, se dissolverá. Cada vez que o expusermos a certa situação poderá sofrer uma certa modificação. Nesse sentido é que podemos extrair uma semelhança entre estados disposicionais naturais e a disposição comportamental de estados mentais. Toda vez que um sistema estiver num certo estado mental poderá apresentar um certo comportamento.

Ainda que o termo disposição tenha uma certa caracterização, falta-lhe uma definição mais rigorosa, afirma Hampshire (1950). Além disso, não existe uma relação precisa dos comportamentos disposicionais de um estado mental. Quantos destes comportamentos um sistema precisa apresentar para ter ou estar num estado mental? Como ordenar tais comportamentos, a fim de poder realmente, através deles, identificar um estado mental? Se o sujeito não tivesse algum destes comportamentos para os quais tem disposição, poderia ter ou estar em certo estado?

Como dissemos, os estados mentais são definidos por meio de sentenças, em sua maioria hipotéticas. Assim sendo, a determinação do comportamento do indivíduo depende de certas condições. Por exemplo, supomos que Paulo acredita que vai chover. De acordo com o behaviorismo, podemos saber de sua crença analisando seu comportamento. Assim, diremos que, se as janelas de seu quarto estiverem abertas, Paulo as fechará; se for sair de casa, levará guarda-chuva e assim por diante. Porém, Paulo poderia não fechar as janelas e mesmo assim ainda ter a crença de que vai chover. Portanto, este condicional seria insuficiente para determinar se este indivíduo tem ou não tal crença. Desse modo, colocamos em dúvida o caráter de determinação do condicional. Podemos saber quando ele é realmente suficiente? Haveria sempre a possibilidade de acontecer o condicional e não



acontecer o comportamento ditado pelo conseqüente e mesmo assim o indivíduo ter o estado mental referido? Ou é possível encontrar um condicional preciso? Se for, como obtê-lo? As respostas a tais questões não ficam claras no behaviorismo.

Além das objeções acima, citamos outras três que diminuem a confiança no behaviorismo em suas duas vertentes. A primeira delas é o argumento do superator/espartano proposto por Putnam (1963). O caso do superator é o do indivíduo que finge estar em um determinado estado mental. Este indivíduo age de acordo com os comportamentos associados a esse estado sem de fato tê-lo. Trata-se de uma simples representação. No caso do super-espartano, lembramos os guerreiros de Esparta, treinados a suportar a dor sem se exaltar. Tal soldado poderia não exibir os comportamentos associados à dor e mesmo assim estar passando por ela. Sendo assim, não é possível determinar o estado mental do sujeito simplesmente analisando seu comportamento.

O argumento acima traz à tona a segunda crítica ao behaviorismo: ele deixa de lado aspectos internos, subjetivos, introspectivos do indivíduo, eliminando dele qualquer característica mental (Searle, 1997). De acordo com os behavioristas, todo estado mental pode ser objetivamente definido e identificado através da análise do comportamento do sistema em questão.

Ao identificar estado mental com comportamento ou com disposição comportamental, o behaviorismo deixa de lado as relações causais entre estados mentais e comportamento. Esta é a terceira crítica a esta teoria da mente. Para tal teoria, desejos, sentimentos, emoções não causam comportamentos. Eles são os próprios comportamentos.

A identificação de estados mentais com comportamento acabou por ser um pouco desacreditada pelos pesquisadores das diversas Ciências Cognitivas. Muitas teorias tentaram

encontrar maneiras distintas de poder caracterizar e identificar estados e faculdades mentais. Porém, todas elas têm como fim caracterizá-los e identificá-los através de algo empiricamente observável e manipulável.

Um dos modos de tornar a pesquisa dos estados e faculdades mentais empírica foi identificá-los com estados cerebrais. Desde o início deste século, muitos estudos sobre o cérebro foram realizados. Descobriu-se nestas pesquisas que danificações neuronais provocam modificações em estados mentais e comportamentos dos indivíduos pesquisados. Dentre as várias teorias criadas sobre o cérebro humano, duas em especial podem ser destacadas: *holista* e *localizacionista*. Localizacionistas como Geschwind (1974) afirmam que cada região ou neurônio do cérebro tem uma função específica e única. Por outro lado, holistas como Lashley (1950) e Lennenberg (1967) dizem que o cérebro é uma unidade e não é o lugar do cérebro danificado que produz seqüelas, mas sim a quantidade, o tamanho da lesão. Alguns cientistas como Hebb (1949) tentaram unir as duas teorias. Para estes, cada parte do cérebro tem uma certa função. Na falta ou danificação desta parte, outras podem substituí-la com talvez alguma perda de qualidade na função da qual ela é responsável.

Podemos criar uma analogia para falar das teorias acima citadas: para os holistas é como se o cérebro fosse uma massa de modelar num formato plano: se tirarmos uma parte dela, pode-se espichar a massa e fechar o buraco. Para os localizacionistas, o cérebro seria como um pedaço de madeira. Uma vez quebrada uma parte da madeira, não tem como recuperá-la. Mesmo repondo o pedaço, notaremos a deformação apresentada. A teoria conciliadora afirma que é possível fecharmos o buraco da massa, porém sua espessura fica mais fina, menos resistente.



Seguindo tais idéias sobre o cérebro e o objetivo de tornar o estudo da mente empírico, alguns cientistas não apenas relacionaram estados mentais com estados cerebrais como os identificaram. Criaram, assim, a teoria da *identidade mente-cérebro*, que pode ser dividida em duas vertentes: a de *tipos* e a de *ocorrência*.

3.2 Teoria da identidade de tipos

A teoria de identidade de tipos tem como característica fundamental identificar tipos de estado mental com tipos de estado cerebral (neurofisiológico). Sendo assim, a dor, por exemplo, de fato nada mais é do que uma relação neuronal. Ao tomar um estado cerebral, é possível saber qual estado mental o sujeito tem. Por outro lado, ao identificar o estado mental de um indivíduo, é possível conhecer seu estado cerebral correspondente. Dentre seus defensores encontramos pesquisadores como Place (1956) e Smart (1959).

Os teóricos da identidade de tipos são acusados de “chauvinistas”. Isso porque, para eles, apenas entidades com um cérebro igual ao humano podem ter estados mentais (Block, 1978). A teoria de identidade dos tipos sofre ainda de algumas críticas já atribuídas ao behaviorismo, como por exemplo o fato de deixar de lado aspectos subjetivos do indivíduo ao tratar-se de estados mentais. Além disso, como no behaviorismo, onde é possível encontrar comportamentos associados à caracterização de um determinado estado mental sem necessariamente tê-lo, na teoria de tipos é possível encontrar estados cerebrais associados à caracterização de um determinado estado mental, sem necessariamente tê-lo. É perfeitamente possível que o indivíduo tenha o mesmo estado cerebral em circunstâncias distintas associado a estados mentais distintos. Isso faria com que a relação um-a-um se



desintegrasse. A teoria da identidade porém, não sofre a crítica de falta de relações causais entre estado mental e comportamento. O estado mental é identificado a estados cerebrais. Porém, o cérebro é quem causa o comportamento.

A objeção mais citada contra a teoria da identidade apoia-se na lei de Leibniz. Segundo esta lei, dois elementos são idênticos se e somente se possuem exatamente as mesmas propriedades. Ou ainda: se dois termos se referem ao mesmo objeto, então, para qualquer propriedade, ela é atribuída ao objeto referido pelo primeiro se e somente se for atribuída a ele também pelo segundo termo. Desse modo, para mostrar que estados mentais não são estados cerebrais, bastaria encontrar uma propriedade que um deles tivesse e outro não. Duas dessas propriedades poderiam ser as seguintes: os estados neuronais ocupam algum lugar no cérebro. Porém, estados mentais não parecem ter tal propriedade. Não seria muito correto afirmar que uma dor ou crença ocupa um lugar no cérebro. Por outro lado, dizemos de nossas crenças que elas são verdadeiras ou falsas ou que nossas dores são mais fortes ou fracas, o que não afirmamos dos estados neurofisiológicos. Os teóricos da identidade de tipos defendem-se desta crítica dizendo que a lei de Leibniz só se aplica àqueles objetos necessariamente idênticos. Porém, a identidade entre tipos de estados mentais e cerebrais é contingente e não necessária.

É comum encontrarmos o seguinte argumento fundado na lei de Leibniz e na introspeção para falsificar a identidade entre estados físicos e mentais:

- 1 Meus estados mentais são introspectivamente conhecidos por mim como estados da minha própria consciência
- 2 Meus estados físicos não são introspectivamente conhecidos por mim como estados da minha própria consciência
- Logo, pela lei de Leibniz,
- 3 Meus estados físicos não são idênticos a meus estados mentais (Churchland, 1984, p. 32)

Tal argumento, porém, revela-se uma falácia. Como é comum em lógica, mostramos por analogia que o argumento é inválido.

- 1 Pelé é conhecido como o jogador do século
- 2 Edson Arantes não é conhecido como o jogador do século
- 3 Logo, Pelé não é idêntico a Edson Arantes.

Portanto, tal raciocínio não serve para falsificar a teoria da identidade. É bem possível que a conclusão seja falsa e as premissas verdadeiras. Observe-se que, nas premissas, estamos falando do nome do objeto. Se ele (objeto nomeado) for o mesmo, é óbvio que uma das premissas é verdadeira e outra é falsa. Porém, parece que no primeiro argumento não é isso o que se deseja. Pretende-se, ao contrário, mostrar que os objetos não são a mesma coisa.

3.3 Teoria de identidade ocorrência

Uma versão mais fraca da teoria da identidade é encontrada na *teoria de ocorrência*. Aqui, ao invés de identificar tipos de estados mentais com tipos de estados cerebrais, seus adeptos, como Sheferd (1983) identificam ocorrências desses estados. Uma *ocorrência* é um elemento particular de um determinado tipo de objetos. A dor, por exemplo, é um tipo de estado mental. Cada dor em particular é uma ocorrência deste tipo. Como no caso da dor, ela pode estar identificada com vários estados cerebrais (um de cada vez), dependendo da circunstância e da dor em questão.

A identificação de ocorrências de estados mentais com ocorrências de estados cerebrais cria uma certa fraqueza na identificação de estados mentais. Se em determinadas

circunstâncias ele pode estar associado a estados cerebrais distintos, como podemos reconhecê-lo via estado físico? No caso da teoria de identidade de tipos, ao saber do estado mental, sabemos também do estado cerebral ao qual está associado. Na teoria da ocorrência isso não acontece. Não há uma relação biunívoca entre estados mentais e cerebrais. Um mesmo estado mental pode estar relacionado a diversos estados cerebrais em momentos, situações ou ocasiões distintas, desde que sejam ocorrências do mesmo tipo. Deste modo, não é possível saber com certeza qual estado mental é o atual apenas observando o estado físico atual.

Como dito, estados físicos distintos podem ser identificados com um mesmo estado mental. Surge daí a dúvida de como explicar que dois estados neurofisiológicos distintos se transformem em ocorrência do mesmo tipo de estados mentais (Searle, 1997, p. 62). O que faz com que estados neurológicos distintos associem-se ao mesmo estado mental? Esta questão tal teoria não consegue responder

3.4 Materialismo eliminativo

Se a teoria da identidade tenta reduzir estados mentais a estados físicos, o materialismo eliminativo, como o próprio nome diz, elimina qualquer característica referente ao mental. Dentre os defensores desta teoria encontram-se Churchland (1981) e Stich (1983). Tais filósofos acreditam na hipótese desta teoria vir a ser uma Psicologia madura, realmente científica. Para tanto, buscam eliminar os termos da Psicologia popular que, segundo eles, tornam os estados mentais incompreensíveis. Tal objetivo será alcançado com o desenvolvimento das neurociências, afirmam eles, buscando novos vocabulários e

conhecimentos sobre o cérebro humano. Quando isso acontecer, não falaremos mais em introspecção, qualidades subjetivas e propriedades internas do sujeito. O homem será tratado como uma simples máquina cujo motor é o cérebro, afirmam os eliminativistas.

A identificação entre estados mentais e cerebrais revelou-se em muitos sentidos pobre em justificações, previsões e satisfação do que pretendia explicar. A doutrina apresentada a seguir procura diminuir um pouco a importância do cérebro em relação aos estados mentais, identificando-os com estados funcionais.

3.5 *Funcionalismo*

Os funcionalistas, dentre os quais encontramos Putnam (1967), Dennett (1978) e Pylyshyn (1984), diminuem um pouco a relevância do cérebro em relação aos estados mentais. Ele é apenas o meio pelo qual estados mentais do ser humano são realizados, mas não é necessário para sua realização. O substrato material de uma entidade qualquer deve apenas ter um certo nível de complexidade para poder “computar” os processos que caracterizam os estados mentais.

A composição física dos sistemas, segundo os funcionalistas, não nos revela muito sobre seus estados mentais (Putnam, 1967). Se ele é constituído de metal, aço ou neurônios não importa para a caracterização de estados e faculdades mentais. Estes, ao contrário, são caracterizados através das relações causais entre si, e entre os *inputs* e *outputs* do sistema no qual estão presentes. Sendo assim, qualquer sistema capaz de apresentar as propriedades causais corretas de um estado mental poderá possuí-lo.



O funcionalismo resolve o problema que a teoria da identidade de ocorrências deixa pendente: o que torna dois estados neurofisiológicos diferentes ocorrências do mesmo tipo de estado mental é o fato de desempenharem a mesma função na vida total do organismo (Searle, 1997, p. 62).

Para tratar dos estados mentais, os funcionalistas utilizam a chamada *Sentença de Ramsey*. Esta determina quais os dados de entrada, relações entre estados mentais e respostas são necessários para a realização de um estado mental. Assim, por exemplo, para dizer que Paulo tem um determinado estado mental x , o fariamos do seguinte modo:

($\exists x$) (Paulo tem $x \wedge x$ é causado pelo *input* $y \wedge x$, juntamente
com y causam os estados mentais z que, juntos causam a ação w)

Desse modo, alguém tem um estado mental quando

... tem um estado que tem certas relações com outros estados, que tem certas relações com outros estados e com *inputs* e *outputs*. Qualquer sistema que satisfaça tais relações, pode ser considerado como tendo o estado mental em questão. (Block, 1980, p. 17)

Ao caracterizar estados mentais como relações funcionais, os funcionalistas em geral tentam escapar de críticas direcionadas ao behaviorismo, como a de não considerar a relação causal entre estados mentais e comportamento.

Os funcionalistas acreditam também terem resolvido o problema da relação entre a mente e o cérebro. Isso porque, para eles, o cérebro é apenas um instrumento pelo qual os estados mentais são realizados. Qualquer entidade com um sistema físico adequado pode ter estados mentais. Por exemplo, podemos supor que um extraterrestre tenha sua constituição física (cerebral) totalmente distinta da nossa e mesmo assim encontrássemos todas as relações de um determinado estado mental do ser humano. Sendo assim, diria o funcionalista, este ET possuiria o estado em questão. Segundo os funcionalistas, “O que é importante para a caracterização da mente não é a matéria da qual a criatura é feita, mas a

estrutura de atividades internas que esta matéria sustenta” (Churchland, 1984, p.37). Para os funcionalistas, sistemas funcionalmente isomórficos (com as mesmas características e relações funcionais) à mente humana, tais como a máquina de Turing, realizada pelo computador digital, têm estados mentais.

O caso acima origina uma primeira crítica ao funcionalismo: o “liberalismo”. Ao contrário dos teóricos da identidade, os funcionalistas permitem uma abrangência muito extensa de sistemas que podem apresentar estados mentais. Citemos o conhecido exemplo da população chinesa (Block, 1978): suponhamos que colocássemos todos os chineses juntos e criássemos entre eles certas relações. Suponhamos ainda que tais relações se constituíssem de tal forma que fossem idênticas às relações de um certo estado mental. Desse modo, tal sistema teria este estado mental. Se este estado fosse uma dor, por exemplo, diríamos que a população chinesa está com dor. É compreensível que os chineses tenham dor, mas não a população como um todo. Assim como neste caso, qualquer outro sistema que possua as mesmas relações entre estados mentais produzidos por determinados *inputs* e produzindo determinados *outputs* tem o estado mental associado a estas relações.

O funcionalismo é também criticado por eliminar algumas características internas ou de natureza qualitativa do sujeito referentes aos estados mentais. Um exemplo geralmente apresentado é o da inversão do espectro. Suponhamos que duas pessoas vejam um mesmo limão em sua frente. Poderia acontecer que uma delas visse tal limão e notasse sua cor verde normalmente. Porém, a segunda pessoa poderia experimentar uma sensação de vermelho ao ver a mesma cor e responder que o limão é verde. O dado de entrada foi o mesmo, as relações internas foram as mesmas e a resposta foi a mesma. Porém, cada um teve uma sensação distinta, contrariando a teoria.

Falamos, na apresentação deste capítulo, que o projeto mecanicista é baseado no funcionalismo e podemos agora justificar tal afirmação. O projeto mecanicista, que é o seguido pela Ciência Cognitiva, procura caracterizar a mente como um sistema mecânico processador de informações. Na Ciência Cognitiva, a mente processa informações ou seguindo regras lógicas e manipulando símbolos ou auto-organizando-se através da relação entre suas partes básicas. Porém, a constituição física do sistema não é fundamental para ele ter ou ser uma mente. O importante é que essa constituição seja de tal modo que é capaz de realizar as relações exigidas para a emergência de estados e faculdades mentais. Desse modo, com o que dissemos acerca do funcionalismo, parece ficar evidente que o mecanicismo funda-se no funcionalismo quando trata da mente.

Segundo a visão mecanicista da mente, um comportamento inteligente pode ser entendido como um processamento adequado de informação. Tal processamento pode ser algorítmico ou auto-organizante, como veremos no quinto capítulo. O que dissemos da mente no parágrafo anterior pode ser dito também da inteligência: um sistema é inteligente quando é capaz de realizar os processos exigidos para a resolução de determinados problemas, não importando essencialmente sua constituição física. Tais processos exigem mais do que a simples análise comportamental ou neurofisiológica de um sistema para saber se ele pode ou não ser considerado inteligente. Exige a análise das relações causais entre os dados de entrada, estados internos e dados de saída desse sistema.

Como dito, a tendência natural é dividir as teorias da mente em dualistas e monistas. Alguns pesquisadores, como Searle (1997), por exemplo, criticam esta divisão. Há uma tendência, afirma ele, de pensar que tudo é ou material ou mental. Mas, o que falar de gols de uma partida de futebol, taxas de juros e governos, por exemplo? (Searle, 1997, p.41) Em



qual destas categorias seriam classificados estes itens? Existem coisas que situam-se em outras categorias além destas duas. O mesmo ponto de vista Searle mantém acerca da mente. Não é necessário que todas as teorias devam ser ou dualistas ou materialistas. Pelo contrário, uma teoria da mente pode ser nem uma nem outra. Um exemplo, é a sua própria, denominada *naturalismo biológico*.

3.6 *Naturalismo biológico*

Um dos principais objetivos de Searle é eliminar a idéia de que tudo deva ser ou material ou imaterial. Para ele, podemos aceitar os fatos óbvios da física - por exemplo, que o mundo é constituído inteiramente de partículas físicas em campos de força -, sem ao mesmo tempo negar os fatos óbvios de nossas próprias experiências - por exemplo que somos todos conscientes e que nossos estados conscientes têm propriedades fenomenológicas irreduzíveis bastante específicas (Searle, 1997).

Para Searle, as visões tradicionais que tentaram caracterizar a mente tomavam a consciência, a intencionalidade, a subjetividade como características que fazem parecer tão difícil a caracterização de estados e faculdades mentais e a resolução do problema mente-corpo. Qualquer explicação satisfatória da mente e da relação entre esta e o corpo deve levar em consideração todas as características acima. Se uma teoria da mente acaba por negar ou deixar de lado alguma delas, deve ter havido algum erro, afirma Searle (1997, p. 45).

O problema da relação mente-corpo, segundo Searle, pode ser resolvido facilmente se adotarmos a postura naturalista biológica. Tal postura implica na criação de uma teoria que



de certa forma é a união entre a teoria molecular dos corpos e a teoria evolucionista da biologia. Para Searle (1984), a mente é causada pelo cérebro, mas não pode ser explicada puramente através dele. Existem características biológicas cerebrais que proporcionam a causa de estados e faculdades mentais. Quais são tais características Searle não explicita.

Desse modo,

... os fenômenos mentais são causados por processos neurofisiológicos no cérebro, e são eles próprios características do cérebro... Os processos e fatos mentais fazem parte de nossa história biológica tanto quanto a digestão, a mitose, a meiose, ou a secreção enzimática. (Searle, 1997, p. 7) O cérebro causa determinados fenômenos 'mentais', tais como estados mentais conscientes e esses estados conscientes são simplesmente características de nível superior do cérebro. (Searle, 1997, p. 25).

De acordo com a postura adotada por Searle, não há uma segunda substância nem propriedades irreduzíveis (por isso não é dualista), tampouco há apenas componentes físicos (por isso não é materialista) quando tratamos de estados e faculdades mentais. Por isto a resolução do problema da relação mente-corpo é facilmente resolvida por ele. A mente é causada pelo cérebro e ao mesmo tempo é uma característica dele².

Pelo fato de não ser dualista, Searle acredita não propor características imateriais, quase incompreensíveis. Por outro lado, por não ser materialista não deixa de lado a mente. Ou seja, leva em consideração estados e faculdades mentais como tais e não apenas como estados físicos ou comportamentais observáveis empiricamente. Dessa forma, acredita ainda não eliminar aspectos introspectivos do indivíduo, podendo explicar o surgimento da consciência e a manutenção de intencionalidade no indivíduo.

² A noção de causalidade deste ponto de vista não deve ser tomada como é tradicionalmente definida, na qual uma de suas características reside no fato de que a causa é diferente do que é causado. É preciso redefinir a noção de causalidade. Searle concorda com isto, mas não propõe uma nova definição. Porém, é possível ainda neste sentido manter esta característica. Segundo Searle (1984), a mente é causada pelas unidades básicas do cérebro, ou seja, os neurônios. Uma vez que ela é causada, é característica não mais destas unidades, mas da rede toda. Ou seja, ela é causada pelo cérebro enquanto microestrutura e é característica da macroestrutura cerebral.

A presença de consciência em uma entidade é uma das características fundamentais para ela poder ser considerada uma mente ou possuir uma mente. Do mesmo modo que as faculdades e estados mentais em geral, a consciência é causada pelo cérebro e é característica dele. Segundo Searle(1997, p. 25-6), a consciência é

... uma propriedade emergente, ou de nível superior, do cérebro, no sentido absolutamente inócuo de 'nível superior' ou 'emergente'... A consciência é uma propriedade mental, e portanto física do cérebro, no sentido em que a liquidez é uma propriedade de sistemas de moléculas.

Como dissemos, Searle afirma que a consciência e estados e faculdades mentais em geral são um produto de certas relações cerebrais. Porém, tais relações não envolvem simplesmente a constituição física do cérebro. Se assim fosse, poderíamos identificar estados mentais a estados físicos. Tal filósofo tampouco concorda que estados mentais podem ser identificados por estados comportamentais ou funcionais. De fato, Searle (1997, p. 98-105) procura mostrar que não existe nenhuma conexão necessária ou suficiente entre estados cerebrais, mentais e comportamentais.

De acordo com Searle (1997), é possível pensar um indivíduo com um cérebro de silício possuindo estados mentais e comportamentais. Ou seja, podemos não ter estados cerebrais (físicos) e mesmo assim ter estados mentais e comportamentos num sistema qualquer. Por outro lado, podemos pensar em um indivíduo em coma, sem nenhum comportamento e mesmo assim sentir dor ou estar consciente. Podemos ainda pensar em um indivíduo que, mesmo exibindo comportamentos associados a um determinado estado mental de fato não apresentasse este estado. É o caso do indivíduo com um cérebro de silício ou uma máquina de Turing, capazes de possuir as relações causais ou funcionais corretas do cérebro que simulassem os *inputs* e *outputs* do indivíduo, porém, sem a presença do estado mental em questão.

Haveria a possibilidade da mente ser simulada por outros sistemas que não o cérebro? Poderiam existir máquinas artificiais com mente de acordo com o naturalismo biológico? Searle afirma que nosso cérebro é uma máquina biológica. Uma vez descobertas as características possuídas pelo cérebro para poder causar estados e faculdades e como isso é feito, ele acredita que é possível construir sistemas artificiais capazes de possuírem mente. Bastaria que esses sistemas fossem capazes de exibir as capacidades cerebrais corretas para a sua causação.

Uma vez que, se o sistema possuir as capacidades cerebrais corretas, pode ter um mente, qual a diferença entre naturalismo biológico e funcionalismo? Assim como nesta teoria, um sistema deveria apenas possuir as relações cerebrais corretas para possuir uma mente.

Uma diferença fundamental entre naturalismo biológico e funcionalismo, ao nosso ver, encontra-se na caracterização dos estados e faculdades mentais. No funcionalismo, os estados mentais são as próprias relações funcionais. No naturalismo biológico, por outro lado, os estados mentais têm como causa as relações cerebrais. Além disso, nesta causação são acrescentadas aos estados mentais certas características subjetivas que não pertencem às relações cerebrais.

Diversas críticas são direcionadas à teoria de Searle. Dentre elas está a de que tal filósofo não explica quais são as características ou capacidades existentes no cérebro responsáveis pela causação da mente. Além disso, ao serem causados, estados e faculdades mentais adquirem certas características próprias. Searle não explica quais e como são adquiridas estas características. Searle concorda com estas críticas e defende-se dizendo que o problema de descobrir estas coisas não faz parte de seu trabalho como filósofo. Tal



incumbência é atribuída às neurociências que devem ser as responsáveis por estas descobertas.

Como afirma Searle (1997), as teorias materialistas em geral deixam de lado a mente quando tratam dela. Todas elas têm seus problemas quando tentam reduzir estados mentais a estados físicos ou comportamentais observáveis empiricamente e objetivamente. Por que então os cientistas tendem a criá-las ou aceitá-las?

Uma primeira justificativa, dentre as três apresentadas aqui, é a tentativa de fuga do dualismo. Não é possível atualmente aceitar crenças nem religiosidade na ciência, quando se quer verdadeiramente descobrir ou conhecer o objeto de pesquisa (Churchland, 1984, p. 39). Estamos num momento científico em que não se pode aceitar algo imaterial, além das possibilidades do limite de conhecimento humano.

Desde Descartes principalmente, houve uma tendência de separar o mundo em duas metades: o *mental* e o *físico* (Searle, 1997, p. 41). Qualquer coisa parece dever estar em uma e apenas uma destas partes. Desse modo, tomamos como contrários termos do tipo *corpo x mente*, *físico x mental*, *matéria x espírito*. Isso torna clara a segunda razão para a adoção de uma teoria materialista da mente: os materialistas precisam escolher um dos lados dos pares. A decisão, obviamente pende para o lado esquerdo das duplas.

Um terceiro motivo para a escolha de uma teoria materialista é a tendência objetivista existente na ciência atual. Pensa-se hoje que a realidade é objetiva. Tudo o que é real pode ser igualmente acessível a todos. Isso faz com que queiramos conhecer os estados mentais como objetivamente observáveis. Com isso, não os definimos como pertencentes a um sujeito, mas como algo objetivo.



Como descrevemos, qualquer uma das caracterizações propostas pelos materialistas faz com que a mente seja materializada. Tornam assim seu objeto de pesquisa possível de manipulação empírica. Porém, mesmo materializando a mente, os cientistas têm problemas para observá-la diretamente. Isso porque, se a mente é o cérebro, seria preciso manipular o corpo humano, o que pode ser eticamente desaconselhável. Se ela é como um programa de computador, o problema é saber onde podemos encontrá-lo, produzi-lo ou criá-lo. Daí surge uma hipótese fundamental de alguns pesquisadores como Newell e Simon (1961): a mente pode ser imitada, simulada e explicada através de *modelos* (estruturas que teriam o mesmo funcionamento da mente). Com a adoção destes modelos, dentre os quais um de grande relevância é o computador digital, surge a Ciência Cognitiva.

4 A mente na Ciência Cognitiva

4.1 A caracterização da Ciência Cognitiva

A Ciência Cognitiva nasce com o intuito de responder empiricamente, através do uso do computador, questões que os primeiros filósofos já haviam colocado sobre o estudo da mente. Dar uma definição precisa deste novo campo de pesquisa não é uma tarefa simples. Trata-se de uma ciência nova e, talvez por isso, seu campo de atuação, método e objeto ainda não estejam bem delimitados.

Gardner (1995, p. 19-20) procura definir Ciência Cognitiva 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 emprego. Embora o termo Ciência Cognitiva seja às vezes ampliado, passando a incluir todas as formas de conhecimento - tanto



animado como inanimado, tanto humano como não humano - aplico o termo sobretudo a esforços para explicar o conhecimento humano. Interessa-me saber se questões que intrigavam nossos ancestrais filosóficos podem ser definitivamente respondidas, ilustrativamente reformuladas, ou permanentemente abandonadas. Hoje a Ciência Cognitiva tem a chave para decidir.

Segundo Churchland (1984, p i), "... uma das funções principais da Ciência Cognitiva é estudar quais são os elementos básicos da atividade cognitiva e como eles podem ser implementados em sistemas físicos reais." Deste modo, a preocupação inicial da Ciência Cognitiva não é a caracterização de estados e faculdades mentais em geral, mas sim, à caracterização e simulação de processos cognitivos.

Dentre as características principais da Ciência Cognitiva podemos citar três em especial: a interdisciplinaridade, o representacionalismo e o uso de modelos computacionais da mente.

É importante nesta nova ciência que as diversas áreas que estudam a mente humana interajam e tentem se comunicar. Cada ciência pode colaborar em algum sentido para o desenvolvimento do conhecimento sobre aspectos cognitivos. Um dos objetivos desta interdisciplinaridade é tentar encontrar um vocabulário e método únicos para o estudo da mente.

Há na Ciência Cognitiva quase o consenso de que nosso conhecimento do mundo é mediado por representações mentais. Deste modo, os pesquisadores desta área de pesquisa acreditam na necessidade do entendimento da natureza destas representações para o entendimento da mente. Daí o fato da segunda característica ser o representacionalismo. Uma representação é uma versão modificada do mundo (Charniak, 1993, p.8) ou um modo de reter conhecimentos sobre ele, que pode ser representado através de símbolos, esquemas, imagens, idéias.



O estudo das representações mentais exige a análise e manipulação da mente, o que pode não ser aconselhável. Este é um dos motivos para a adoção da terceira característica da Ciência Cognitiva: os modelos computacionais, abordados nos próximos dois capítulos, podem servir como sistemas que simulam e explicam processos cognitivos. Com eles, não é preciso observar a mente propriamente dita.

A caracterização da mente na Ciência Cognitiva adequa-se ao moldes mecanicistas há pouco comentados. Ela é basicamente um sistema processador de informações e funciona mecanicamente, como veremos nas próximas duas seções, onde introduzimos a IA e RNA.

4.2 A Inteligência Artificial

Podemos dizer que as pesquisas mais intensas para a construção e adoção de uma máquina concreta como modelo da mente humana surgiram com o início da IA. Segundo Gardner (1995, p. 159),

... o nome Inteligência Artificial foi pronunciado em 1956, quando alguns cientistas como John McCarthy, Marvin Minsky, Allen Newell e Herbert Simon discutiram as possibilidades de se produzir programas computacionais que pudessem 'se comportar' ou 'pensar' de maneira inteligente, como por exemplo solucionar problemas, reconhecer padrões, tomar parte em jogos e raciocinar logicamente. Estes cientistas baseavam-se na hipótese de que todo aspecto de aprendizagem ou de qualquer outra faceta da inteligência pode ser descrito de forma tão precisa que se pode fazer com que uma máquina o simule.

Para Charniak (1985, p. 113), "IA é o estudo de faculdades mentais através do uso de modelos computacionais. Seu objeto é a mente, entendida como um sistema processador de informação." Já para Dreyfus (1993, p. 39), um dos críticos desta teoria, "A IA é a tentativa de simular o comportamento humano inteligente utilizando-se técnicas de programação que precisam demonstrar pouca ou nenhuma semelhança com os processos mentais humanos".

A grande motivação dos primeiros cientistas desta corrente da Ciência Cognitiva foi o fato de que, pela primeira vez, sistemas artificiais conseguiam realizar com sucesso comportamentos que até então eram de exclusividade humana. Dentre estavam a atividade de resolução de problemas do tipo lógico-matemáticos e a participação em jogos como xadrez.

A Inteligência Artificial pode ser historicamente dividida em duas versões: IA *forte* e *fraca*. De acordo com a *Inteligência Artificial Forte*, a mente é um programa de computador. Ambos, computador e mente, devem ser concebidos como um sistema simbólico - entidade que processa, transforma, elabora e manipula símbolos de vários tipos, processando informações no decorrer do tempo, numa ordem mais ou menos lógica. Para esta versão, a mente está para o cérebro assim como o *software* está para o *hardware* do computador. Como afirmam Newell & Simon (1972, p. 19), "... o homem é um sistema de processamento de informação, pelo menos quando está resolvendo problemas" e um computador pode perfeitamente simular tal sistema³.

A *Inteligência Artificial Fraca* não adota a hipótese da identidade entre mente e programa. Seus adeptos afirmam existir apenas uma semelhança entre mente humana e programa computacional. Os programas de computador são um bom modelo da mente, podendo explicar seu funcionamento e suas características. Tanto a mente quanto o programa manipulam símbolos e seguem regras lógicas. Porém, não são exatamente o mesmo objeto.

³ Newell & Simon (1972), porém, algumas vezes parecem pender um pouco para a IA fraca, por afirmações do tipo: "O fato de que existem realizações computacionais de um sistema de processamento de informação não prova, com certeza, que os homens são bem modelados por um tal sistema".

A Inteligência Artificial não teve em seus primórdios muito inclinada a simular estados mentais como dor, sentimentos, crenças. Talvez porque este trabalho seja o mais complicado dentre todos os referentes ao estudo da mente. A preocupação dos cientistas da IA esteve voltada até o momento para a simulação cognitiva. Seu objetivo é saber como o homem e sistemas em geral conhecem, raciocinam, pensam. Como afirmam Newell & Simon (1972, p.1), "O objetivo deste livro é avançar nosso entendimento de como o homem pensa". Tanto esta dupla de cientistas, como a grande maioria dos pesquisadores da IA estão interessados fundamentalmente em compreender a atividade de resolução de problemas. Para eles, o pensamento pode ser explicado por meio de uma teoria de processamento de informações, que envolve a manipulação de símbolos através do seguimento de regras.

Para explicar os processos pelos quais o ser humano passa ao tentar resolver problemas, muitos cientistas da IA, tais como Newell, Shaw & Simon (1958) constróem sistemas artificiais de processamento de informações. Estes sistemas (modelos) são construídos a partir dos pressupostos representacionalistas da IA. Uma vez que os modelos conseguem resolver problemas precisamente descritos, seus construtores tentarão explicar os processos pelos quais indivíduos passam quando realizam este tipo de problemas, tais como jogar xadrez, provar teoremas lógicos e resolver quebra-cabeças. Falaremos mais destes modelos no quinto capítulo deste trabalho.

4.3 O Conexionismo

O *Conexionismo* é a outra grande vertente da Ciência Cognitiva a ser comentada neste trabalho. Dentre muitos nomes, alguns dos principais ligados a esta área de estudos estão Rosenblatt (1962), Hopfield (1982), Kohonen (1987), McClelland & Rumelhart (1988), Caudill & Butler (1992). Segundo os conexionistas, um cérebro humano é formado por mais de uma centena de bilhão de elementos computadores chamados neurônios. Esta rede de neurônios é responsável por todos os fenômenos que chamamos pensamento, emoção e cognição. Desse modo, para simular ou explicar processos cognitivos humanos, é preciso levar em consideração o estudo do cérebro, suas características, funcionamento, suas partes constituintes e as relações estabelecidas entre elas.

Conexionistas como os citados no parágrafo anterior têm por fim criar sistemas inspirados no cérebro humano para simular e explicar, dentre outras coisas, comportamentos humanos inteligentes. Porém, mesmo assim, tais cientistas estão conscientes que as redes neurais artificiais são apenas uma aproximação muito limitada do cérebro humano (Caudill & Butler, 1992, v. 1, p. 4; McCulloch & Pitts, 1943, p. 117). Um simples exemplo disso é o fato dos conexionistas chamarem as partes constituintes das RNA de nódulos (neurodes) e não neurônios (neuron).

A mente, para os conexionistas, não é simplesmente um sistema manipulador de símbolos e seguidor de regras lógicas, como afirmam os cientistas da IA. Em vez disto, ela é entendida como um conjunto de neurônios relacionados entre si, produzindo estados mentais, originando conhecimento, aprendizagem, comportamento inteligente.

Um dos objetivos da IA, é construir programas que simulem comportamentos humanos de um modo algorítmico (seguindo regras lógicas e manipulando símbolos), desconsiderando aspectos físicos e biológicos do ser humano. Os connexionistas, por sua vez, afirmam que o processo de simulação deve ser construído tomando-se por base as conexões entre neurônios.

Segundo Gardner (1995, p. 414), o Conexionismo ou PDP (processamento distribuído em paralelo) pode ser assim caracterizado:

... em vez de operações seriais ou computações sobre símbolos ou cadeias de símbolos, em vez de 'executivos', 'interpretes' e 'unidades centrais de controle', a abordagem PDP tipicamente postula milhares de conexões entre centenas de unidades (em princípio, a abordagem pode ser estendida a milhões ou mesmo bilhões de conexões). As redes resultantes apresentam a sinalização de excitações e inibições de uma unidade para outra. 'Percepção', 'ação' ou 'pensamento' ocorrem em consequência da alteração das forças (ou pesos) das conexões entre estas unidades. Uma tarefa é concluída ou um *input* processado quando o sistema finalmente se 'acomoda' ou 'relaxa' (pelo menos provisoriamente) em um conjunto satisfatório de valores ou 'estados estáveis' - em suma, em uma 'solução.'

A abordagem connexionista da mente, segundo seus defensores, proporciona um novo modo de pensar sobre a percepção, memória, aprendizagem, pensamento e sobre os mecanismos computacionais básicos para o processamento inteligente de informações em geral.

O início da história do Conexionismo deve talvez ser contada tomando como um dos personagens principais o neuropsiquiatra McCulloch (1965). A preocupação de McCulloch era, por assim dizer, filosófica: buscava descobrir como sabemos e como desejamos. Para ele, os eventos mentais não são compreendidos como submissos a uma entidade controladora abstrata. Ao contrário, são explicados a partir de conjunções sinápticas. Com esse objetivo e tendo em vista a idéia da encarnação da mente (tentativa de identificação entre mente e cérebro), McCulloch & Pitts (1943) constróem uma rede neural artificial, ainda seguidora de regras lógicas, tomada como um modelo do cérebro.

A idéia de McCulloch era representar cada atividade mental por alguma proposição lógica. No caso da dor, por exemplo, a rede a simulará através da conexão entre seus nódulos. Estas conexões serão equivalentes a uma determinada proposição temporal do cálculo proposicional da lógica clássica, como veremos mais detalhadamente na segunda seção do capítulo 3.

Segundo McCulloch & Pitts (1943), grande parte das atividades mentais poderiam ser descritas em termos de conexões e estas em termos de proposições lógicas. Logo, por transitividade, tais atividades poderiam ser descritas por meio de proposições lógicas. A idéia que realmente liga McCulloch ao Conexionismo é a de análise dos fenômenos mentais através de conexões neuronais.

Os conexionistas admitem que seus modelos operam mais eficientemente com a percepção e outros processos de nível inferior. Segundo Gardner (1995, p. 417),

Mesmo aqueles que simpatizam com abordagens PDP admitem que elas operam mais eficientemente com a percepção e outros processos 'de nível inferior' (subsimbólicos) do que com solução de problemas de grande escala, detecção de problemas, invenção e outros empreendimentos 'simbolicamente carregados'. Como Rumelhart e seus colegas colocam sucintamente, o que é difícil descrever na estrutura PDP são 'o processo do pensamento, os conteúdos da consciência, o papel dos processos seriais, a natureza dos modelos mentais, as razões para as simulações mentais e o importante papel sinérgico da linguagem no pensar e na formação de nosso pensamento.'

Nos modelos conexionistas, as informações são codificadas não em estruturas simbólicas, mas através dos padrões de ativação das conexões entre as unidades. Smolensky (1987) usa o termo subsimbólico para designar processos (estados) físicos que, de alguma forma, participam como substratos dos estados simbólicos abstratos. Ou seja, os subsímbolos são constituintes básicos dos processos simbólicos. São também menos primitivos na escala de abstração do sistema cognitivo. De outra forma, constituem uma estrutura intermediária entre os planos neural e simbólico.

Como dissemos, tanto a IA quanto o Conexionismo acreditam na possibilidade de simulação, de um modo ou de outro, de estados e faculdades mentais, como o pensamento, através da utilização de modelos computacionais. Além de simular propriedades mentais, alguns pesquisadores, principalmente da IA, acreditam estar explicando-as. Os modelos criados pelos cientistas cognitivos, segundo eles, podem também exibir comportamentos considerados inteligentes. Nos próximos dois capítulos, apresentaremos alguns desses modelos, com o fim de conhecê-los melhor, expondo o modo de funcionamento, características, potencialidades e limitações. Com isso, acreditamos poder entender mais claramente os requisitos de cada corrente para um comportamento ser inteligente e porque afirmam, cada uma a seu modo, que suas máquinas são inteligentes. Os modelos da mente apresentados serão a máquina de Turing, a rede de McCulloch-Pitts, o perceptron e a rede de Kohonen.



Cap. 2

Modelos algorítmicos da mente

1 Apresentação

O termo máquina suscita na mente de muitas pessoas a lembrança de um inimigo, uma ameaça à ordem política, social, familiar. Isso é compreensível por, no mínimo, dois motivos: a sensação de estar sendo tirado do centro do universo o tocante a certas habilidades. A máquina é capaz de apresentar comportamentos antes de exclusividade humana. Por isso, o ser humano coloca-se na condição de “perda de privilégios”. O segundo motivo está relacionado ao fato de que muitas pessoas já perderam seus empregos para alguma máquina. Como consequência, foram privados de amigos, família, condição social e econômica.

É indubitável que as máquinas trazem muito benefícios ao ser humano – como substituir em trabalhos perigosos e proporcionar-lhe um certo conforto. Por outro lado, elas podem lhe causar algum prejuízo, como o desemprego, por exemplo. Porém, é claro que elas não podem ser responsabilizadas por nenhum dano provocado aos seus criadores. Elas



não têm vontade própria ou liberdade para querer ou escolher fazer ou deixar de fazer algo. São simplesmente criadas e manipuladas pelo ser humano. Portanto, se há um lobo do homem nesta história, ele é o próprio homem, que não as utiliza unicamente para o bem da humanidade.

Ainda que os temas acima sejam interessantes, não trataremos aqui dos problemas sociais, morais ou econômicos criados pelo surgimento das máquinas em geral. Em vez disso, pretendemos apenas apresentar uma classe específica delas, que foram tomadas como modelos da mente para o estudo de estados e faculdades mentais.

Neste capítulo, apresentamos um tipo de modelos da mente, que denominamos *modelo algorítmico* em oposição ao *modelo conexionista* da mente, apresentado no capítulo seguinte. O modelo apresentado a seguir, a *máquina de Turing*, funciona de um modo mecânico, determinado, seguindo regras e manipulando símbolos. Em primeiro lugar, discutimos a origem da máquina de Turing. Em seguida, apresentamos sua definição e alguns exemplos de máquinas. Definimos então a *máquina de Turing Universal*, capaz de realizar qualquer comportamento computável. Por fim, discutimos algumas questões referentes a estas máquinas, como a necessidade de um agente e de uma memória ilimitada pertencentes a ela.

Não visamos a construção física ou concreta da máquina, como o faria um engenheiro, por exemplo. Ao contrário, a construiremos abstratamente. Esta é uma característica do cientista teórico. Para ele, a parte física da máquina está num plano secundário. O teórico a constrói abstratamente e testa sua eficiência. Se a máquina concreta (como o computador digital, por exemplo) não conseguir realizar a ação da máquina abstrata, o problema deverá ser atribuído ao aparato físico. Poderia ser um problema tecnológico, como por exemplo a

incapacidade atual de criação (física) de uma memória potencialmente infinita. Isto, porém, não teria nenhuma consequência sobre a máquina abstrata, uma vez testada e aprovada no plano teórico. Estaríamos simplesmente diante de um problema de engenharia.

2 A origem da Máquina de Turing

A tentativa de formalização dos processos de pensamento humano⁴ pode ser encontrada já na Grécia, há mais de dois milênios e meio. Formalizar o pensamento humano significa criar sistemas nos quais o conhecimento é obtido mecanicamente (sintaticamente) ou seja, a partir de verdades necessárias, mediante aplicação de regras lógicas e manipulação de símbolos. Euclides, por exemplo, criou um sistema formal para a obtenção de verdades geométricas, demonstradas a partir de verdades mais básicas através de regras bem determinadas. Pouco tempo depois, Aristóteles montou uma teoria com as mesmas características para a demonstração de silogismos.

Na Idade Moderna, pensadores como Hobbes também acreditavam que o pensamento devia dar-se mecanicamente. Hobbes chegou a ser considerado o avô da IA, por afirmar que *pensar é calcular*.

Neste século, surgem o logicismo e o formalismo, com o intuito de formalizar logicamente o pensamento, principalmente o matemático. Para isso, deveriam criar um sistema com o qual se pudesse demonstrar toda verdade matemática, seguindo regras bem determinadas e manipulando símbolos. Segundo Penrose (1993, p. 57), "... a opinião dos



formalistas é a de que é possível deixar de lado os significados dos enunciados matemáticos, considerando-os apenas como seqüências de símbolos.”

O objetivo destes cientistas, dentre os quais encontramos matemáticos como Russell & Witehead (1925), Hilbert & Ackermann (1962) e Frege (1950), era inicialmente mostrar que os sistemas formais seriam consistentes e completos. Além disso, acreditavam que a matemática, e em especial, a aritmética, era passível a uma formalização lógica.

Alguns dos primeiros cientistas desse programa foram matemáticos como Cantor, Poincaré e Frege. Este último, por exemplo, ficou muitos anos redigindo um trabalho para assentar a matemática em bases lógicas seguras. Com isso, pretendia garantir a consistência e completude desta ciência⁴. Frege pretendia definir os conceitos matemáticos através de conjuntos. Quando este matemático estava para terminar sua pesquisa, Russell produziu um paradoxo a partir do sistema fregeano, através da construção de um conjunto específico. Tal conjunto, denominado R, é definido como *o conjunto de todos os conjuntos que não pertencem a si mesmos*.

Pensando em sanar o problema encontrado na teoria de Frege, Russell & Witehead (1925), construíram um sistema matemático de axiomas altamente formalizado. Tal sistema foi apresentado na obra monumental *Principia matemática*. As regras do sistema foram introduzidas com o cuidado de não ser possível gerar paradoxos. Porém, seu sistema mostrou-se muito complexo e complicado de manusear. Pensando em sanar esta dificuldade, Hilbert & Ackermann (1962) construíram um sistema mais funcional e abrangente

⁴ O pensamento a que nos referimos aqui é aquele relacionado à obtenção do conhecimento. Ao falar do estudo do pensamento, estamos falando do estudo dos processos cognitivos humanos pelos quais o conhecimento é obtido.

⁵ Uma teoria é consistente quando não demonstra contradições, ou seja, uma fórmula e sua negação como sendo teoremas. A teoria é completa quando demonstra todas as suas verdades.

objetivando sustentar a matemática em bases inatacavelmente seguras. As esperanças dos formalistas e logicistas acabaram quando Gödel mostrou que o projeto destes cientistas estava fadado ao fracasso (como falaremos à frente).

Segundo os cientistas das correntes citadas acima, se a matemática, ou o conhecimento matemático, pudesse ser reduzida à lógica e formalizada, então o pensamento matemático também o seria. Desse modo, a prova de qualquer afirmação sobre a realidade, que pudesse ser transcrita em linguagem matemática, reduziria-se à mera computação. Como dissemos, o pensamento aqui está relacionado com a obtenção do conhecimento matemático. Provar afirmações sobre a realidade é obter novos conhecimentos sobre ela. Assim, o pensamento matemático seria mecânico e poderia ser reduzido a um cálculo.

Para saber se as verdades matemáticas poderiam ser resolvidas mecanicamente (ou seja, provadas simplesmente através da manipulação simbólica e seguimento de regras lógicas), Hilbert propôs o seguinte problema: dada qualquer fórmula do cálculo de predicados, existe um método (mecânico) que estabeleça se ela é um teorema ou não deste cálculo e em sua linguagem? Se houvesse este método, seria possível provar sintaticamente as verdades lógicas. Como a matemática seria reduzida à lógica, seria também possível demonstrar sintaticamente as verdades matemáticas.

O método pelo qual os cientistas acreditavam que as verdades devessem ser demonstradas (denominado algoritmo ou procedimento efetivo) pode ser exposto informalmente nos cinco seguintes pressupostos:

- *1. Um algoritmo é um conjunto finito de instruções. (Qualquer algoritmo matemático clássico, por exemplo, pode ser descrito num número finito de palavras em português.)
- *2. Há um agente computacional, geralmente humano, capaz de reagir às instruções e realizar as computações.
- *3. Há facilidades para criar, armazenar e recuperar passos numa computação.



*4. Seja P um conjunto de instruções como em *1 e L um agente computacional como em *2. Então L reage a P de tal modo que, dado qualquer *input*, a computação é realizada num padrão de passos discretos, sem o uso de métodos contínuos ou criações análogas.

*5. L reage a P de tal forma que a computação é realizada deterministicamente, sem recorrer a métodos randômicos ou artificios, como dados por exemplo. (Rogers, 1964, p. 19)

Em outras palavras, um algoritmo é um conjunto finito de operações que determinam o comportamento de um *agente*. Este deve ser capaz de segui-las, uma de cada vez, realizando, assim, mecanicamente alguma função. Embora o número de instruções do algoritmo deva ser finito, não podemos limitá-lo. É sempre possível haver um algoritmo que necessite uma ou mais instruções que o número limitado para computar a função. O mesmo acontece com o tempo de computação do algoritmo e com o tamanho de seus dados de entrada.

Alguns autores definem algoritmo como uma seqüência (e não um conjunto) de instruções. Isso significa que o comportamento do agente depende da ordem das instruções, como é o caso da máquina URM de Sheferdson & Sturgis (1963). Esse não é o caso da máquina de Turing, cujo comportamento é determinado pelo seu estado atual e do símbolo lido, como comentaremos na próxima seção.

Com o propósito de provar a completude e a consistência da lógica e da matemática, alguns cientistas criaram ou aperfeiçoaram algumas teorias. Uma delas foi a teoria das funções recursivas, cuja noção básica é a de função. Uma função pode ser definida como uma relação entre dois conjuntos A e B . Os elementos de A (alguns ou todos) estão relacionados com os elementos de B de acordo com alguma regra. Cada elemento de A pode estar relacionado com apenas um elemento de B . Nos textos de matemática é comum definir uma função f como um conjunto de pares ordenados tal que, se $\langle x, y \rangle \in f$ e $\langle x, z \rangle \in f$ então $z = y$, e $f(x)$ (resultado de x na função f) é definido por meio de uma regra.



O primeiro teórico das funções recursivas foi o matemático Dedekind. Contudo, foi Gödel quem criou, na década de 1920, a primeira grande teoria formal das funções recursivas. Em seu sistema, uma função é recursiva se pode ser construída a partir de funções básicas através de um conjunto de regras.

Depois de Gödel, surgiram muitos outros critérios formais para determinar quais funções são recursivas. Nestes critérios, porém, foram criados métodos para computar estas funções e não apenas demonstrar quais eram elas. Dentre esses sistemas, podemos citar o de Post, Markov, Turing e o de Church (todos eles descritos por Cutland, 1992). A idéia de cada sistema é a mesma: uma função é recursiva ou computável se, dado qualquer elemento do conjunto A , é possível encontrar, através de um método mecânico, formalmente definido em cada sistema, um elemento de B ao qual aquele esteja relacionado. Um desses métodos foi o do matemático Turing (1936), segundo o qual uma função é computável, se pode ser realizada por uma máquina de círculo livre (máquina de Turing). Isso significa que, dado qualquer elemento do domínio de uma função, a máquina encontra mecanicamente uma resposta correta para ele.⁶

Turing (1936) afirmou que toda função naturalmente (intuitivamente) computável o é também computável por uma máquina de círculo livre. Como Turing estava relacionando os planos intuitivo e formal não podia provar tal afirmação. Um vez que Church criou a mesma tese, tal afirmação foi denominada *tese de Church-Turing*.

As afirmações sobre o mundo resumem-se à atribuição de predicados aos objetos desse mundo. Sendo assim, a relação entre a teoria das funções recursivas com o logicismo

⁶Além de pensar na matemática e no problema de Hilbert, Turing compara o homem com a máquina, proporcionando um novo campo de discussões em filosofia da mente e criando um modelo da mente, posteriormente adotado pela Inteligência Artificial.

e formalismo situa-se no critério (mecânico) de decisão desses predicados (ou seja, saber quais objetos estão ou não sujeitos a eles). Cada um deles está relacionado com uma função denominada *função característica*. Para um predicado n -ário (relaciona n indivíduos), sua função característica será:

$$f(x_1 \dots x_n) = \begin{cases} 1, & \text{se } x_1 \dots x_n \text{ satisfaz o predicado} \\ 0, & \text{caso contrário} \end{cases}$$

Para uma sequência de objetos satisfazer o predicado é preciso que possamos construir, através do método mecânico, a relação prescrita por esse predicado a tal sequência. Se assim não fosse, poderíamos ter uma circularidade no sentido de que, para saber se qualquer sequência de objetos satisfaz ou não ao predicado, analisamos a computabilidade de sua função característica. Por outro lado, para analisar a computabilidade de sua função característica, precisamos saber se cada sequência de objetos satisfaz ou não ao predicado.

A decidibilidade do predicado é analisada do seguinte modo: um predicado é decidível se e somente se sua função característica é recursiva (computável). Se há um método mecânico para calcular a função característica de um predicado P , há também um método para decidir em qualquer ocasião, dado um elemento qualquer, se ele satisfaz (ou não) a P , sem dar margem à dúvidas ou arbitrariedades no critério de decisão. Dada a função, se houver um método mecânico segundo o qual, para qualquer elemento do domínio da função, é possível, através desse método responder 0 ou 1, o predicado é decidível.

Cada sistema para decisão das funções computáveis foi denominado uma *noção de computabilidade*. Um dos resultados interessantes provados a respeito desses sistemas é que

todos eles computam exatamente as mesmas funções. Porém, nenhum deles computa a função característica do problema de Hilbert (cujo predicado é *ser teorema*). Em 1936, Turing mostrou que esse problema era insolúvel.

A idéia de Turing era substituir a questão inicial do problema de Hilbert “existe um procedimento efetivo capaz de decidir se qualquer fórmula do cálculo de predicados é ou não teorema” por “existe uma máquina de Turing...”. Para isso, buscou encontrar um algoritmo para determinar se uma máquina qualquer pararia ou não quando aplicada a um número qualquer. Isso porque, se fosse possível encontrá-lo, para toda fórmula do cálculo de predicados apresentada à uma determinada máquina, saberíamos com certeza se ela é ou não teorema. Turing (1936) demonstrou que não existe um procedimento para resolver tal problema, denominado *problema da parada*. Ou seja, não podemos saber, de um modo algorítmico, se a máquina responderá se uma fórmula é ou não teorema.

Embora não seja possível existir um procedimento mecânico para resolver o problema da parada, podemos muitas vezes saber (não algorítmicamente) quando uma máquina pára ou não. De fato, Turing mostrou que, para qualquer fórmula apresentada à máquina, se ela for um teorema, a computação pára e a máquina encontra sua demonstração. Porém, se a fórmula não for um teorema, a máquina nunca pára. É por isso que o problema de Hilbert é insolúvel. Não é possível encontrar um procedimento efetivo em que podemos ter certeza de que, para cada fórmula, podemos saber se ela é ou não teorema.

O sonho logicista ruiu quando Gödel demonstrou que, qualquer sistema formal suficientemente complexo para abranger a aritmética, se é completo, é inconsistente. Em outras palavras, se o sistema não gerar contradições, então haverá no mínimo uma sentença verdadeira indemonstrável nesse sistema. A verdade destas proposições é indecidível apenas



pela utilização de axiomas e regras desse sistema. Como consequência, não é possível fazer matemática apenas sintaticamente, simplesmente seguindo regras lógicas e manipulando símbolos. Conclui-se daí que nem toda verdade matemática pode ser resolvida algoritmicamente. Além disso, o pensamento matemático não pode ser puramente mecânico. Segue-se daí que, “para demonstrar a verdade de certos enunciados, é necessário mais do que ser capaz de seguir regras e manipular símbolos” (Penrose, 1993).

Estes resultados originaram uma objeção (matemática) à suposição de Turing de que a máquina pensa, uma vez que a máquina de Turing tem seu modo de funcionamento totalmente algorítmico. Existem verdades que ela não é capaz de provar. Turing (1950) defende-se dizendo que isso não representa um problema à hipótese de que o pensamento possa ser mecanizado. Isso porque existem muitas verdades que o ser humano também não é capaz de demonstrar. Esta deficiência é até favorável à hipótese mecanicista sobre o pensamento, como veremos no quinto capítulo.

Se, por um lado Turing ajudou a abalar o sonho logicista, por outro, provocou um grande progresso na ciência da computação e na própria lógica e matemática. Foi ele um dos primeiros a produzir uma teoria formal para a criação de programas computacionais. Originou com isso, a criação dos atuais computadores digitais, adotados como modelo concreto da mente pela Inteligência Artificial. Para os cientistas desta corrente, a máquina exibe um comportamento criativo, pensa, é inteligente, tem estados mentais e futuramente poucos compartimentos do intelecto permanecerão fora do seu domínio (Minsky, 1967, p. 2).

Como dissemos, os resultados apresentados por Gödel e pelo próprio Turing abalaram o ideal logicista e formalista. Porém, nem por isso acabou com o ideal de mecanização dos

processos de pensamento humano. Se o pensamento pudesse ser mecanizado, a máquina de Turing poderia ser capaz de pensar e exibir comportamento inteligente. Isso porque ela funciona mecanicamente e porque o comportamento inteligente está relacionado à resolução de problemas (computação de funções). Passemos a seguir a investigar o funcionamento desta máquina.

3 A definição da Máquina de Turing

As máquinas, também denominadas *autômata* (que significa pôr-se em movimento por meios mecânicos) dividem-se em *finitas* e *infinitas*. As finitas são assim chamadas porque possuem limites, como por exemplo o de possuírem uma memória limitada. As infinitas, por sua vez, possuem uma memória potencialmente infinita. São capazes de computar funções mais complexas do que as outras. Apresentamos, a seguir esses dois tipos de máquina, detendo-nos mais no segundo, por dois motivos: por ser mais potente e, com isso dar margem a discussões mais relevantes sobre suas limitações e capacidades; e porque os autômata finitos podem ser explicados como máquinas “infinitas” limitadas.

3.1 Autômata finitos

As máquinas limitadas são geralmente definidas, do ponto de vista do usuário ou do ambiente, como uma caixa fechada com canais de entrada e saída, tal como visualizado na fig. 1. A caixa é preta se o seu interior é desconhecido. A parte física ou a composição física que corre pelos canais não é de grande importância na teoria dos autômata. O que interessa é a definição abstrata da máquina e não sua construção material. É de maior relevância o seu

funcionamento, a realização (algorítmica) de uma determinada função, não importando se tal máquina seja feita de fios, neurônios ou latas de cerveja.

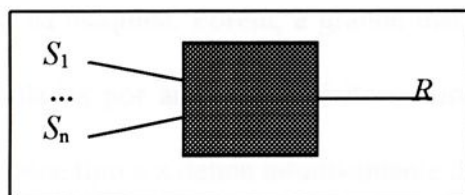


Fig. 1 - Estrutura de um autômato finito

A entrada S da máquina, visualizada na fig. 1, possui certos estados ou sinais $s_1 \dots s_n$ e a saída R possui um canal de saída r_i . O usuário, ou operador da máquina, no momento inicial de sua operação, determina seu estado interno E , que num momento t qualquer é chamado $E(t)$. O estado da máquina é necessário para definir seu comportamento. A partir do dado de entrada apresentado a ela e de seu estado atual é que a máquina “saberá” qual o próximo passo a dar e para qual estado deve ir. Esse novo estado, juntamente com o dado de entrada atual determinará o novo comportamento, novo estado e, assim sucessivamente. A resposta r_i da máquina depende unicamente de seus estados internos e de seus sinais $s_1 \dots s_n$ (dados de entrada).

Os autômatos finitos, assim como os infinitos, devem agir discretamente, ou seja, dando saltos de uma instrução a outra. Desse modo, o tempo é discreto e pode ser representado pelos números naturais, não havendo nada entre um tempo e seu sucessor. Isso faz com que a máquina, ao realizar uma instrução, opere ou *comporte-se* de acordo com o que a instrução determina e pule de uma para outra, sem passos intermediários.

Como todo autômato finito é uma máquina de Turing particular com uma capacidade limitada de memória, passamos a seguir à explicação dos autômatos infinitos. Ao explicar o seu funcionamento, estaremos também explicando o funcionamento dos autômatos finitos.

3.2 Autômatas finitos

Os autômatas finitos podem realizar apenas aquelas funções que não exigem uma capacidade infinita de memória na máquina. Porém, a grande maioria das funções têm esta exigência e só podem ser calculadas por autômatas infinitos. Turing foi o primeiro a criar (abstratamente) uma máquina desse tipo e a define intuitivamente do seguinte modo:

Podemos comparar um homem no processo de computação de um número real [uma função] com uma máquina que é apenas capaz de um número finito de condições $q_1, q_2, \dots, q_R$ chamadas ' m -configurações' [estados internos]. À máquina pertence uma 'fita' (análoga a um papel) que a percorre, e é dividida em seções (chamadas quadrados) cada uma capaz de armazenar um 'símbolo'. Em cada momento há apenas um quadrado, o r -ésimo, guardando o símbolo $\#(r)$ que está 'na máquina', chamado 'quadrado lido'. O símbolo sobre ele é denominado o 'símbolo lido', que é o único com o qual a máquina está, por assim dizer, 'diretamente ligada'. Contudo, ao alterar uma m -configuração a máquina pode efetivamente relembrar algum dos símbolos que ela 'viu' (escaneou) anteriormente. O comportamento possível da máquina em qualquer momento é determinado pela m -configuração q_n e o símbolo lido $\#(r)$. Este par será chamado de 'configuração', que determina o possível comportamento da máquina. Em algumas das configurações nas quais o quadrado lido está vazio (não carrega nenhum símbolo), a máquina escreve sobre ele um novo símbolo: em outras configurações ela apaga o símbolo lido. A máquina pode também mudar o quadrado que está sendo lido, mas somente deslocando-se um lugar a direita ou a esquerda. Em adição a qualquer uma destas operações, a m -configuração pode ser mudada... Se para cada estágio o movimento da máquina (no sentido acima explicitado) é completamente determinado pela configuração, podemos chamá-la de uma *máquina automática*... Se uma máquina automática imprime dois tipos de símbolos, dos quais o primeiro tipo (chamado figuras) consiste inteiramente de 0 e 1, (os outros sendo chamados símbolos do segundo tipo), então a máquina será chamada uma *máquina computadora*. (Turing, 1936)

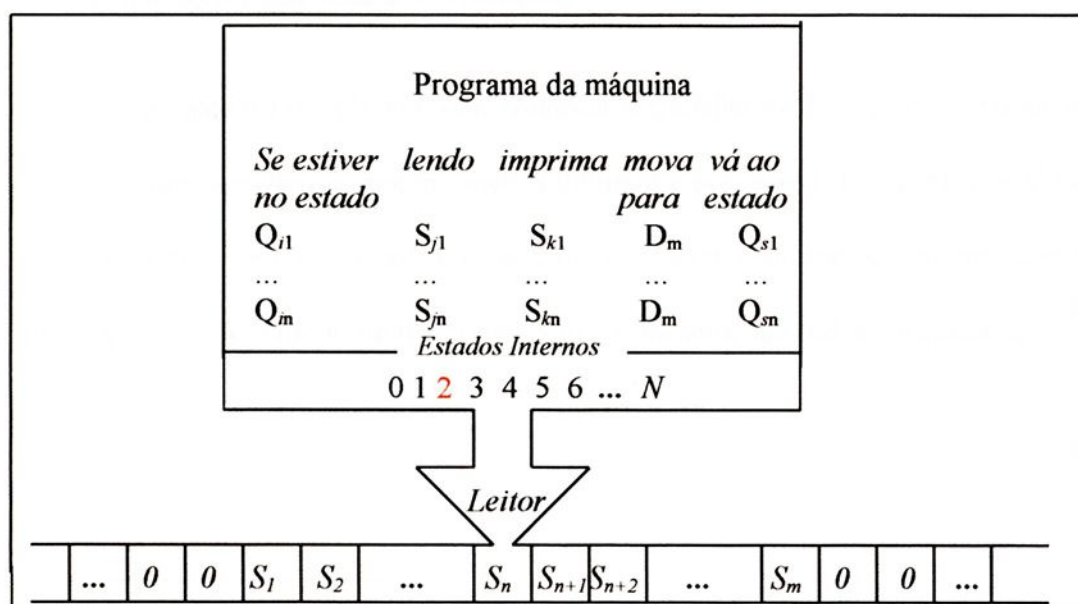


Fig. 2 - Estrutura de uma máquina de Turing

Cada configuração completa (descrição do comportamento da máquina em um dado momento) terá uma das seguintes formas:

$$q_i S_j q_s S_k D \qquad q_i S_j q_s S_k E \qquad q_i S_j q_s S_k N$$

Cada componente da quintupla representa:

(estado atual; símbolo lido; novo estado; novo símbolo; direção)⁷.

Como a máquina computa funções, uma mesma entrada só pode produzir respostas distintas quando apresentadas a máquinas diferentes (a diferença deve estar no programa). Dado um elemento da função à máquina, ela encontrará a resposta adequada para ele, caso haja esta resposta. Para iniciar sua computação é preciso definir o estado inicial da máquina, o conteúdo de sua fita e o quadrado onde deve começar a ler.

Toda máquina deverá ter as seguintes características:

- i) um conjunto finito de símbolos (alfabeto)
- ii) um conjunto finito de estados internos
- iii) uma fita (memória) potencialmente infinita
- iv) um conjunto finito de instruções (programa)
- v) um *agente* que realiza as instruções.

O matemático Post (1943), sem conhecer o trabalho de Turing, criou também em 1936 uma máquina semelhante, porém, menos intuitiva e por isso a de Turing prevaleceu. Como todas as outras máquinas criadas posteriormente fazem exatamente o mesmo que a máquina de Turing, pois seu modo de funcionamento é o mesmo, são todas consideradas como tal.

⁷ A ordem para a quintupla segundo Turing é $q_i S_j S_k D q_s$. Preferimos não segui-la com o intuito de simplificar a exposição da máquina universal. Há autores que definem as configurações completas como sendo quádruplas. Isso faz com que o número de estados e de instruções da máquina sejam maiores.

3.3 Exemplos de Máquinas de Turing

Todos os exemplos a seguir são de máquinas infinitas. Notamos neles que, para computar algumas das funções mais elementares da aritmética, uma máquina não pode ter uma memória limitada. Para produzir resposta aos seus dados de entrada é necessário um número ilimitado de memória (o tamanho da memória é igual ao número de quadrados da fita). Os dois primeiros exemplos são de máquinas que computam a função soma na notação unária (base 1) e binária (base 2). O terceiro exemplo é de uma máquina que computa a função multiplicação. Os outros dois exemplos são partes da máquina de Turing Universal. Todas elas começam no estado zero e o quadrado lido inicialmente é o que possui o gráfico  acima dele.

3.3.1 Somadora unária

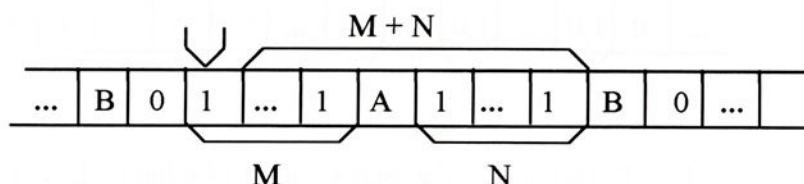
Esta é uma das máquinas de Turing cujo programa é um dos mais simples possível de construção. Ela soma dois números quaisquer M e N no sistema unário (onde $0 = 0$, $1 = 1$, $2 = 11$, $3 = 111$, ...). O programa é constituído de quatro instruções:

00201 01101 11111 1A211 2A201

Seu funcionamento é muito simples. Pode ser assim entendido: a máquina começa sempre no estado 0, lendo o primeiro símbolo à esquerda de M . Se M for 0, a máquina substitui A por 0 e pára (quinta instrução da esquerda para a direita do programa acima). A resposta neste caso será o próprio N . Caso M não seja 0, a máquina substitui o primeiro símbolo de M por 0 e vai à direita (instrução 2) até encontrar a letra A (instrução 3). Feito

isto, substitui esta letra pelo número 1 e pára a computação (instrução 4). A resposta será a seqüência de 1's que restou na fita.

A configuração inicial geral (símbolos iniciais da máquina distribuídos na seqüência dada e símbolo lido - o quadrado com o gráfico  sobre ele) será:



A máquina simplesmente substitui o primeiro 1 de M e o imprime no lugar de A, juntando os dois números M e N. Exatamente o objetivo para a função soma.

Mesmo nesta máquina tão simples é necessária a presença de uma fita potencialmente infinita (ver discussão na seção 3.2)

Os símbolos além de 0 e 1 são os denominados por Turing de *símbolos do segundo tipo*. Eles são postos unicamente por questões de simplicidade na apresentação da máquina. Porém, como mostrou Turing (1936), não são necessários mais do que dois dígitos em qualquer computação (ver seção 3.3)

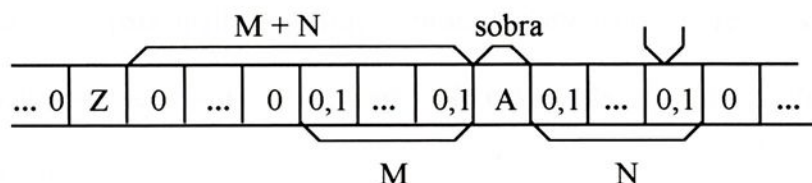
3.3.2 Somadora binária

Esta máquina soma dois números M e N quaisquer na notação binária. Simplesmente por questões de conveniência, como veremos na seção 3.3, tal notação foi a escolhida por Turing. O programa e a configuração inicial geral desta máquina serão os seguintes⁸:

⁸ Na soma com os números representados na notação binária, $0+0=0$ e sobra 0; $1+0=1$ e sobra 0; $0+1=1$ e sobra 0; $1+1=0$ e sobra 1; $1+1+1=1$ e sobra 1)

001Z0	10100	20200	3X3X0	4X4X0	5X5X1	6X6X1	7X7X0	8X800
012Z0	11110	21210	3Y3Y0	4Y4Y0	5Y5Y1	6Y6Y1	7Y7Y0	8Y810
0A8A0	1A3B0	2A4A0	305X1	405Y1	5A5A1	6A6B1	706X1	8Z ψ Z0
0B4B0	1B4B0	2B7B0	315Y1	416X1	5B5A1	6B6B1	716Y1	
					5Z0Z0	6Z0Z0		

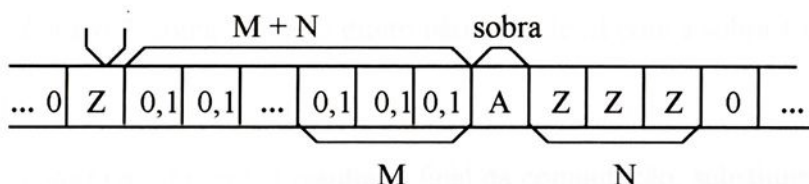
onde ψ é na verdade o estado 13 da máquina seguinte.



onde '0,1' em cada quadrado significa que nele contém um 0 ou 1.

A máquina começa lendo o último dígito não computado de N e o substitui por Z . Feito isto, busca a sobra, que pode ser A ou B (0 ou 1), dependendo dos números da computação anterior. A partir daí, a máquina busca o primeiro dígito ainda não lido de M para fazer a soma com o dígito de N e a sobra. Se der resultado 1, imprime Y . Se der resultado 0, imprime X . Isto é necessário para a máquina identificar os dígitos já somados. No final da computação, cada X é substituído por 0 e cada Y por 1. Ao terminar a computação atual, a máquina imprimirá a sobra deste último cálculo e iniciará um novo, ou seja, lerá o próximo dígito de N ainda não computado. Quando não encontrar mais dígitos em N , a máquina analisará a sobra. Se ela for B (ou seja 1), continuará somando com o próximo dígito de M . Se a sobra for A (0), a máquina imprimirá a resposta toda e parará.

Na configuração inicial acima, o tamanho de memória para ' $M+N$ ' é potencialmente infinito. Fizemos isso para que a máquina fosse realmente capaz de computar quaisquer dois números naturais. Sendo assim, não importa quão pequeno seja M , ou quão grande seja N , a resposta sempre será impressa corretamente. A configuração geral final desta máquina terá a seguinte forma:



Cada estado da máquina tem uma função. Nós apresentamos o programa em colunas exatamente para melhor ilustração desta característica. Porém, as instruções poderiam ter sido apresentadas desordenadamente. No caso deste programa, a função de cada estado será a seguinte:

Estado 0: lê (a máquina) o último dígito de N. Ao encontrá-lo, vai ao estado 1 ou 2. Se não houver mais nenhum dígito de N a ser lido, analisa a sobra da última computação. Se for A, manda imprimir a resposta final (estado 8). Caso contrário, manda encontrar o próximo dígito de M para somar com a sobra (vai ao estado 4);

Estado 1: busca a sobra quando o último dígito lido de N no estado 0 foi 0. Se a sobra foi A, vai ao estado 3, caso contrário, vai ao estado 4;

Estado 2: busca a sobra quando o último dígito lido de N no estado 0 foi 1. Se a sobra foi A, vai ao estado 3, caso contrário, vai ao estado 4;

Estado 3: vai somar o último dígito não lido de M com a sobra 0 e dígito de N se seu valor for 0. Feito isto, vai ao estado 5;

Estado 4: vai somar o último dígito não lido de M com a sobra 0 e o último dígito de N se seu valor for 1;

Estado 5: registra a inexistência de sobra na última computação (escreve A) e encontra o último dígito lido de N;

Estado 6: registra a existência de sobra na última computação (escreve B) e encontra o último dígito lido de N;

Estado 7: soma o último dígito não lido de M com a sobra 1 e dígito de N quando seu valor é 1;

Estado 8: imprime o resultado final da computação, substituindo cada X de M por 0 e cada Y por 1. feito isto, a máquina pára. o resultado final será composto pelos dígitos do dado de entrada M.

Esta máquina, ao terminar a computação, não memoriza os números de entrada. Porém, é perfeitamente possível construir um programa de uma máquina para fazer isso.

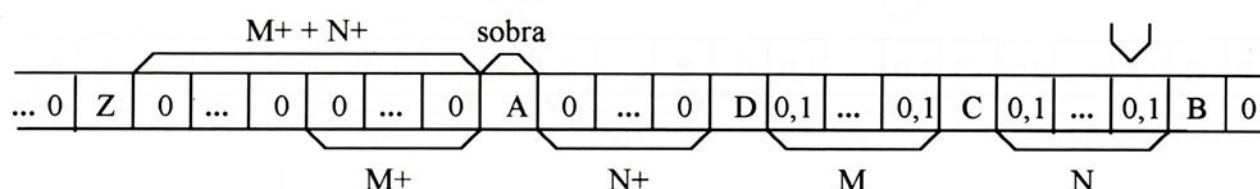
3.3.3 Multiplicadora binária

Esta máquina multiplica dois números quaisquer M e N na notação binária. A multiplicação de qualquer número por 0 é 0 e $1 \times 1 = 1$. O número zero multiplicado por qualquer outro número produz também o resultado zero. O programa e a configuração geral inicial desta máquina serão:

001X0	10100	20200	3X3X0	4X4X0	50500	60600	7X7X0	8X8X0
012Y0	11110	21210	3Y3Y0	4Y4Y0	51510	61610	7Y7Y0	8Y8Y0
0C15C0	1C3C0	2C4C0	305X0	405X0	5D7D0	6D8D0	709X1	809Y1
			315Y0	416Y0			719X1	819Y1
			3D11D0	4D11D0				

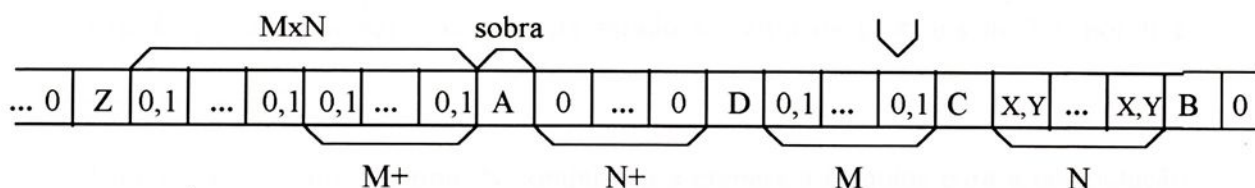
9X9X1	<i>1001001</i>	<i>11X1100</i>	<i>1201201</i>	<i>1301301</i>	<i>13D13D1</i>	<i>1401401</i>
9Y9Y1	<i>1011011</i>	<i>11Y1110</i>	<i>1211211</i>	<i>1311311</i>	<i>13X1301</i>	<i>1411411</i>
9D10D1	<i>10X1X0</i>	<i>11A12A1</i>	<i>12DφD0</i>	<i>13A13A1</i>	<i>13Y13Y1</i>	<i>14X0X0</i>
90901	<i>10Y2Y0</i>	<i>11B12B1</i>		<i>13Z1301</i>	<i>13C14C1</i>	<i>14Y0Y0</i>
91911						
9C10C1						

No programa acima, os números em *itálico* maiores que dez representam um estado interno da máquina. O estado φ é o estado 0 da máquina anterior.



O procedimento da máquina será o seguinte: inicialmente, lê o dígito de N que está sendo multiplicado por todos os dígitos de M. Feito isto, encontra o próximo dígito de M ainda não multiplicado pelo dígito de N em evidência. Ao encontrá-lo, multiplica e o imprime em N+ como X ou Y (0 ou 1). É necessário que ela imprima X e Y para poder saber qual o próximo quadrado ainda não impresso em N+. Se não houver mais dígitos em M para multiplicar pelo N multiplicado atualmente, a máquina substitui cada X e Y de N+ por 0 e 1 (estado 11). Feito isto, começa a soma da última multiplicação impressa em N+ com o resultado parcial de todas as outras, impresso em M+. O estado 12 deixa o leitor da máquina no exato quadrado em que a somadora binária começa lendo. Nesta máquina, a fita à direita de D é exatamente a fita da máquina anterior. Embora não seja necessária a presença de um número infinito de 0's a esquerda de M+, preferimos fazê-lo por questões ilustrativas

Terminada a soma, a somadora envia a máquina ao estado 13. Este tem a função de zerar todos os quadrados de N+ e reimprimir em M seus dígitos originais (0 e 1). Feito isto, vai ao estado 14, que encontra o último dígito de N multiplicado. Após envia a máquina para o próximo dígito a ser multiplicado e começa a multiplicação deste com cada dígito M. Se não houver mais nenhum dígito em N a ser multiplicado, a máquina pára (estado 0, lendo C). O resultado final é a última soma realizada. A configuração final desta máquina será:



O papel de cada estado neste programa será o seguinte:

Estado 0: lê o próximo dígito de N a ser multiplicado. Se não houver, manda parar a computação

Estado 1: encontra a letra C quando o valor de N é 0;

Estado 2: encontra a letra C quando o valor de N é 1;

Estado 3: encontra o próximo dígito de M a ser multiplicado por N quando seu valor é 0. Se não houver, manda preparar a máquina para começar a soma da multiplicação de cada dígito de M pelo dígito atual de N (estados 11 e 12);

Estado 4: encontra o próximo dígito de M a ser multiplicado por N quando seu valor é 1. Se não houver, manda preparar a máquina para começar a soma da multiplicação de cada dígito de M pelo dígito atual de N (estados 11 e 12);

Estado 5: encontra D e vai ao estado 7;

Estado 6: encontra D e vai ao estado 8;

Estado 7: escreve o resultado da última multiplicação em N+ quando for 0;

Estado 8: escreve o resultado da última multiplicação em N+ quando for 1;

Estado 9: encontra a letra C;

Estado 10: encontra o dígito de N que está sendo multiplicado para computá-lo com o próximo dígito de M ainda não multiplicado por ele;

Estado 11: troca X por 0 e Y por 1 em N+, ou seja, prepara a máquina para a soma;

Estado 12: leva o leitor da máquina ao quadrado inicial para a soma;

Estado 13: ao terminar a soma, este estado substitui os símbolos de N^+ por 0 e reimprime 0 e 1 em M;

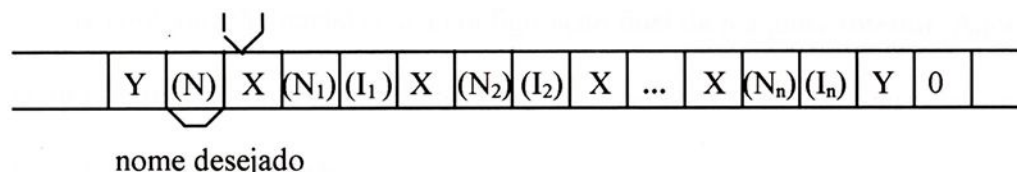
Estado 14: encontra o último N computado e prepara a máquina para a computação do próximo (vai ao estado 0).

3.3.4 máquina de memória endereçada

Esta máquina tem a função de encontrar um nome idêntico a um outro desejado. Cada nome N_i está associado a um item I_i . Para simplificação do programa, todos os nomes têm o mesmo tamanho. Se assim não fosse, seria necessário separá-los por algum sinal, exigindo assim mais símbolos e estados e consequentemente mais instruções. Seu programa será:

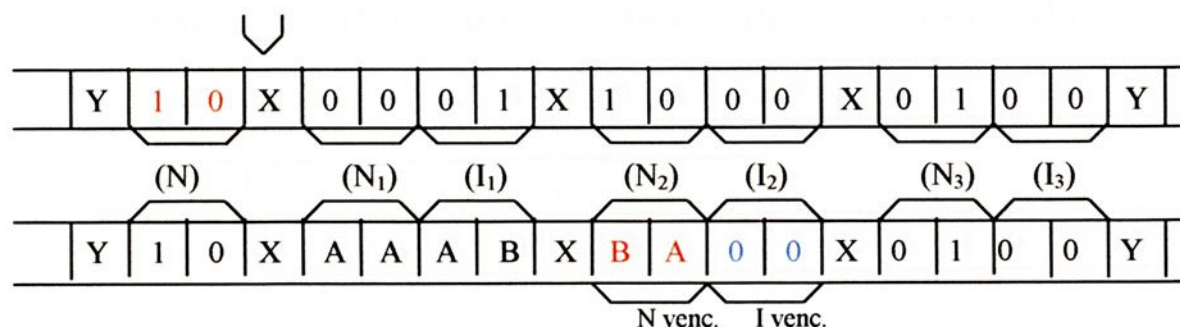
000A0; 010B0; 0Y1Y1; 0X0X0; 0A0A0; 0B0B0; 1A201; 1B311; 1X6X1; 11111; 10101;
215B1; 204A0; 2X2X1; 2A2A1; 2B2B1; 314B0; 305A1; 3A3A1; 3B3B1; 3X3X1; 4Y1Y1;
40400; 41410; 4X4X0; 4A4A0; 4B4B0; 5X0X0; 50501; 51511; 5Y ξ Y0
onde ξ é o estado 0 da máquina seguinte.

Sua configuração inicial geral será (N) e (I) são seqüências de dígitos binários 0,1)



No estado 0, esta máquina substitui os elementos de (N) por A e B. Ao terminar (estado 0 símbolo lido Y) vai ao estado 1 e memoriza o primeiro elemento do N_i ainda não analisado. Se este elemento for igual ao do N, a máquina busca o próximo elemento de N. Se for diferente, elimina o N_i e I_i e recomeça o processo de identificação. O nome vencedor será o associado ao primeiro I não eliminado.

Um exemplo de uma computação com suas configurações inicial e final pode ser:



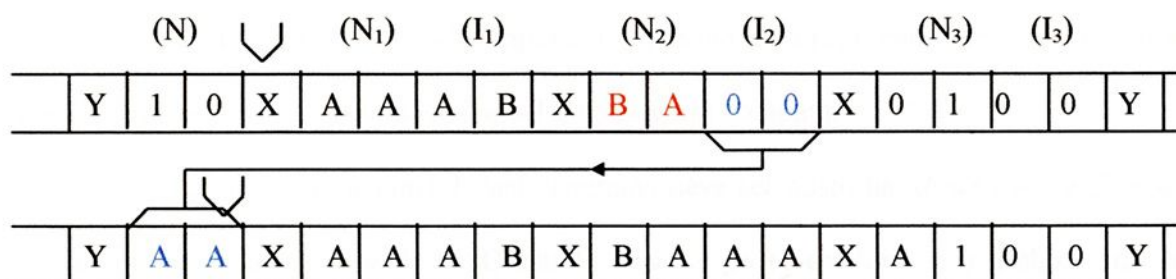
3.3.5 Máquina copiadora

Uma vez identificado o nome do item procurado pela máquina anterior, pode-se querer transferi-lo para o lugar do nome desejado, transformando-o em (N). É isto que faz esta máquina, cujo programa é constituído das seguintes instruções:

001A0; 012B0; 0X0X1; 0A0A1; 0B0B1; 1Y3Y1 1A1A0; 1B1B0; 1X1X0; 10100; 11110; 2Y4Y1; 20201; 2X2X1; 2A2A1; 2B2B1; 21211; 315A1; 305A1; 3A3A1; 3B3B1; 3XζX0; 4XτX0; 405B1; 415B1; 4A4A1; 4B4B1; 5X0X1; 50501; 51511; 5A5A1; 5B5B1.
onde ζ é o estado 0 de 3.4.1 e τ o estado 1 de 3.4.1.

Sua configuração inicial será a configuração final da máquina anterior. Aqui o item e o nome devem ter o mesmo tamanho. Se assim não fosse, poderia faltar ou sobrar espaço para escrever o item vencedor em N.

No estado 0, a máquina busca o primeiro dígito do I vencedor. Ao encontrá-lo, o memoriza (vai ao estado 2, se for 0 ou estado 1, se for 1) e volta até o Y da esquerda. Feito isto, volta percorrendo N até encontrar seu primeiro dígito ainda não substituído por um dígito do I. Se houver, o substitui pelo dígito memorizado pela máquina e repete o processo. Se não houver mais nenhum dígito, a máquina pára (vai ao programa 3.4.1), memorizando o elemento que não foi posto em N. Eis um exemplo, com as configurações inicial e final de uma computação, cujo item vencedor é I_2 :

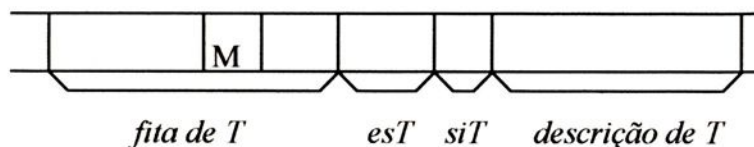


Neste caso, I_2 foi transferido para a posição de nome do item procurado na configuração final da computação)

3. 4 Máquina de Turing Universal

Cada máquina apresentada até o momento computa uma função particular. Turing (1936) porém, criou uma máquina universal capaz de realizar qualquer função computável. Dado um *input* qualquer e a função que o computa, esta máquina é capaz de calculá-lo. A máquina apresentada a seguir segue a construção feita por Minsky (1967).

Para realizar uma máquina T , a máquina universal U precisa ter a descrição de T (seu programa) registrada em sua memória para poder recuperá-las a todo momento que precisar. Deve ter também memorizados o estado atual, o símbolo lido e a fita de T , ou seja, sua configuração inicial. A fita de U deverá ser composta então de quatro partes:



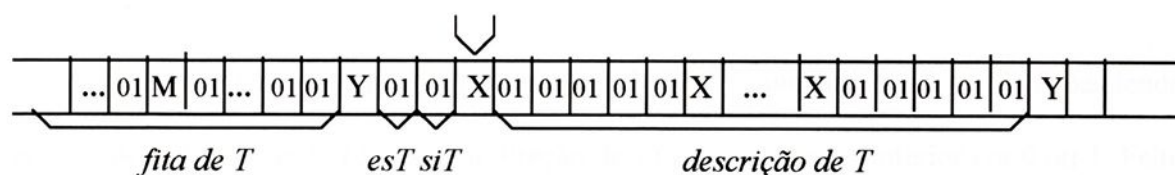
Neste caso, 'fita de T ' é como dito acima. A letra M representa o quadrado lido de T ; 'es T ' é o estado atual de T ; 'si T ' é seu símbolo lido e 'descrição de T ' é um conjunto finito de

quíntuplas separadas por 'X'. Não importa o tamanho do programa, nem seu número de símbolos ou estados, a máquina universal o conseguirá computar.

Para simular uma máquina T , seu programa deve ser posto na '*descrição de T* ' e sua configuração inicial na '*fita de T* '. O estado inicial é posto em '*es T* ' e o símbolo lido em '*si T* '. Aí então U está pronta para começar a primeira instrução do programa de T . A



máquina começa realizando o programa da máquina 3.3.4 com a seguinte configuração inicial geral



Se tirarmos a 'fita de T ', a configuração inicial geral de U , é idêntica à configuração inicial geral de 3.3.4. Dado o estado inicial de T e seu símbolo lido (N em 3.3.4), a máquina encontrará o nome (configurações do programa de T) idêntico a este par.

Uma vez identificado o nome, a máquina passa a realizar o programa da máquina 3.3.5 e sua fita segue como terminou na máquina anterior. Aqui U transfere o novo estado de T e o novo símbolo da instrução (posteriormente impresso em M) que formam o item vencedor da máquina 3.3.4 para esT e siT (o (N) em 3.3.4). O nome procurado e o item vencedor parecem não ter o mesmo tamanho, como exigido em 3.3.5. De fato, o item vencedor é composto pelo novo estado e novo símbolo da instrução. A direção será memorizada pela máquina 3.3.5 e irá ao terceiro passo, que é a realização do seguinte conjunto de instruções:

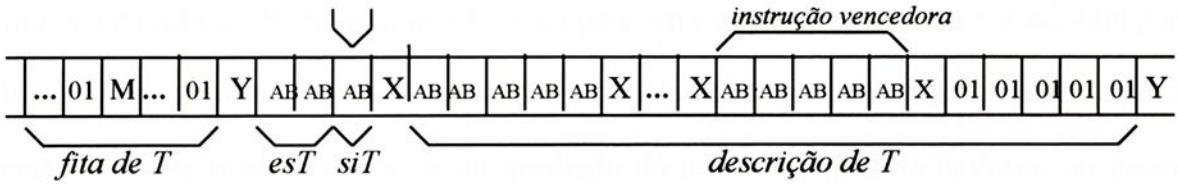
Máquina 3.4.1

01011; 00001; 0M2A1; 0A0A1; 0B0B1; 0Y0Y1; 11111; 10101; 1M2B1; 1A1A1; 1B1B1; 1Y1Y1; 21211; 20201; 2Y2Y1; 2A201; 2B121; 2X3X1; 3A3A1; 3B3B1; 3X3X1; 30400; 31410; 40 $\cup$ S0; 41 ∂ S0; 4A400; 4B410; 4X4X0⁹

No programa acima, $\cup$ é estado 0 de 3.4.2; ∂ seu estado 1.

A configuração inicial geral de U neste passo será

⁹ A função de cada estado será: 0: encontra M , o substitui por A e o leva a esquerda; 1: encontra M , o substitui por B e o leva a direita; 2: transforma os elementos do condicional da máquina em 0 e 1; 3: encontra 0 ou 1 para função 4; 4: transforma cada instrução em 0 e 1 e substitui o novo símbolo por S , levando a computação para o programa 3.4.2.

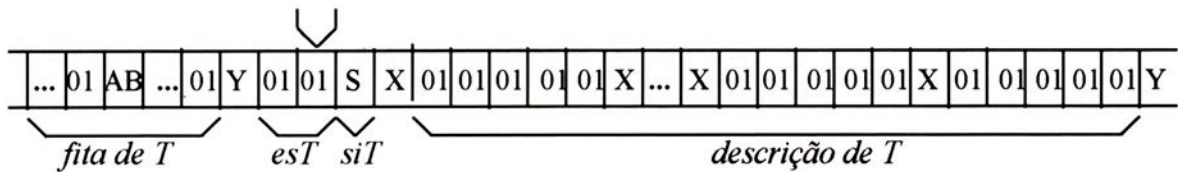


Ao realizar este programa, a máquina vai até M e o substitui por A ou B , dependendo se o número memorizado (que será a direção de T) pelo programa anterior era 0 ou 1. Feito isto, volta e transforma o símbolo no quadrado sobre esT e siT (que será a próxima configuração). Vai até o final da instrução vencedora e volta recolocando as instruções de T em 0,1 até encontrar 0 ou 1 (que será o símbolo a ser posto em A ou B na *fita de T* , ou seja, o novo símbolo do quadrado lido por T nesta instrução). Ao encontrar o número, substitui por S e o memoriza, ou seja, passa ao último conjunto de instruções:

Máquina 3.4.2

00500; 01510; 0Y5Y0; 0B201; 0A300; 10100; 11110; 1Y1Y0; 1B211; 1A310; 204M1; 215M1; 304M1; 315M1; 50501; 51511; 5Y5Y1; 5S $\perp$ B0; 4041; 41411; 4Y4Y1; 4S $\perp$ A0¹⁰
onde $\perp$ é o estado 0 de 3.3.4

A configuração de U ao iniciar esta última parte de seu programa é exatamente a configuração final da execução da parte anterior:

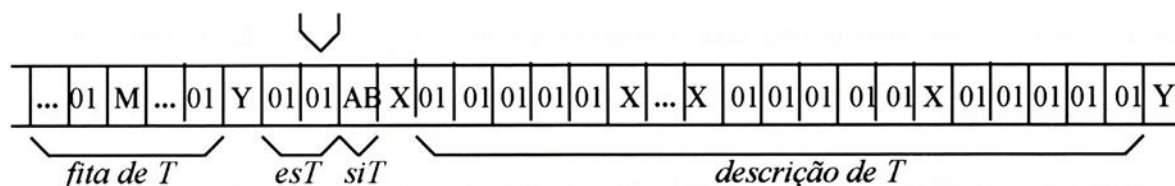


Se o número memorizado pela máquina anteriormente foi 0, ela começa este programa no estado 0 e se foi 1, começa no estado 1. Ambos buscam o A ou B na fita da máquina

¹⁰ Função de cada estado: 0: encontra direção e a substitui por 0; 1: encontra direção e a substitui por 1; 2: move a máquina para a direita e substitui o símbolo lido por M ; 3: leva a máquina a esquerda e substitui o símbolo lido por M ; 4 e 5: substitui S pelo símbolo atual e leva a computação para o programa da máquina 3.3.4

(que é a direção de T). Ao encontrá-lo, a máquina vai a esquerda, se tiver A e o substitui por 0, ou direita, se tiver B substituindo-o por 1 (novo símbolo da instrução atual de T). Ai então encontra o símbolo que há no quadrado do qual a máquina se deslocou no passo anterior, o substitui por M e o imprime no lugar de S (siT) como A ou B e vai à máquina 3.3.4. Sendo assim, U realizou perfeitamente uma instrução de T e está preparada para realizar a próxima.

A máquina pára quando 3.3.4 não encontrar nenhum nome idêntico ao procurado, ou seja, quando o programa de T produzir uma nova configuração (par formado pelo estado e símbolo atual) que não seja nenhuma das pertencentes às suas instruções. A configuração geral final de U será idêntica à configuração geral inicial de 3.3.4, exceto pelo fato de U já ter transformado o último dígito de (N) em 0 ou 1 e fazer com que 3.3.4 comece no penúltimo dígito de (N) e não em X. Ela será então a seguinte::



3 Discussão sobre o conceito de Máquina

Ao falar em Máquinas de Turing, muitos cientistas, principalmente os matemáticos, a entendem como sendo simplesmente um conjunto de instruções. Eliminam dela a memória (concreta) e o agente. Procuraremos mostrar neste tópico que não faz muito sentido definir máquina desse modo. Discutiremos ainda algumas questões referentes ao alfabeto e estados internos da máquina.

3.1 O agente

Geralmente a máquina de Turing é definida como sendo simplesmente um programa computacional. Com isso, deixa-se de lado uma característica, do nosso ponto de vista, fundamental, qual seja, o agente. Como afirma Cutland (1992, p. 124), “não é comum construir máquinas fixas, exceto por questões ilustrativas”. Podemos pensar desse modo quando estamos num plano formal. Porém, quando saímos desse domínio, não faz sentido pensar na máquina como sendo simplesmente um programa. Isso por, no mínimo, dois motivos. Um deles diz respeito ao fato de dizermos que a máquina computa uma função. Sem um agente para realizar suas instruções, o programa nada computa. O segundo motivo é que a máquina precisa de memória para guardar dados. O ato de guardar dados aqui é entendido como um ‘local’ onde eles ficam armazenados. Nesse sentido, sem um agente não há como memorizar estes dados. Como poderíamos fazer isso apenas com um conjunto de instruções?

Na definição de máquina proposta por Turing (1936), podemos encontrar a necessidade de algo a mais do que o programa quando afirma, por exemplo,

... que a máquina escreve, apaga, muda de lugar, que o comportamento da máquina é descrito por uma tabela onde R significa que a máquina move a direita, que a máquina começa na *m*-configuração *b* com uma fita em branco (Turing, 1936, pg. 119)¹¹.

Diríamos então que a máquina é um par composto por um agente e um programa. Agente sem programa não faz nada, pois precisa das instruções para realizar um cálculo. Por outro lado, programa sem agente, não tem serventia, pois para realizar qualquer função precisa do agente para realizar suas instruções.

Se o agente parece ser necessário na caracterização de máquina, porque os cientistas o deixam de lado, quando a caracterizam formalmente? Uma das possíveis respostas pode ser as complicações originadas por esta noção. Podem surgir questões do tipo quem deve ser o agente? Quais habilidades deve ter? O que deve fazer? Pode ser criativo ou impulsivo?

Como dissemos na caracterização informal de algoritmo, não há para ele um limite no tamanho de dados de entrada nem no número de instruções. Desse modo, o agente pode precisar ser muito sofisticado para realizar as operações contidas no programa. Questionamo-nos então sobre o grau de habilidade que deve ser apresentado pelo agente.

Na realidade, para computar qualquer função algorítmica, basta que o agente realize instrução por instrução do programa, uma de cada vez. Para realizar qualquer instrução, deve possuir apenas as simples habilidades de mover-se ou permanecer no lugar; ler, escrever ou apagar símbolos e mudar de estado. Sendo assim, o comportamento inteligente da máquina, para a IA, resume-se ao fato dela poder realizar estas instruções e que, ao final desta realização tenha resolvido algum problema.

Na máquina de Turing, o agente possivelmente não entende o que está fazendo. Quando lê um símbolo, não precisa relacioná-lo com mais nenhum outro. Basta saber em que estado está e realizar o que a configuração determina. Com isso, simplesmente manipula símbolos, sem atribuir a eles um significado.

O agente não é um sujeito com vontade, não é livre e nem tem decisão de ação. Seu comportamento é determinístico, ou seja, não escolhe o que faz nem age de acordo com a sorte ou acaso. Além disso, age discretamente, ou seja, realiza uma ação por vez em cada momento e ao terminá-la, passa para outra, sem nenhum passo intermediário entre elas. O

¹¹ É claro que quando fala de fita, Turing o faz metaforicamente para falar de memória e quando fala de



agente pode ser qualquer sistema capaz de realizar corretamente as instruções de um programa e não deve ser confundido com o que chamamos anteriormente de operador da máquina. Este está fora da máquina enquanto que aquele faz parte dela.

3.2 A memória

A maioria das máquinas interessantes precisam de uma memória ilimitada para computar suas funções. Isso significa que é necessária a existência de uma fita infinita na máquina. No caso dos exemplos 3.3.1 e 3.3.2, a impossibilidade de limite encontra-se nos dados de entrada. Uma vez que estamos no domínio dos números naturais, é sempre possível aumentar o seu tamanho. Desse modo, se limitássemos o número de quadrados da fita em um tamanho qualquer, seria perfeitamente possível encontrar um dado de entrada cuja extensão excedesse qualquer limite estipulado.

Uma das saídas adotadas para eliminar ou resolver este problema de máquinas como as acima citadas, seria a possibilidade da máquina apagar o dígito de entrada após tê-lo lido. Sendo assim, a máquina lê os pares de elementos de M e N , os computa e apaga em seguida. No lugar deles seria posto o próximo par de elementos. Isso faria com que a máquina não necessitasse de uma memória infinita. Porém, exigiria um operador externo para enviar os símbolos a ela par a par e esperar pelo próximo momento de poder acessar os próximos dígitos. Isso seria desinteressante para os propósitos da máquina. Pretende-se que ela compute sozinha as funções, sem a necessidade de um auxiliar externo.

mover-se, ler número, etc, está falando da capacidade do agente de identificar e manipular símbolos.



Ao limitar a memória para os dados de entrada nestas máquinas, estamos também limitando sua memória de saída, onde são impressos os resultados da máquina para cada computação. Isso porque a resposta e o dado de entrada M são os mesmos no caso da primeira. No caso da segunda, os dígitos finais da resposta são dados exatamente nos quadrados dos dígitos de M . Em ambos os casos, a máquina não seria capaz de memorizar a resposta e apresentá-la de uma só vez. Novamente, ela precisaria de um auxiliar externo, tornando-a como anteriormente, desinteressante.

No caso da máquina 3.3.3, nem limitando a memória para os dados de entrada poderíamos ter uma memória finita. Isso porque as respostas produzidas pela máquina são parciais. Para cada computação, a máquina precisa guardar seu resultado. Esse será somado com o resultado da nova multiplicação. A resposta total será dada quando todas as multiplicações tiverem sido feitas.

Notamos, assim, que as funções que uma máquina de memória finita pode realizar são geralmente aquelas cujos objetos do domínio são finitos. Assim, uma máquina finita pode calcular a função soma, por exemplo, quando os números podem ser apenas de um tamanho limitado. Isso é muito desinteressante

3.3 O alfabeto e os estados internos

Há uma exigência que o alfabeto da máquina e seu número de estados devem ser finitos. Por quê? O alfabeto e o número de estados da máquina são os símbolos e os estados que aparecem no programa. Desse modo, se qualquer um deles for infinito, o número de instruções do programa também o será, pois cada instrução pode conter no máximo dois

símbolos e estados distintos. Porém, o número de instruções do programa deve ser finito, caso contrário, a máquina poderia não terminar sua computação.

Porque o alfabeto e o número de estados da máquina são os símbolos e os estados que aparecem no programa? Não poderia pertencer apenas ao agente? Se assim fosse, poderíamos ter o problema de símbolos indesejados aparecerem na fita ou do agente não possuir os símbolos ou estados apropriados para a realização de uma função, mesmo que o programa estivesse correto. Assim, a máquina poderia produzir um resultado indesejado ou não produzir nenhum quando devesse apresentar. Turing (1936) afirma que se o número de símbolos ou estados fossem infinitos, haveria símbolos e estados que diferiram tão pouco um do outro que seriam indistinguíveis.

Turing (1936) afirma que não são necessários mais do que dois símbolos para uma máquina computar qualquer função. Isso pode ser feito criando uma tradução para cada símbolo utilizado na máquina para uma notação digital. Na máquina 3.3.2, por exemplo, utilizamos os símbolos 0, 1 Z, A e B. A tradução destes símbolos para um sistema binário poderia dar-se do seguinte modo:

$$0 = 0; 1 = 10; Z = 110; A = 1110; B = 11110$$

Desse modo, substituindo cada símbolo por seu correspondente na notação binária, teríamos a fita da máquina com apenas dois símbolos.

A utilização de apenas dois símbolos torna a máquina um pouco mais complexa. Aumenta o tamanho da fita, o tamanho do número de instruções e o de estados internos. É fácil ver que o tamanho da fita é maior quando utilizamos apenas dois símbolos nela. Enquanto que na máquina 3.3.2 utilizamos apenas um quadrado para registrar a letra B, precisamos de cinco quadrados na notação digital.

O número de estados internos aumenta porque o processo para encontrar ou reconhecer um símbolo na notação binária é mais complexo do que na outra notação. Por exemplo, suponhamos que temos o seguinte programa na notação com símbolos adicionais (os mesmos da máquina 3.3.2) para encontrar a letra B:

0B8B0 0A0A0 01010 00000 0Z0Z0

O comportamento da máquina cujo programa é o acima é simplesmente ir à esquerda até encontrar um B. Na notação binária, seria preciso, para uma máquina ter o mesmo comportamento, que ela lesse um 0 e quatro 1's seguidos. Precisaria contar cada 1 que ela lesse a fim de poder saber quando está realmente lendo o símbolo 'B'. O modo de poder armazenar (memorizar) a quantidade de 1's lidos é através da mudança de estados. Assim, por exemplo, na notação binária, teríamos o seguinte programa:

00100 10000 20000 30000 40000 11210 21310 31410 41810

Observamos neste programa a necessidade de cinco estados para fazer o mesmo que o programa anterior fazia com apenas três.

O aumento do número de instruções está evidente neste exemplo. Para o reconhecimento de cada símbolo é necessário, na notação binária, o aumento de estados. Porém, cada novo estado é uma nova instrução. Desse modo, o número de instruções aumenta no mínimo tanto quanto o número adicional de estados.

Por que adotar uma notação binária, se esta produz novas complicações? A justificativa é simples. Quanto menos símbolos a máquina tiver, menos dificuldade tem para diferenciá-los. Assim, não há o problema de realizar uma computação erroneamente causada pela confusão da máquina no reconhecimento de símbolos.

Para que servem os estados internos na máquina de Turing? Basicamente, sua função é determinar o comportamento da máquina. Os estados internos são necessários para o agente poder saber qual o próximo passo a dar. Algumas vezes a presença dos estados também serve como “contadores”, como no caso do programa acima apresentado.

Os estados internos da máquina poderiam ser totalmente eliminados de sua definição. Isso seria possível, se a construíssemos de outro modo. Um destes modos é a Máquina de Registro Ilimitado (URM) apresentada por Shepherson & Sturgis (1963). Ao invés de definir o programa como um conjunto de instruções, eles o definem como uma sequência. Desse modo, a máquina realiza sempre a próxima instrução ou a que a instrução que está sendo realizada indicar. O processo de contagem que os estados internos exercem na máquina de Turing, também são substituídos por outros mecanismos na URM. A URM faz exatamente o que a máquina de Turing e do mesmo modo, qual seja, mecânico.

Como visto, Turing utiliza a notação binária para a representação dos elementos das funções. Por que prefere utilizá-la e não fazer uso, por exemplo, de uma notação unária ou decimal? O motivo é simples: a notação decimal exige muito mais instruções para os programas. Na notação binária, a máquina poder ler 0 ou 1. Sendo assim, precisamos apenas de duas instruções para cada estado. Desse modo, podemos determinar corretamente o funcionamento da máquina. No caso da notação decimal, precisaríamos de dez instruções para cada estado. Em segundo lugar, esta notação exige um conjunto maior de símbolos no alfabeto. Como já dissemos, Turing quer uma economia na quantidade de símbolos para a máquina. Isso torna seu reconhecimento mais preciso.

No caso da notação unária, poderia acontecer uma simplificação no número de instruções, se seguirmos o mesmo argumento apresentado acima. O problema é quando

temos um dado de entrada muito elevado para computar. Mesmo que existissem menos instruções no programa, o número de vezes que elas deveriam ser executadas seria muito grande. Tome o exemplo da máquina 3.3.1. Se M fosse 1 milhão, precisaríamos aplicar a segunda instrução 1 milhão de vezes. Na notação binária, os números tornam-se muito menos extensos. Com isso, sua computação é muito mais rápida.

Como visto, cada estado tem uma determinada função. Porém, esta função é simplesmente dada para o entendimento da máquina por parte do usuário ou do ambiente. O agente, porém, não tem “consciência” destas funções de cada estado. Ele apenas reconhece o símbolo atual e com o estado atual opera de acordo com a regra. Não precisa entender o que está fazendo, basta ser capaz de diferenciar um símbolo do outro e em qual estado está no momento. Esta é uma das críticas feitas à tese de que as máquinas pensam e são inteligentes. O sujeito inteligente não parece agir desse modo, como veremos no último capítulo.

3.4 Algumas limitações da máquina de Turing

A máquina de Turing poderia realizar qualquer comportamento ou resolver qualquer problema? Uma vez que ela é capaz de realizar apenas o que pode ser precisamente descrito, a primeira resposta a esta pergunta é negativa. Se pensarmos do ponto de vista da computabilidade, ela pode resolver qualquer função que possa ser precisamente descrita? Mesmo neste campo, a resposta é não. Foram provadas várias funções que parecem ser precisamente descritas e que a máquina não consegue resolver. Algumas delas são citadas a seguir:



Problema da parada: Existe uma máquina que, para um programa de uma máquina qualquer e um dado de entrada a ela apresentado, pode decidir se sua computação é finita? Até hoje, não existe nenhuma máquina capaz de computar a função característica desse problema.

A insolubilidade do problema da parada foi utilizado por Turing para mostrar a indecidibilidade do problema de Hilbert, como comentamos na seção 2 deste capítulo. Turing (1936) supôs a existência de um programa para computar a função característica deste problema. A partir daí, chegou a uma contradição. Como consequência, pelo famoso método de lógico de *redução ao absurdo*, mostrou que a suposição deve ser falsa.

Se o problema da parada fosse decidível, muitas conjecturas como a de Goldbach ¹² e o próprio problema de Hilbert poderia ser resolvidos. O problema é que não sabemos com certeza se a máquina pára com relação a qualquer número. Não temos garantia da resposta da máquina. Pode ser que ela nunca pare. Com isso, nunca saberemos do final da computação. Consequentemente, não temos certeza da decidibilidade do problema.

Problema da identificação: Haveria uma máquina que, para uma função qualquer, pudesse encontrar todos os programas pelos quais ela pode ser computada? Isso também não é possível e traz complicações para a computação, pois não é possível encontrar algoritmicamente, dentre uma lista de programas, qual o que computa uma determinada função.

Problema da igualdade: Existe um método mecânico para, dados dois programas quaisquer, decidir se suas regras propiciam a computação de uma mesma função? Se fosse

¹² Segundo a conjectura de Goldbach, todo número par maior do que dois é a soma de dois números primos.



possível responder positivamente a este problema, poderíamos decidir *mecanicamente* qual deles seria o mais simples para computar um mesmo número.

Problema lógico-matemático: Este é o problema pelo qual o ideal logicista desmoronou. Não existe um procedimento efetivo que decida se uma fórmula qualquer do cálculo de predicados é teorema deste cálculo. A insolubilidade deste problema, como já dissemos, é consequência do problema da parada. Esse predicado é parcialmente decidível. Se a fórmula é teorema, a máquina encontra sua demonstração. Caso contrário, nunca produzirá uma resposta.

Em resumo, a máquina de Turing pode ser pensada como uma caixa quase preta com uma capacidade de memória potencialmente infinita. A máquina recebe dados de entrada, os processa algoritmicamente e produz uma resposta. Ao criar a máquina abstratamente, Turing (1936) afirmou que o comportamento dela poderia ser comparado ao processo de cálculo realizado pelo homem ao computar uma função. Como dissemos, uma grande parcela dos cientistas da IA acreditam que os processos de conhecimento humano podem ser explicados analisando o comportamento da máquina de Turing. Para eles, a inteligência e demais características da mente podem ser simulados e explicados através desta máquina. Nesse caso, o comportamento dá-se através do seguimento de instruções de programas computacionais. O comportamento inteligente será aquele cujo resultado final da realização de um conjunto de instruções resolva algum problema relevante, como do tipo lógico-matemático, por exemplo.

Para os conexionistas, o comportamento inteligente não pode ser explicado simplesmente através do seguimento de regras e manipulação de símbolos. Ao contrário, é preciso levar em consideração as partes básicas da mente e suas relações. Para os cientistas

da corrente em questão, a mente é, de fato, o cérebro e suas partes básicas são os neurônios. Desse modo, uma vez que a máquina de Turing não possui as características físicas e relacionais do cérebro, não é um bom modelo da mente. Os conexionistas criaram então modelos de inspiração cerebral, denominados *redes neurais artificiais*.



Cap. 3

Modelos conexionistas da mente

1 Apresentação

Segundo conexionistas como Rosenblatt (1962), Kohonen (1987), McClelland & Rumelhart (1988) e Caudill & Butler (1992), a máquina de Turing não pode ser considerada um bom modelo da mente. Nela não são levados em consideração algumas características fundamentais da mente como as físicas, por exemplo. Um bom modelo da mente deve ser de inspiração neuronal, constituído de “neurônios” interconectados entre si, afirmam eles. É a partir destas conexões que emergirá o comportamento inteligente da rede. A rede não pode ter seu funcionamento simplesmente algorítmico.

De acordo com os conexionistas, suas redes não funcionam como as máquinas de Turing, cuja essência é seguir regras e manipular símbolos. Ao contrário, são formadas por nódulos que, ao se relacionarem, fazem com que a rede aprenda e comporte-se inteligentemente. Porém, tal comportamento não é puramente algorítmico (Caudill & Butler, 1992).



Uma das grandes inovações no projeto conexionista foi introduzir na pesquisa sobre os processos cognitivos a noção de auto-organização. Não são necessários mais um conjunto de regras fixas preestabelecidas tampouco um centro controlador para a máquina realizar um determinado comportamento. Ao contrário, suas partes básicas se interagem de tal modo que possibilitam a organização do sistema. É a partir desta dinâmica que a rede aprende, resolve problemas, reconhece padrões etc. Embora saibamos que a teoria da auto-organização seja fundamental ao se pesquisar o Conexionismo e seus modelos, infelizmente, por questões de espaço e tempo, não nos aprofundaremos neste tema.

Uma das primeiras *redes neurais artificiais*, a serem construídas foi a de McCulloch & Pitts (1943), apresentada a seguir. Após sua exposição, apresentamos o *perceptron*, uma rede mais sofisticada, com uma regra de aprendizagem, e a *rede de Kohonen*, capaz de se auto-organizar.

2 A rede de McCulloch-Pitts

2.1 Introdução

Uma das hipóteses fundamentais de McCulloch e Pitts no estudo sobre a mente era a da identidade mente-cérebro e, por isso, sua rede deveria ser construída com alguma inspiração cerebral. Embora a rede tivesse algumas semelhanças com o cérebro, estes pesquisadores sabiam que aquela era uma aproximação muito limitada deste. Mesmo assim,



acreditavam que esse modelo poderia simular eventos mentais e comportamentos humanos inteligentes, como o cálculo de funções matemáticas.

2.2 O cérebro segundo McCulloch & Pitts.

De acordo com McCulloch & Pitts (1943), o sistema nervoso é uma rede de neurônios compostos por um soma e um axônio. O soma do neurônio é seu corpo celular e o axônio é um cilindro-eixo. As conexões entre os neurônios, denominadas sinapses, acontecem sempre entre o axônio de um e o soma de outro, como ilustrado na fig. 3 abaixo. Cada neurônio pode receber inúmeros *inputs* de outros neurônios ou do ambiente externo simultaneamente. Porém, pode produzir apenas uma resposta, possivelmente transmitida a inúmeros outros neurônios ou para o ambiente externo da rede. A resposta do neurônio é enviada pelo axônio e pode ser “repartida” (não dividida) através das terminações axonais.

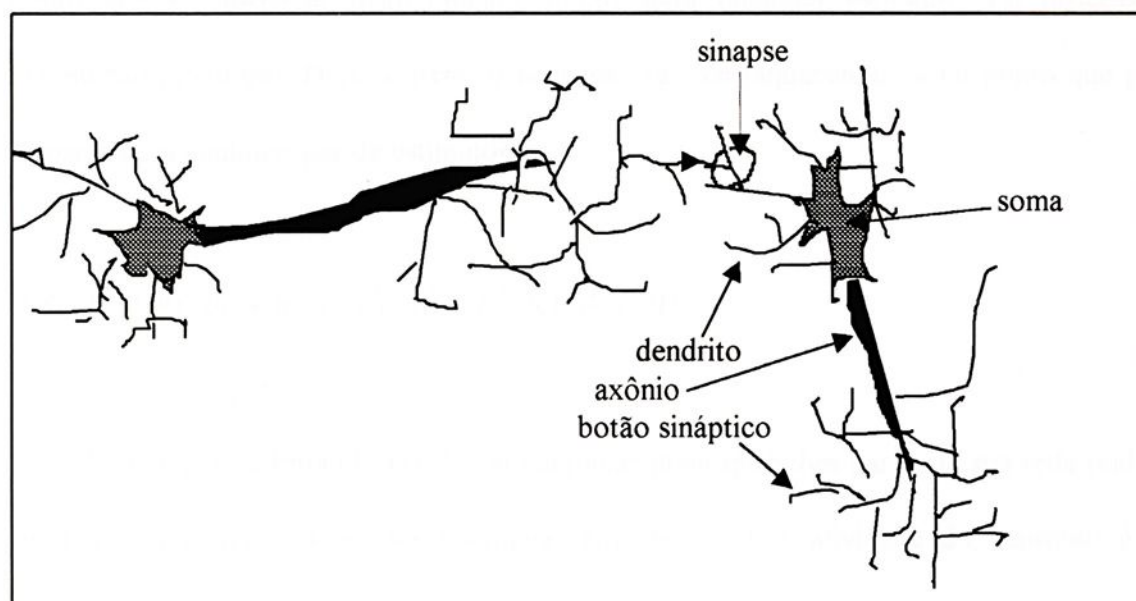


Fig 3 – Descrição simplificada de neurônios biológicos

As sinapses inibitórias, aquelas que diminuem a possibilidade do neurônio disparar, podem inibir completamente o impulso do neurônio (sinapse inibitória absoluta) ou diminuir a possibilidade de seu disparo, aumentando o seu limiar (sinapse inibitória relativa). Por outro lado, as sinapses excitatórias auxiliam no disparo do neurônio. As conexões entre os neurônios do cérebro são feitas através de transmissão de substâncias químicas e elétricas. O disparo do neurônios depende da relação entre estas substâncias com algumas características próprias do neurônio, como seu limiar

O neurônio tem sempre algum limiar, cujo estímulo precisa ultrapassar para dar início a um impulso. Para McCulloch & Pitts (1943), o limiar que, na rede artificial é representado por valores numéricos naturais, precisa ser maior do que zero, pois um estímulo sozinho não excita o neurônio. Efetuado o estímulo, o impulso é propagado para todas as partes do neurônio.

Durante a primeira parte do impulso nervoso, o limiar do neurônio aumenta a tal ponto que ele torna-se absolutamente imune para qualquer estímulo. Tal processo é denominado *extinção*. Depois disso, o neurônio vai “enfraquecendo” a tal ponto que pode disparar com qualquer par de estímulos.

2.3 As pressuposições de McCulloch & Pitts

McCulloch & Pitts (1943) fizeram algumas pressuposições para que sua rede pudesse ser construída, que podem ser resumidas em cinco: 1) A atividade do neurônio é um processo de tudo-ou-nada; 2) um certo número fixo de sinapses precisa ser estimulado dentro do período de adição latente para excitar um neurônio em qualquer tempo, e este



número independe de atividade prévia e posição do neurônio; 3) a única demora significativa é a demora sináptica; 4) a atividade de qualquer sinapse inibitória impede a excitação de um neurônio em qualquer tempo; 5) a estrutura da rede não muda com o tempo.

A primeira pressuposição diz respeito ao disparo do neurônio. Ele pode ter apenas dois estados: ativado ou desativado. De acordo com o que entendiam sobre o cérebro, para o neurônio disparar, precisa haver uma determinada soma de substâncias químicas e elétricas sobre ele. Esta quantidade é transformada num valor numérico no neurônio artificial que, para disparar, precisa exceder seu limiar. Isso independe de sua posição ou atividade prévia, como posto na segunda pressuposição.

No neurônio biológico, a distância do estímulo para o neurônio ao qual deverá estar relacionado não tem importância para o tempo de chegada do estímulo ao alvo. Por isso, a única demora significativa são as ocorridas entre as sinapses, como posto na terceira pressuposição. Um impulso que está a uma distância muito maior de um certo neurônio do que um outro impulso, pode chegar primeiro ao alvo, se o número de neurônios entre eles for menor do que entre o outro estímulo. Cada sinapse é entendida como uma demora

McCulloch & Pitts (1943) criaram um sistema onde o disparo de cada neurônio é determinado por uma fórmula lógica. Desse modo, o neurônio dispara se e somente se sua fórmula correspondente for verdadeira. Uma vez que a sinapse inibitória relativa aumenta o limiar do neurônio, pode ocorrer no próximo tempo que a fórmula seja verdadeira e o neurônio não dispare e com isso a relação de equivalência não é assegurada. Por isso precisam da quarta pressuposição, onde permitem apenas a utilização da sinapse inibitória absoluta.

A quinta pressuposição está, de certa forma, relacionada com a anterior. Ela exige que nem as conexões entre os neurônios, nem o limiar de cada um sejam modificados. Se assim não fosse, a relação entre o disparo do neurônio e o valor de verdade da fórmula lógica correspondente poderia não ser assegurada. Por causa desta limitação, a rede sofreu a crítica que não era capaz de aprendizagem: uma vez criada a rede, dado o limiar e os *inputs* dos neurônios aferentes, as respostas dos neurônios serão sempre as mesmas para os mesmos dados de entrada. Não há como adequar a resposta da rede, sem mudar sua topologia. Por isso, McCulloch foi muitas vezes relacionado antes com o projeto da IA que com o do Conexionismo.

Embora os conexionistas afirmam que a rede de McCulloch-Pitts é uma máquina cujo modo de funcionamento está estritamente assemelhado ao da máquina de Turing, não se pode negar que McCulloch foi o primeiro grande conexionista. Muitas das idéias presentes nesta corrente podem ser encontradas, ainda que imaturas, nos textos de McCulloch. A mais evidente é a que identifica a mente ao cérebro.

2.4 A rede neural artificial de McCulloch-Pitts

A rede construída por McCulloch & Pitts (1943) pode ser intuitivamente descrita do seguinte modo: é composta por um conjunto de neurônios (artificiais) capazes de formar sinapses. Cada neurônio produz uma ação através das sinapses sobre os outros. Ela pode ajudar (sinapse excitatória) ou impedir (sinapse inibitória) a produção de uma resposta dos neurônios sobre os quais está conectada. Cada neurônio ou dispara ou não dispara, sem

haver uma terceira possibilidade. A resposta de um neurônio é determinada pelo cálculo de seus dados de entrada relacionados com seu limiar.

Os neurônios da rede são compreendidos como uma espécie de calculadora que relaciona seu limiar com seus dados de entrada. Se o número de sinapses excitatórias ultrapassar o valor do limiar e nenhuma sinapse inibitória disparar sobre ele, o neurônio deverá disparar. Caso contrário, não disparará.

A idéia de McCulloch & Pitts (1943, p. 115) pode ser resumida no seguinte:

Devido ao caráter 'tudo-ou-nada' da atividade nervosa, os eventos neurais e as relações entre eles podem ser tratados através da lógica proposicional. ... o comportamento de qualquer rede pode ser descrita nestes termos, com a adição de significados lógicos mais complexos para redes que contém círculos; e que para qualquer expressão lógica que satisfaça certas condições, podemos encontrar uma rede agindo no padrão descrito por ela.

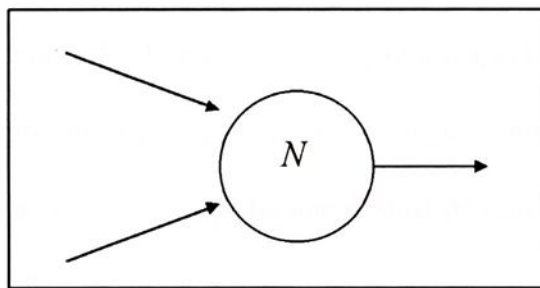


Fig. 4 - Estrutura de um neurônio de McCulloch-Pitts. As setas a esquerda do círculo (soma) representam os dados de entrada do neurônio e a seta a direita representa a sua resposta.

Como dito acima, McCulloch e Pitts pretendiam equivaler o disparo dos neurônios com proposições lógicas denominadas *Expressões Proposicionais Temporais* (EPT). Uma EPT é uma sentença do cálculo proposicional da lógica clássica e é formada por ações de neurônios. Pode estar unida com outras proposições através de conjunções ou disjunções e cada ação pode conter uma negação (representam geralmente as sinapses inibitórias).

A rede é formalmente definida do seguinte modo (McCulloch & Pitts, 1943): os neurônios de uma dada rede N serão designados pelos sinais ' c_1 ', ' c_2 ', ... ' c_n '. $N_i(t)$ significa

que c_i dispara no tempo t . N_i é chamado a *ação* de c_i . Os neurônios de $\tilde{N}$ sem nenhum axônio em sinapse com eles (mais externos a esquerda) são denominados seus *aferentes periféricos*. $N_1, \dots, N_p$ denotarão as ações de tais neurônios e $N_{p+1}, N_{p+2}, \dots, N_n$ as ações dos demais, ou seja, os neurônios de saída (mais externos à direita) e os internunciais ou intermediários, situados entre os de entrada e de saída. Uma *solução* de $\tilde{N}$ será um conjunto de sentenças que regulam o disparo de cada neurônio não aferente periférico da rede.

2.5 Redes e Expressões Proposicionais Temporais

Para cada neurônio, é possível encontrar sua EPT do seguinte modo: seja c_i qualquer neurônio de $\tilde{N}$ com um limiar $\theta_i > 0$, e $c_{i1}, c_{i2}, \dots, c_{ip}$ têm respectivamente $n_{i1}, n_{i2}, \dots, n_{ip}$ de sinapses excitatórias sobre ele. $c_{j1}, c_{j2}, \dots, c_{jq}$ têm sinapses inibitórias sobre c_i . Seja K_i o conjunto de subclasses de $\{n_{i1}, n_{i2}, \dots, n_{ip}\}$ tal que a soma de seus membros excedem θ_i . De acordo com estas suposições, mostramos que

$$N_i(z_1) \equiv S \left\{ \bigwedge_{m=1}^q \neg N_{jm}(z_1) \wedge \sum_{a \in K_i} \prod_{s \in a} N_{is}(z_1) \right\}^{13} \quad (1)$$

onde ' $\sum$ ' ' $\prod$ ' são símbolos sintáticos para disjunções e conjunções, as quais são finitas em cada caso. Esta expressão significa que o neurônio c_i dispara se e somente se nenhuma sinapse inibitória dispara sobre ele e que pelo menos um número somado de sinapses excitatórias maior do que o limiar de c_i dispara sobre ele. Uma expressão desta forma pode

¹³ No cálculo lógico, McCulloch e Pitts criam um *functor* S , em que: sendo $P_i(t)$ uma propriedade sobre um ou mais neurônios quaisquer num tempo determinado, o acréscimo do S significará que a mesma propriedade P vale para o momento anterior ao tempo t . Por exemplo, $N_i(t)$ significa que um determinado neurônio dispara no tempo t e $S N_i(t)$ significa que este mesmo neurônio dispara no momento $t-1$.

ser escrita para cada c_i que não seja um aferente periférico. Substituindo a expressão correspondente em (1) para cada N_{jm} e N_{is} que não seja um aferente periférico (e repetindo o mesmo processo do resultado para todos os não aferentes periféricos), construímos uma expressão para N_i que contenha apenas aferentes periféricos.

Também é possível encontrar, para qualquer EPT, a rede cujo último neurônio terá sua ação equivalente ao valor de verdade da expressão. Para isso, deve-se levar em conta a forma da EPT e o número de demoras sinápticas existentes na expressão (para maiores detalhes, veja McCulloch & Pitts, 1943).

Cada EPT terá uma das seguintes formas:

- (a) $N_i(t)$ - atômica.
- (b) $N_i(t-n)$ - demora
- (c) $p_1(z_1) \wedge p_2(z_1)$ - conjuntiva
- (d) $p_1(z_1) \vee p_2(z_1)$ disjuntiva
- (e) $p_1(z_1) \wedge \neg p_2(z_1)$ conjuntiva negada

O símbolo 'n' na segunda EPT é um número natural qualquer. Os símbolos $p_i(z_1)$ são variáveis metalinguísticas que representam uma EPT qualquer. Assim, $p_1(z_1)$ significa que o neurônio 1 dispara no tempo z_1 . Este neurônio pode estar ligado, direta ou indiretamente, com uma infinidade de outros neurônios que irão determinar o seu disparo. As redes cujo último neurônio têm seu disparo equivalente ao valor de verdade destas sentenças acima são as da fig. 5. Nestas redes, cada círculo sobre o neurônio significa uma sinapse inibitória e cada seta uma excitatória. O número no interior dos neurônios significa sua ordem e não seu limiar, que é sempre igual a 1.

Na rede (b) existe uma demora sináptica. Para as expressões com n demoras sinápticas, basta acrescentar n neurônios à esquerda do último neurônio da rede. Sendo

assim, seu disparo é equivalente ao valor de verdade desta EPT. Nas redes (a)-(e), a variável 'z₁' pertencente às formas de EPT apresentadas há pouco, é neste caso, igual a $t-1$. As redes da fig. 5 são as básicas para o cálculo de McCulloch-Pitts. Toda rede pode ser construída a partir delas.

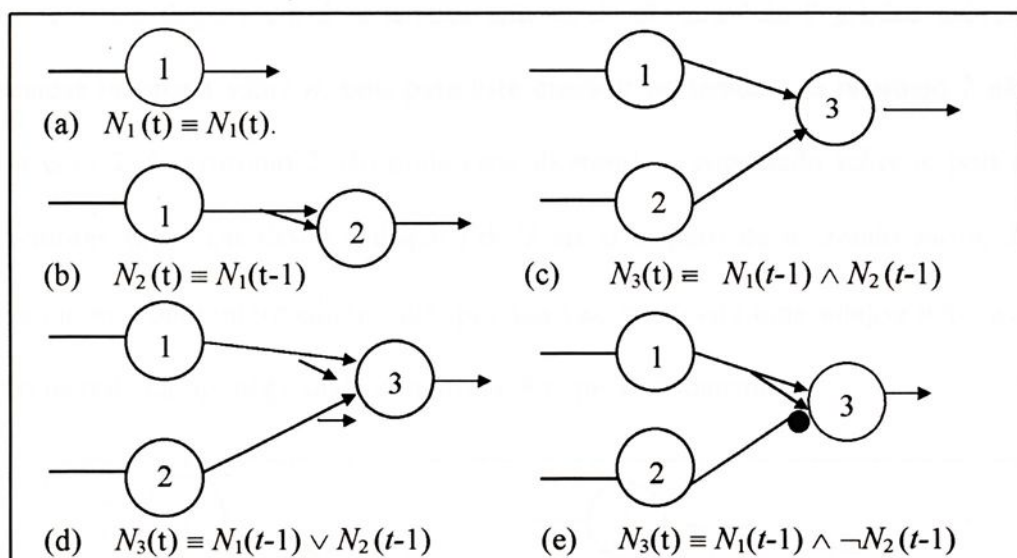


Fig 5 - Redes básicas com sua EPT equivalente.

A seguir, apresentamos dois exemplos de como é feita a construção de redes cujo neurônio de saída é equivalente ao valor de verdade das seguintes EPTs:

$$S_1: N_1(t-1) \wedge \neg N_2(t-1)$$

$$S_2: N_1(t-1) \wedge \neg N_2(t-2)$$

A forma da primeira sentença é conjuntiva negada. Sendo assim, o neurônio de saída da rede que temos de construir irá disparar no tempo t se e somente se forem satisfeitas as duas partes da fórmula. Ou seja, terá de acontecer que o neurônio 1 dispare em $t-1$ e 2 não dispare neste mesmo tempo. Se 2 não pode disparar, significa que ele está produzindo uma sinapse inibitória sobre o neurônio de saída da rede. Além disso, a ação de 1 precisa exceder

o limiar do neurônio de saída, ou seja, precisa ter no mínimo duas terminações axonais sobre ele. É exatamente assim a rede da fig. 6(f).

A sentença S_2 tem a forma conjuntiva negada. De acordo com esta sentença, o disparo do neurônio 1 no tempo $t-1$ é suficiente para disparar um neurônio qualquer n no tempo t , desde que 2 não dispare em dois tempos anteriores. O neurônio 2 produz indiretamente uma sinapse inibitória sobre n , pois para este disparar no tempo t , o neurônio 2 não pode disparar em $t-2$. O neurônio 2 não pode estar diretamente conectado sobre n , pois existem duas demoras sinápticas desde o disparo de 2 até o disparo de n . Sendo assim, 2 estará ligado a um neurônio intermediário que, por sua vez, produzirá uma sinapse inibitória sobre n , como na rede da fig. 6(g), onde o neurônio 4 é que denominamos n .

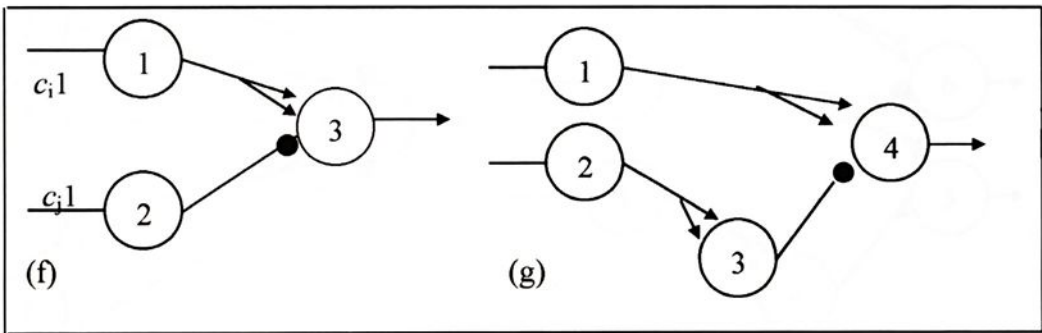


Fig. 6 - Redes cujo neurônio de saída satisfaz S_1 e S_2

Os dois exemplos acima podem ilustrar como podemos construir redes cujo disparo do neurônio de saída é determinado por uma EPT. O processo inverso é dado utilizando as redes da fig. 7.

Para as redes da fig.7 abaixo, podemos encontrar sua solução do seguinte modo: na rede (h), os neurônios que produzem uma sinapse inibitória sobre o neurônio 6, que é o único não aferente periférico, são 1 e 5. Sendo assim, estes neurônios deverão estar com suas ações negadas e unidas através de conjunção. Esta conjunção estará em conjunção com

uma outra expressão formada por disjunções de conjuntos de neurônios cuja soma de seus estímulos excedem o limiar de 6. Sendo assim, a solução de (h) será constituída pela seguinte EPT:

$$N_6(t) \equiv \{ [\neg N_1 \wedge \neg N_5] \wedge [N_2 \vee N_4 \vee (N_2 \wedge N_3) \vee (N_3 \wedge N_4) \vee (N_2 \wedge N_4) \vee (N_2 \wedge N_3 \wedge N_4)](t-1) \}$$

onde $K_i = \{ \{2\}, \{4\}, \{2,3\}, \{3,4\}, \{2,4\}, \{2,3,4\} \}$.

A figura (i) tem sua solução composta de quatro fórmulas

$$S_1: N_4(t) \equiv N_3(t-1)$$

$$S_2: N_5(t) \equiv \neg N_2(t-1) \wedge N_4(t-1)$$

$$S_3: N_6(t) \equiv \neg N_2(t-1) \wedge N_1(t-1)$$

$$S_4: N_7(t) \equiv N_1(t-1)$$

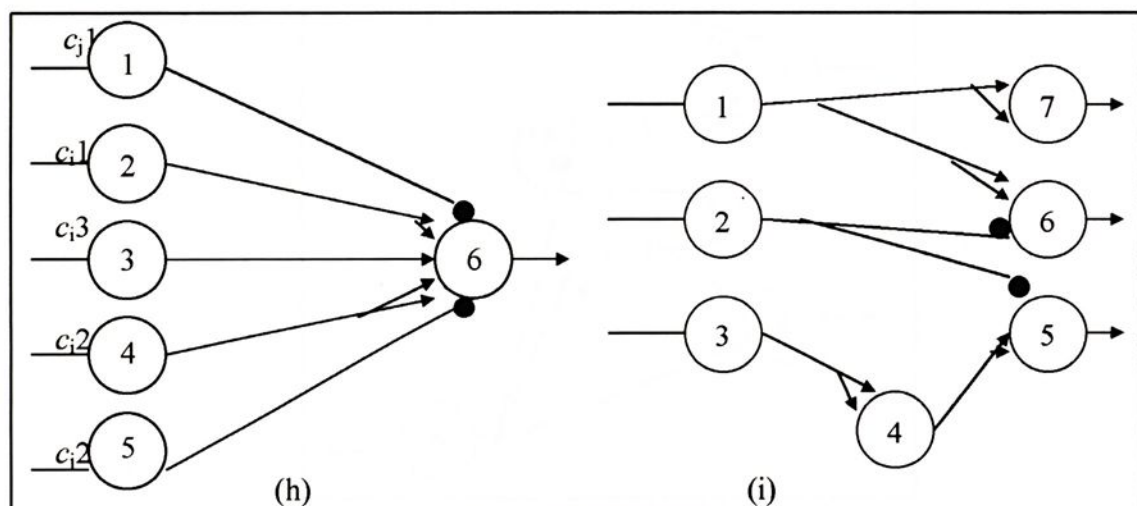


fig. 7 - Exemplos de rede sem solução

A idéia de McCulloch é a de que cada neurônio de saída das redes represente uma atividade mental ou nervosa como por exemplo sensação de frio, calor, dor, etc. Desse modo, todo evento mental pode ser calculado por meio de uma TPE e cada TPE, por sua vez, representa um evento.

Citaremos, a seguir, um exemplo de uma rede que simula a sensação de frio e calor. O evento é assim descrito por McCulloch & Pitts (1943): se um objeto frio é tocado na pele

por algum tempo e removido, uma sensação de calor será percebida; se ele for aplicado por um longo tempo, a sensação será de frio. Sabe-se que um receptor cutâneo é afetado pelo calor, e outro pelo frio. Sejam N_1 e N_2 as ações dos respectivos receptores e N_3 e N_4 de neurônios cuja atividade implica a sensação de calor e frio, estas sensações podem ser simuladas pela rede da fig. 8, cuja solução é:

$$N_3(t) \equiv N_1(t-1) \vee N_2(t-3) \wedge \neg N_2(t-2)$$

$$N_4(t) \equiv N_2(t-2) \wedge N_2(t-1)$$

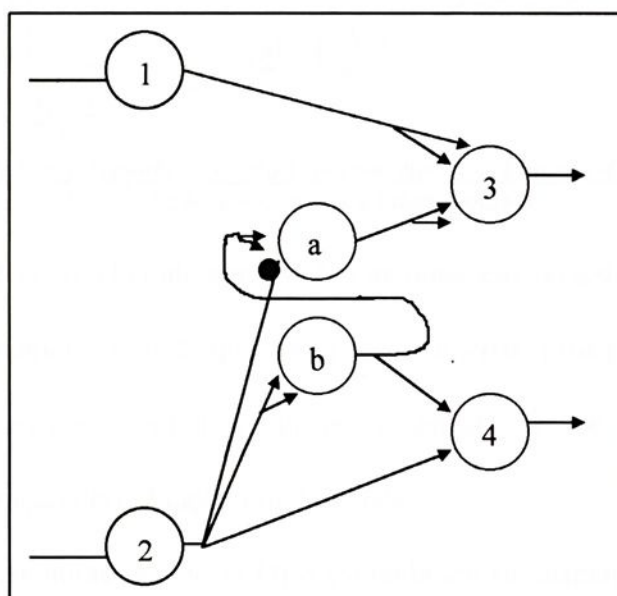


Fig. 8 - Rede que simula sensação de frio e calor.

2.6 Redes com círculos

As redes apresentadas até o momento podem simular diversos eventos mentais. Porém, são ineficientes para a realização de alguns comportamentos tais como a soma de dois números quaisquer. Para isso, é preciso criar redes mais complexas, chamadas *redes com círculos ou circulares*. Sua caracterização é a mesma das redes simples, exceto pela

seguinte definição: um círculo é uma cadeia de $c_i, c_{i+1} \dots$ neurônios, cada membro da cadeia em sinapse sobre o próximo, com o mesmo começo e fim, ou seja, o último membro da cadeia está em sinapse com o primeiro.

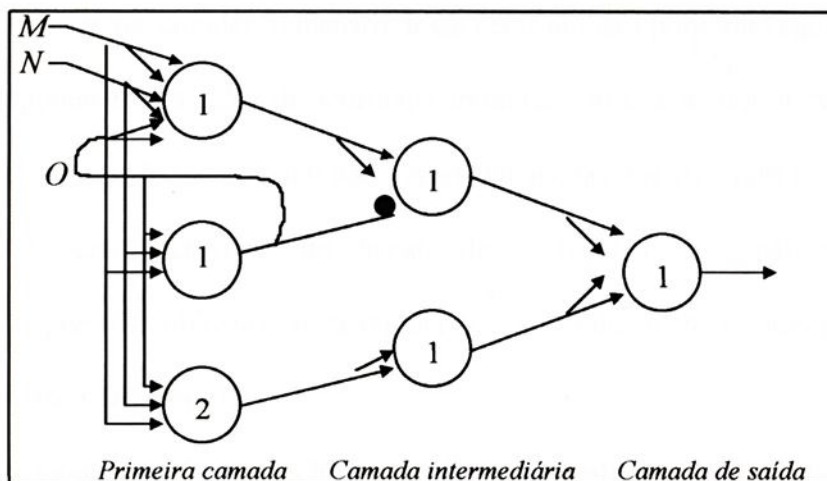


Fig. 9 - Rede que computa a função soma

Com estas redes é possível computar funções matemáticas no sistema binário, como a soma. A fig. 9 é um exemplo de rede que soma dois números quaisquer. O círculo desta rede encontra-se no neurônio central da primeira camada. É ele o responsável pela possibilidade da computação desta função por esta rede.

Se pelo menos duas fibras (M, N ou O) o estimularem ele dispara, inibindo o disparo do neurônio intermediário superior. Se apenas duas fibras forem estimuladas no tempo t a resposta da rede no tempo $t+3$ será 0 (o neurônio de saída não dispara). Se três fibras forem estimuladas, a resposta será 1, porque o neurônio inferior da primeira camada dispara. Caso apenas uma fibra seja estimulada, a resposta é 1 e se nenhuma fibra for estimulada, a resposta é 0.

O neurônio central da primeira camada é o responsável pela “memorização” da soma em cada cálculo de M e N. Se testarmos a rede, veremos que ela realmente soma dois

números quaisquer, assim como o exemplo de máquina de Turing 3.3.2. Como naquela máquina, aqui os dados de entrada e a resposta não são memorizados

Nas redes com círculo não é possível construir uma EPT para o neurônio de saída da rede. Pelo fato da rede ser circular, o disparo desse neurônio não pode ser regulado por uma EPT, cujos componentes são ações de neurônios aferentes periféricos. Isso porque o disparo dos neurônios que pertencem ao círculo não dependem apenas dos aferentes periféricos, mas direta ou indiretamente, dependem do disparo deles próprios, que não são aferentes periféricos. As expressões utilizadas para regular o disparo desses neurônios pertencem ao cálculo de predicados (Kleene, 1956).

Embora o modelo de McCulloch-Pitts tenha sua construção distinta da máquina de Turing, podemos notar algumas semelhanças no modo de funcionamento de ambas. Para computar uma função ou representar uma atividade nervosa, a rede deve ser construída de tal forma que, apresentado um dado de entrada a ela, deve produzir a resposta correta. Se produzir uma resposta errada, deve-se construir uma nova rede. Não existe, tanto na rede de McCulloch-Pitts quanto na máquina de Turing, uma regra de aprendizagem.

Pensando em tomar o erro como um aliado e não como um problema, surgem na década de 1950 novas redes neurais artificiais. Ao invés de ter de construir uma nova rede sempre que houvesse um erro, ele seria utilizado para uma mesma rede corrigi-lo até responder corretamente ao dado de entrada. Para os criadores destas redes, esta é uma das características que distinguem a aprendizagem e a possibilidade de comportamento inteligente por parte da máquina. Além disso, dizem, o modo de funcionamento do modelo não é algorítmico, como na máquina de Turing e, de certa forma, na rede de McCulloch-Pitts (Rosenblatt, 1962, Kohonen, 1987).

3 Caracterização das redes neurais artificiais modernas

3.1 O cérebro e as redes neurais segundo os conexionistas atuais

Embora o conhecimento sobre o cérebro humano tenha evoluído muito desde 1943, podemos dizer que o conhecimento base sobre ele ainda é o mesmo. Por isso, a caracterização feita na seção 2.2 vale também para os conexionistas atuais.

A estrutura de uma rede neural artificial é inspirada em uma visão simplificada da arquitetura neural biológica. As RNA são formadas por nódulos, interconexões e sinapses. Estes correspondem aos neurônios, axônios e sinapses da rede neural biológica.

A figura 10 abaixo ilustra a estrutura topológica de uma rede neural artificial padrão:

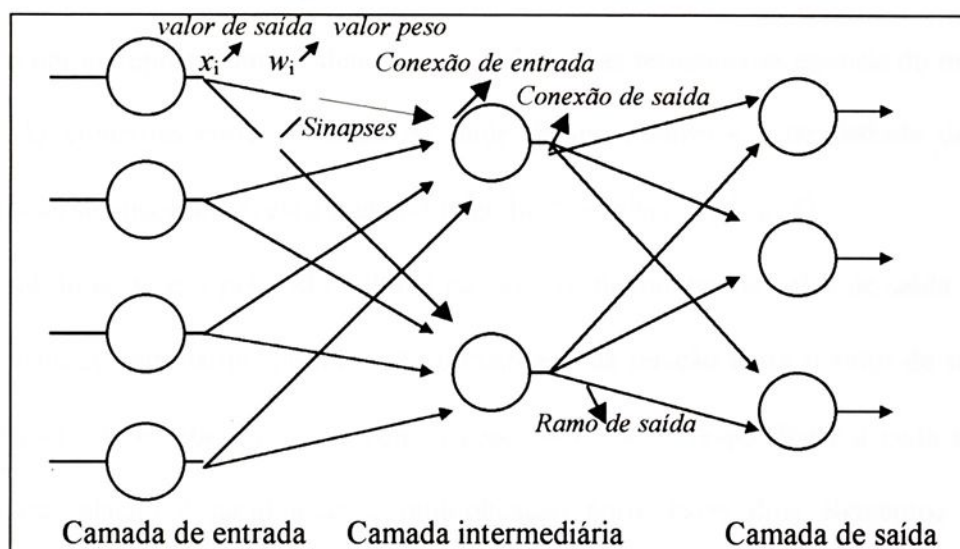


Fig. 10 - Estrutura topológica de uma rede neural artificial

Segundo Caudill & Butler (1992, p. 3) uma RNA é

...um sistema de processamento de informação que é não-algorítmico, não-digital e intensamente paralelo. Não é um computador no sentido de como pensamos atualmente, nem é programado como um computador. De fato, consiste de um número de processadores muito simples e altamente interconectados chamados nódulos, análogos às células neurais biológicas, ou neurônios, no cérebro. Os nódulos conectam-se uns com os outros através de linhas com pesos, sob as quais os sinais podem passar. Cada nódulo recebe muitos sinais através de sua

conexões de entrada; alguns destes sinais de entrada podem provir de outros nódulos, e outros podem vir do ambiente externo. O nódulo geralmente tem muitas destas conexões de entrada; contudo, nunca produz mais que um único sinal de saída. Este sinal é transmitido através da conexão de saída do nódulo [correspondente ao axônio de um neurônio biológico], que geralmente se divide num grande número de conexões menores [botões sinápticos do neurônio biológico] cada uma das quais termina num destino distinto. Cada um destes ramos da conexão de saída única transmite o mesmo sinal; este sinal não é dividido entre eles de modo algum. Muitos destes ramos de saída terminam nas conexões de entrada de algum outro nódulo na rede; outros podem terminar no lado de fora da rede e gerar padrões de controle ou de resposta.

A definição de rede apresentada acima cita características como conexões de entrada, linhas com peso, sinais de entrada e de saída, conexão de saída, ramos da conexão de saída. Estes termos são usados quando desejamos conceituar “formalmente” a rede. Porém, quando vamos realmente construí-las, tais características transformam-se em valores numéricos ou linhas, constituindo a estrutura “real” de uma RNA. Como pode-se notar na fig. 10, as conexões de entrada de um nódulo são as linhas ligadas a ele à sua esquerda. As conexões de saída são a linha ligada a ele à sua direita e suas sublinhas são os ramos de saída. Estas partes representam os dendritos, o axônio e as terminações axonais do neurônio biológico. As conexões entre as linhas de saída de um nódulo e a de entrada de outro representam as sinapses do cérebro humano (Caudill & Butler, 1992, p. 4).

O sinal de saída e o peso são valores numéricos, denominados valor de saída e valor peso. Os sinais de entrada dos nódulos são o resultado da relação entre o valor de saída do nódulo do qual ele recebe um *input* com o peso da linha correspondente a cada sinal de entrada. Esta relação é geralmente a multiplicação entre estes dois elementos. Como consequência, na “construção” da rede o sinal de entrada é também um valor numérico.

O processamento de informações na rede não é feito de maneira algorítmica, como na máquina de Turing, mas sim, através da ativação dos nódulos. A memória da rede não é um lugar separado desta, como uma fita, no caso da máquina de Turing, mas é constituída pelos pesos da rede.



Não existe uma definição formal de RNA como existe para a máquina de Turing, por exemplo. Cada rede tem suas particularidades e a definição geral não é algumas vezes respeitada. Os próprios conexionistas admitem isso. Observe a citação de Caudill & Butler (1993, p. 7), por exemplo: “Note que estamos listando as características da rede neural ideal. As distinções entre a ideal e a real serão apontadas ao examinarmos diferentes modos para implementar redes neurais.”

O valor de saída de um nóculo é obtido através do seguimento de três passos: cálculo da *entrada total* do nóculo, cálculo de sua *ativação*, e a relação deste resultado com o limiar do nóculo, ou seja, o cálculo da *saída* do nóculo. Estas três etapas constituem o que geralmente denominamos de *função transferência* e seu cálculo é realizado como descrito a seguir (Caudill e Butler, 1992, p. 15-6).

Como dito, a primeira etapa da função transferência consiste em calcular a entrada total do nóculo. Tal cálculo, para um nóculo i , é comumente encontrado através da somatória das multiplicações de cada valor de saída dos nósculos ligados a i com o respectivo valor peso. A multiplicação do valor peso pelo valor de saída pode originar um resultado negativo ou positivo, que representarão as sinapses inibitórias ou excitatórias, respectivamente. Em símbolos, a entrada total do nóculo é calculada pela seguinte função

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

Nesta função, Σ significa a somatória, I_i é a entrada total do nóculo i a partir de n nósculos. O valor peso do nóculo j sobre i é representado por w_{ij} . O valor de saída de um nóculo j sobre i (que é o mesmo para qualquer outro nóculo) é representado por x_j .

Na segunda etapa da função transferência, a entrada total é convertida num nível de ativação do nóculo. A ativação de um nóculo num determinado momento depende exclusivamente de sua ativação no momento anterior e de sua entrada total, relacionadas por alguma função, ou seja,

$$a_i(t) = F_i[a_i(t-1), I_i(t)]$$

Na função acima, $a_i(t)$ é a ativação do nóculo i no tempo t .

A terceira e última etapa da função transferência consiste no cálculo da resposta do nóculo, dada pela seguinte função (onde T é o limiar do neurônio)

$$y_i = \begin{cases} a_i, & \text{se } a_i > T \\ 0, & \text{caso contrário} \end{cases}$$

A função ativação é geralmente aplicada a todos os nósculos da rede. Todos eles têm sua resposta calculada através de uma mesma função transferência. Esta é uma caracterização formal da função transferência das redes neurais. Porém, há diferenças entre a função transferência de alguns modelos e esta formal. No perceptron, por exemplo, não há o cálculo da função ativação, que de fato é igual à sua entrada total. Por outro lado, a resposta do nóculo pode ser bipolar (+1 ou -1) ou binária (1 ou 0), isto depende da rede. Como já dissemos, os conexionistas ainda não encontraram uma definição formal de rede neural tal como existe para a máquina de Turing.

3.2 O reconhecimento de padrões

Um dos objetivos fundamentais dos conexionistas é fazer com que as RNA reconheçam padrões. Existem dois tipos de padrão: o primeiro é caracterizado como uma

propriedade emergente da dinâmica da interação entre os nódulos de uma rede (Caudill & Butler (1992). O segundo, aqueles que as redes são capazes de reconhecer, é um conjunto de elementos que representa algum objeto. Nas simulações de RNA, um padrão é um conjunto de valores numéricos distribuídos matricialmente. Desse modo, suponhamos que desejamos apresentar para a rede uma letra do alfabeto. Tal letra seria representada por um conjunto de valores numéricos. Cada nódulo da camada de entrada representa uma pequena parcela do padrão de entrada. Cada uma destas partes é transmitida a todos os nódulos das camadas intermediárias. Desse modo, cada nódulo recebe uma versão modificada do padrão de entrada (Caudill e Butler, 1992). É modificada porque cada parte do padrão é multiplicada por um valor peso. Como cada nódulo das camadas intermediárias estão ligados com todos os outros da camada posterior, cada um destes também recebe o padrão total modificado. Pelo mesmo raciocínio, cada nódulo da camada de saída também recebe o padrão total modificado. Não importa o número de camadas que a rede possui nem quantos nódulos possui cada camada. Se todos eles estiverem conectados, todos receberão o padrão total. Para aquelas redes que não estão totalmente conectadas, tal característica não lhes é atribuída.

Tanto nas redes cujos nódulos conectam-se a todos os nódulos da camada posterior a ela, quanto para as redes que não possuem tal características, o objetivo é fazer com que ela reconheça o padrão apresentado. Isso significa que a rede deve classificá-lo corretamente. Ou seja, a resposta da rede deve estar de acordo com a classe à qual o padrão pertence. Em outros casos, a resposta de cada nódulo de saída deve ser uma característica particular do padrão. Para que isso seja possível nas RNA, deve-se representar os padrões de tal modo



que a rede seja capaz de recebê-los. O meio pelo qual isso é feito é pela transformação de cada elemento de um padrão por um valor numérico.

Para visualizar a localização física dos padrões, podemos distribuí-los em espaços geométricos, de acordo com o número de elementos pertencentes a eles (um padrão com n elementos é representado num espaço n dimensional). Por exemplo: suponhamos que queremos visualizar a posição de padrões com dois elementos. Podemos distribuí-los num espaço euclidiano bidimensional, como na fig. 11. A posição de cada figura geométrica no eixo x representa o primeiro elemento do padrão e no eixo y o segundo elemento. Uma vez apresentado o padrão à rede ela poderá, em princípio, classificá-los corretamente. No caso da fig. 11 temos duas classes (quadrados e círculos). Para cada padrão apresentado à rede, ela o reconhecerá se classificá-lo de acordo com a classe à qual pertence. Como neste caso, as classes podem ser separadas por uma reta e por isso são *linearmente separáveis*.

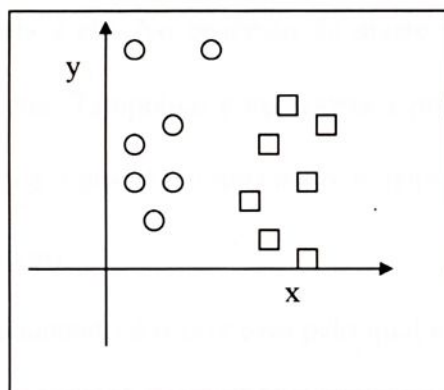


Fig. 11 - Visualização de padrões com dois elementos num espaço euclidiano

Para uma rede aprender a reconhecer padrões, é preciso que ela responda corretamente a um dado padrão de entrada. Desse modo, precisa ir se adequando (modificando-se) até que responda corretamente ao padrão atual. O que pode mudar a resposta da rede são os valores peso e de entrada dos nódulos. É através da sua modificação

que podemos adequar a resposta da rede ao padrão. Porém, não podemos mudar os valores de entrada da rede, uma vez que são eles os valores do padrão de entrada. Sendo assim, se os modificássemos, a rede não estaria mais reconhecendo este mesmo padrão. Portanto, o que resta para mudar são os valores peso dos nódulos.

O processo pelo qual modificamos os pesos da rede a fim de que ela reconheça um determinado padrão é denominado *treinamento*. Dois dos principais modos de treinamento da rede são os seguintes: *a) supervisionado*: leva em consideração a resposta desejada para os padrões e a relaciona com a resposta atual da rede. Daí, calcula o erro entre elas para, a partir deste erro, modificar os pesos dos neurônios. Nesse tipo de treinamento, há ainda a necessidade de um supervisor ou “professor”. Sua função é modificar corretamente os pesos e analisar se a rede conseguirá ou não aprender; *b) não-supervisionado* ou *auto-organização*: nesse tipo de treinamento, a rede recebe apenas o padrão de input e nenhuma resposta desejada é apresentada a ela. No processo de ajuste dos pesos não é levada em consideração a resposta desejada. Tampouco é necessária a presença de um supervisor. O ajuste dos pesos em ambos os tipos de treinamento é feito através de uma *regra de treinamento* (ou de aprendizagem).

Deve-se notar que o treinamento é o processo pelo qual a rede ajusta seus pesos para reconhecer determinados padrões. O resultado final deste processo é a aprendizagem (Caudill & Butler, 1992, p. 8).

A seguir, apresentamos dois modelos conexionistas: o *perceptron* e a *rede de Kohonen*. O primeiro tem seu treinamento supervisionado e o segundo apresenta elementos de auto-organização.



4 O perceptron e a rede de kohonen

4.1 A definição do perceptron

O perceptron, criado na década de 1950 por Rosenblatt (1962), foi uma das primeiras redes a surgir depois de 1943. Rosenblatt afirmava que o perceptron não deveria ser comparado à rede de McCulloch-Pitts. Isso porque estes cientistas não levaram realmente em consideração as características do cérebro na construção de sua rede. Em outras palavras, foram pouco biológicos ao construí-la. Tais críticas referem-se ao fato das redes possuírem um funcionamento quase algorítmico. Além disso, como já afirmamos, McCulloch e Pitts não tinham uma regra de treinamento para a aprendizagem da rede.

O perceptron foi criado fundamentalmente para modelar a percepção visual. É formado por apenas um nódulo, geralmente com apenas dois canais de entrada e um de saída, como representado na fig. 12. O objetivo para esta rede é classificar padrões em duas classes distintas.

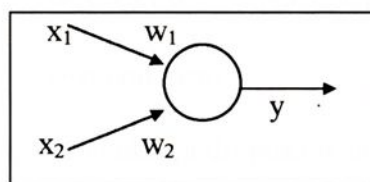


Fig. 12 - estrutura de um perceptron simples

O nódulo do perceptron tem basicamente o mesmo funcionamento do proposto por McCulloch-Pitts. Em primeiro lugar, “computa” sua entrada total. Se ele exceder o limiar, a resposta do neurônio é +1 (dispara). Caso contrário, é -1 (não dispara).

A função transferência do perceptron é formada por dois passos (Caudill e Butler, 1992, p. 35-6): o primeiro é o cálculo da entrada total, encontrado pela seguinte função

$$I = \sum_{i=1}^n w_i x_i$$

onde w_i x_i são vetores peso e input¹⁴. Ou seja, o *input* total da rede é a somatória da multiplicação de cada valor de entrada pelo seu valor peso.

O segundo passo da função transferência do perceptron é o cálculo da resposta do nóculo, encontrada pela seguinte função

$$y = \begin{cases} +1, & \text{se } I \geq T \\ -1, & \text{se } I < T \end{cases}$$

Ou seja, o nóculo dispara quando a entrada total é maior ou igual do que seu limiar. Caso contrário, não dispara. No caso do perceptron, o limiar é geralmente assumido como sendo 0 e a resposta é bipolar por facilidades matemáticas.

A regra de treinamento do perceptron é a seguinte:

$$w_{\text{novo}} = w_{\text{velho}} + \beta yx$$

$$\beta = \begin{cases} +1, & \text{se a resposta do perceptron está correta} \\ -1, & \text{caso contrário} \end{cases}$$

Podemos notar, na função acima, que a mudança do peso necessita da resposta desejada do nóculo. Por isso dizemos que sua aprendizagem é supervisionada.

O treinamento deve ser feito de uma forma organizada e ordenada. Para que a rede aprenda a reconhecer um determinado grupo de padrões, será preciso seguir um conjunto de regras, reunidas no *algoritmo de treinamento*. No caso do perceptron, ele é o seguinte:

¹⁴ Um vetor é um conjunto ordenado de elementos. Assim, para o vetor w , w_1 significa o primeiro elemento deste conjunto, w_2 o segundo, e assim por diante.

1. Para cada padrão no conjunto de treinamento
 - 1.1 aplicar o próximo padrão para o perceptron
 - 1.2 gravar a resposta do perceptron
 - 1.3 se a resposta do perceptron está correta,
 - e a resposta foi +1, então
o novo vetor peso = velho vetor peso + o vetor do padrão de input
 - e a resposta foi -1, então
o novo vetor peso = velho vetor peso - o vetor do padrão de input
 - 1.4 se a resposta do perceptron está incorreta.
 - e a resposta foi +1, então
o novo vetor peso = velho vetor peso - o vetor do padrão de input
 - e a resposta foi -1, então
o novo vetor peso = velho vetor peso + o vetor do padrão de input
2. finalize para cada padrão no conjunto de treinamento. (Caudill e Butler, 1992, p. 29)

Ao fazer os cálculos para todos os padrões, deve-se analisar se o perceptron classifica corretamente cada um deles. Se classifica, ele aprendeu, caso contrário, deve-se recomeçar a fazer os cálculos para cada padrão até que a rede aprenda a reconhecê-los.

A seguir, apresentamos um exemplo de um perceptron que, segundo Caudill e Butler (1992) aprendeu a reconhecer quatro padrões, pertencentes a duas classes A e B. Se o padrão pertence à classe A, o perceptron deve disparar. Se o padrão pertence ao conjunto B, a resposta deve ser -1, ou seja, não dispara. As classes são formadas pelos seguintes elementos, que podem ser visualizados na fig. 13 (cada reta nos eixos representa 0,2):

$$\begin{array}{ll} A_1 = (0,3;0,7) & B_1 = (-0,6;0,3) \\ A_2 = (0,7; 0,3) & B_2 = (-0,2;-0,8) \end{array}$$

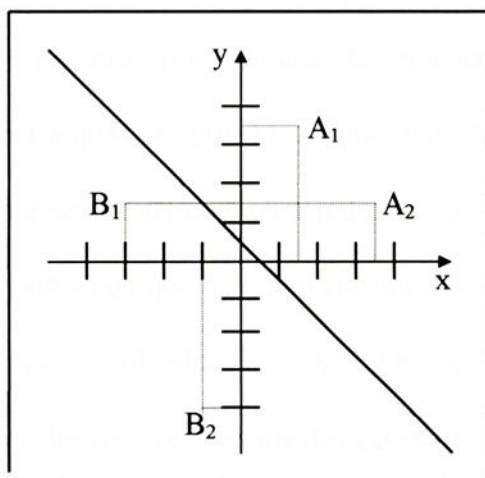


Fig. 13 - visualização dos elementos dos conjuntos A e B

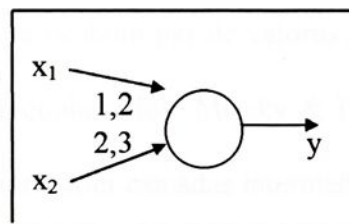


Fig. 14 - Perceptron já treinado

Observe-se que as classes acima são linearmente separáveis. O perceptron que aprendeu a reconhecer os padrões, com seus pesos corretos é o da fig. 14, onde x_1 representa o primeiro elemento do padrão (ligado ao eixo x da fig. 13) e x_2 representa o segundo elemento do padrão, (ligado ao eixo y da fig. 13). Os pesos iniciais neste perceptron foram $(-0,6; 0,8)$. Esses pesos foram escolhidos aleatoriamente. Através de alguns cálculos, obtemos os valores peso que estão no perceptron da fig. 14. Com eles, o nóculo é capaz de identificar todos os objetos à sua classe (dos quatro que foram utilizados no treinamento). Ou seja, o perceptron aprendeu a reconhecê-los, afirmam Caudill & Butler (1992). Porém, se os pesos iniciais fossem outros, poderia perfeitamente haver pesos finais distintos dos da fig. 14 com os quais o perceptron reconheceria todos os padrões do conjunto de treinamento.

Por ser capaz de realizar comportamentos como o acima citado, o projeto conexionista começou com muita euforia na década de 1950. Tal euforia, porém, durou pouco tempo. Minsky & Papert (1969) demonstraram que o perceptron não podia resolver problemas não linearmente separáveis. Um dos exemplos, destes problemas é o do exclusivo da lógica proposicional, cuja tabela de verdade é mostrada na fig. 15. Os elementos dos padrões e a resposta são digitais (0 ou 1). Note pela fig. 16 que não é possível separar ambas as classes com uma reta. Não existirá nenhum par de valores pesos tais que, ao apresentar os objetos à rede, ela conseguirá reconhecê-los. Minsky & Papert (1969) afirmaram ainda que nem perceptrons mais sofisticados com camadas intermediárias seriam capazes de resolvê-los. Isso porque não haveria um bom modo de saber quais as respostas desejadas para os nósculos das camadas intermediárias da rede. Além de não poder

resolver problemas linearmente inseparáveis, Minsky & Papert (1969) mostraram que o perceptron não é capaz de separar mais do que duas classes distintas.

	A	B	$A \vee B$
B_1	1	1	0
A_1	1	0	1
A_2	0	1	1
B_2	0	0	0
x	y		

Fig. 15 - Tabela de verdade do ou-exclusivo

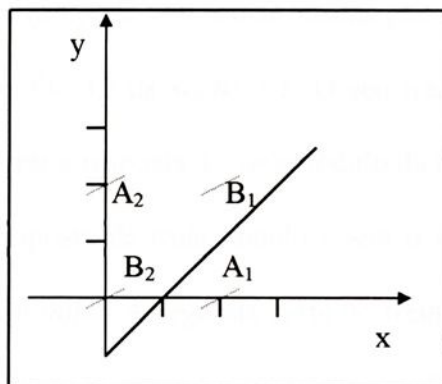


Fig. 16 - Visualização dos padrões do ou-exclusivo

As críticas de Minsky & Papert fizeram com que o projeto conexionista tivesse seu progresso abalado por um longo tempo. Abalado, mas não eliminado. Alguns pesquisadores, como Hopfield (1982), Kohonen (1987), Rumelhart & McClelland (1988), continuaram trabalhando neste projeto. Na década de oitenta, criaram redes capazes de resolver problemas não linearmente separáveis e reconhecer objetos, separando-os em mais do que apenas duas classes distintas. Eles concordavam com o argumento de Minsky & Papert de que as redes de multi-camadas não eram capazes de resolver estes problemas. Isso se a regra de treinamento fosse a mesma do perceptron simples. Porém, se esta regra fosse modificada ou aperfeiçoada, tal argumento seria refutado. Foi exatamente o que fizeram os conexionistas: criaram redes cuja regra de treinamento é uma extensão da regra do perceptron simples, denominada *backpropagation*. Apresentamos um exemplo destas redes a seguir

4.2 O perceptron multi-camadas

As redes que empregam a regra *backpropagation* são formadas por um conjunto de nódulos ordenados em camadas, como na fig. 10 da seção 3.1. O seu treinamento possui duas partes. A primeira consiste em encontrar a resposta de cada nódulo da rede. A começar pelos nódulos da camada de entrada, a resposta de cada nódulo j será o valor de entrada para os nódulos sobre os quais j produz um *input*. A segunda parte do treinamento consiste em comparar as respostas atuais dos nódulos de saída em relação ao padrão atual com a resposta desejada para ele e calcular o erro existente entre estas respostas. A partir deste erro, o valor peso de todos os nódulos da rede serão modificados. O erro é propagado para os nódulos da camadas anteriores, sendo utilizado para a modificação dos pesos.

A primeira parte do processo de treinamento das redes *backpropagation* é feita do mesmo modo como no perceptron simples. A diferença fundamental encontra-se na segunda parte, onde a mudança dos pesos dos nódulos das camadas intermediárias não depende de sua resposta desejada. Na verdade, ela depende da resposta desejada dos nódulos de saída, aos quais aqueles estão direta ou indiretamente conectados. A regra de treinamento será:

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_{pj} o_{pj}$$

onde, $w_{ij}(t)$ é o peso do neurônio i sobre o neurônio j no tempo t

η é uma constante de aprendizagem que varia de 0 a 1

δ_{pj} erro do neurônio j com respeito ao padrão p

o_{pj} resposta do neurônio j com respeito ao padrão p .

Para as unidades de saída, o erro de um neurônio j para o padrão p é assim calculado:

$$\delta_{pj} = K o_{pj} (1 - o_{pj}) (t_{pj} - o_{pj})$$

onde K é uma constante de aprendizagem

o_{pj} é a resposta atual do neurônio j em relação ao padrão p

t_{pj} é a resposta desejada do neurônio j em relação ao padrão p .

Para as unidades intermediárias,

$$\delta_{pj} = K o_{pj} (1 - o_{pj}) \sum \delta_{ps} w_{js}$$

onde s representa os neurônios sobre os quais j produz um *input*.

Sendo assim, a mudança dos pesos de um nóculo intermediário j é calculada pela multiplicação de uma constante de aprendizagem pela resposta atual de j . Isso tudo é multiplicado pelo resultado da subtração desta resposta por 1. Por fim, multiplica-se este resultado pela somatória da multiplicação do erro de cada nóculo s sobre o qual j produz um *input* pelo peso de s em relação a j . Note que, em última análise, cada nóculo terá seus pesos modificados de acordo com o erro dos nós de saída. Sendo assim, este erro é, de certa forma, propagado para todos os outros nós da rede que não sejam da camada de saída.

Com estas redes é possível resolver problemas linearmente inseparáveis, como o do ou-exclusivo. Além disso, pode-se reconhecer padrões classificando-os em diversas classes. Para o primeiro tipo de problemas, podemos citar as redes (já treinadas) da fig.17 que reconhecem os padrões do problema do ou-exclusivo. Para o segundo tipo, deve-se construir redes com o número de nós de saída igual ao número de classes a serem utilizadas. Cada nóculo representa uma classe e o disparo de um deles significa que o padrão atual pertence à classe correspondente a este nóculo.

Na primeira rede da fig. 17, o reconhecimento é possível graças ao nóculo intermediário. Ele só dispara se ambos os elementos do padrão de entrada forem



estimulados (ou seja, se sua resposta for 1). O disparo dos dois nódulos de entrada faz com que o nódulo de saída da rede não dispare (a resposta é 0). Na segunda rede, o padrão de entrada não está diretamente ligado ao nódulo de saída da rede. O nódulo intermediário inferior tem mais ou menos a mesma função do intermediário da rede anterior. Se ambos os elementos do padrão forem estimulados, então o nódulo de saída não dispara.

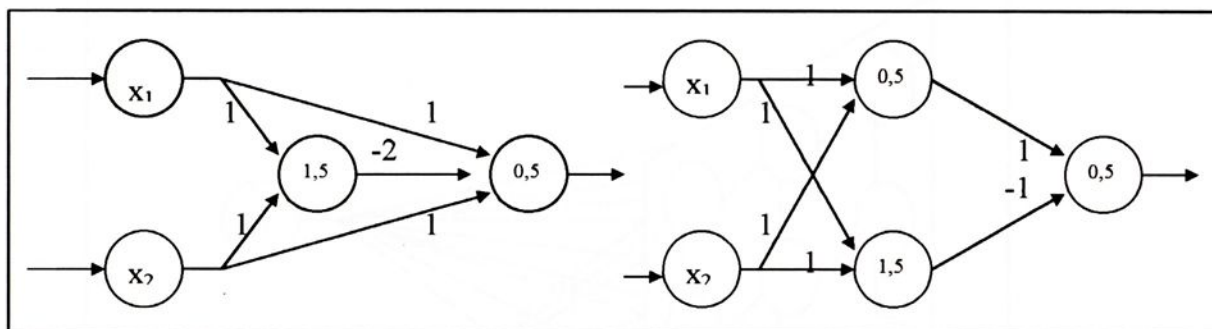


Fig 17 - Redes que resolvem o problema do ou-exclusivo

x_1 representa o primeiro elemento do padrão (A da fig. 16) e x_2 representa o segundo elemento (B da fig. 16). Neste caso, a resposta do neurônio é binária (0 ou 1)

O perceptron multicamadas proporcionou um avanço significativo no estudo sobre redes neurais. Porém, seu treinamento ainda é supervisionado, pois precisa da resposta desejada para aprender a reconhecer padrões. Um dos primeiros modelos que não têm este tipo de treinamento é o modelo de Kohonen. Nesta rede, o ajuste dos pesos depende apenas de seu padrão de entrada e dos próprios pesos de seus nódulos. Além disso, a rede é capaz de reconhecer padrões mais complexos do que o perceptron. Segundo Kohonen, seu modo de aprendizagem está mais assemelhado ao processo de aprendizagem humana.

4.3 A Rede de Kohonen

A rede de Kohonen, criada na década de 1980 por Kohonen (1987), é formada por duas camadas: a de entrada e a de Kohonen. Esta última é a própria camada de saída da rede, como mostrado na fig. 18.

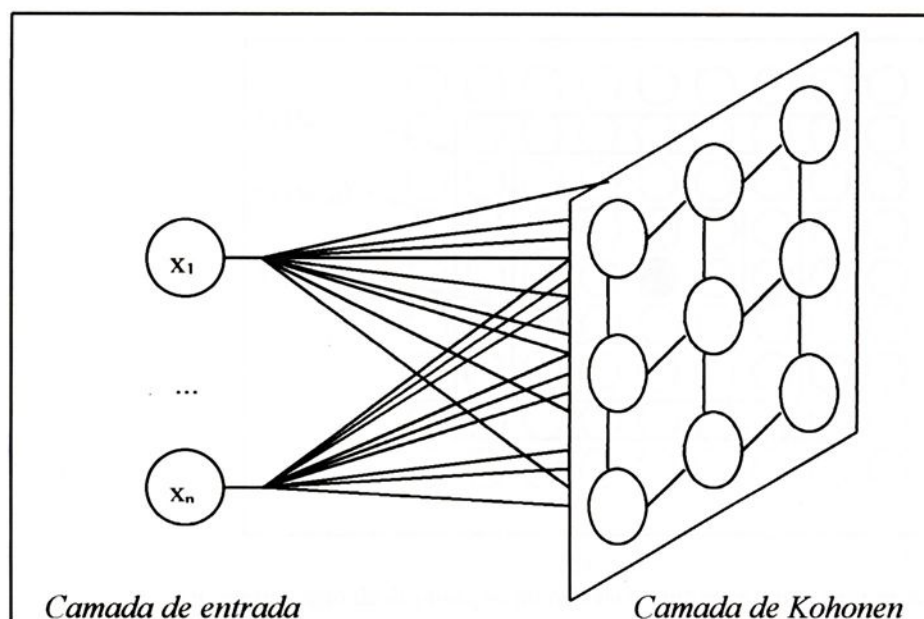


Fig.18 - Estrutura de uma rede de Kohonen

O funcionamento da rede pode ser assim explicado: primeiramente definimos os padrões de entrada e os pesos da rede e uma vizinhança para o nódulo vencedor. O *nódulo vencedor* é determinado pelo cálculo da menor distância euclidiana entre os vetores peso e de entrada dos nódulos. Ele será o responsável pelo reconhecimento de um certo padrão.

Os pesos iniciais serão aleatórios. Sugere-se, porém, que eles sejam dados de tal forma que sua distância com o padrão de entrada seja a menor possível. Como isso pode ser muito difícil e exigiria de certa forma um supervisor, definimos uma vizinhança para o neurônio vencedor. A *vizinhança* é a demarcação de um conjunto de neurônios próximos ao neurônio vencedor. Ela será responsável pelo reconhecimento mais rápido dos padrões. Além disso, a

presença da vizinhança na rede de Kohonen faz com que, se o nóculo responsável pelo reconhecimento de um padrão for danificado, outros vizinhos seus poderão reconhecê-lo, talvez com alguma deficiência. Tal característica, segundo Kohonen (1987), é também encontrada no cérebro humano. Observe-se a semelhança com as visões holista e localizacionista, citadas na seção 3.1 do primeiro capítulo.

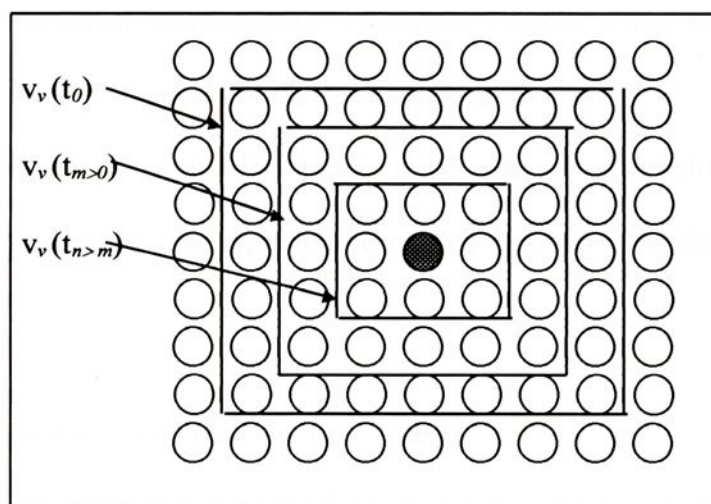


Fig. 19

Representação da diminuição do raio da vizinhança numa rede de Kohonen

Com o passar do tempo (um tempo representa uma mudança de peso dos nós da rede), o raio da vizinhança vai diminuindo. Esta diminuição é determinada por uma função de tempo, como mostrado na fig. 19 acima.

Para o treinamento da rede ser mais eficiente, Kohonen (1987) define uma constante de aprendizagem variando entre 0 e 1. Ao começar o treinamento, a constante pode ter um valor elevado, fazendo com que a distância entre o vetor peso e o vetor de entrada diminua mais rapidamente. Esta constante deve ser diminuída com o tempo. Isso porque a mudança dos pesos deve ser menor a cada momento, à medida que a distância vai diminuindo. A

redução do valor peso e da constante de aprendizagem devem ser feitos através de alguma função.

O vetor peso de um nóculo i no tempo t pode ser denominado por $\mathbf{w}_i(t)$. O vetor de entrada da rede é representado por $\mathbf{x}(t)$ e o vetor peso do nóculo vencedor por $\mathbf{w}_v$. No tempo inicial t_0 o vetor peso pode ser dado aleatoriamente à rede.

A função segundo a qual encontraremos o neurônio vencedor pode ser assim definida:

$$\| \mathbf{x}(t) - \mathbf{w}_v(t) \| = \min_i (\| \mathbf{x}(t) - \mathbf{w}_i(t) \|)$$

$\min_i$ é a menor diferença entre o valor de entrada e o valor peso do neurônio i .

A função que regulará a mudança dos pesos dos neurônios da vizinhança será a seguinte:

$$\mathbf{w}_i(t+1) = \begin{cases} \mathbf{w}_i(t) + \alpha(t) (\mathbf{x}(t) - \mathbf{w}_i(t)) & \text{se } i \in v_v \\ \mathbf{w}_i(t) & \text{se } i \notin v_v \end{cases}$$

No processo de treinamento, os nóculos que têm seus pesos modificados serão apenas os pertencentes à vizinhança. A rede tende a diminuir a distância entre o vetor peso e o vetor input dos nóculos mais próximos ao vencedor e aumentar os mais distantes. Este processo é chamado de inibição lateral.

Ao treinar a rede, o objetivo a alcançar é encontrar a menor distância euclidiana entre o vetor peso de algum nóculo e o vetor de entrada atual. Quanto menor for esta distância, melhor a rede identifica o padrão. Ao final do processo, o único nóculo a disparar é o vencedor.

Se fizermos uns poucos cálculos, notaremos que isso realmente acontece. A maior ou menor rapidez da convergência (diferença inexistente entre o vetor peso de um nóculo e o

vetor de entrada, quando, idealmente, ocorre o reconhecimento do padrão) dependerá da função de aprendizagem $\alpha(t)$. Quanto maior o seu valor, mais rápida é a possibilidade de convergência da rede.

Na função acima, não temos, aparentemente, nenhuma necessidade da presença de um valor desejado. Por isso, a rede de Kohonen é capaz de auto-organização. Porém, o vetor de entrada já parece ser uma espécie de vetor alvo. Isso porque o objetivo é fazer com que a rede obtenha a menor distância entre o vetor peso e o de entrada. Ou seja, o vetor de entrada é a própria resposta desejada que deve ser tomada como vetor de saída. Ainda assim, afirmam conexionistas como o próprio Kohonen (1987), não existe uma regra explícita pela qual podemos explicar a organização da rede.

O processo de treinamento da rede é dado pelo seguinte algoritmo:

1. Inicialize a rede

Defina $w_{ij}(t)$ ($0 \leq i \leq n-1$) como sendo o peso a partir do *input* i para o neurônio j no tempo t . Inicialize os pesos a partir dos n dados de entrada para os neurônios para os menores valores randômicos. Defina o raio inicial da vizinhança ao redor do neurônio j , $N_j(0)$, o mais amplo possível.

2. Apresente o *input*

Apresente o *input* $x_0(t)$, $x_1(t)$, $x_2(t)$, $x_{n-1}(t)$, onde $x_i(t)$ é o *input* para o neurônio i no tempo t .

3. Calcule as distâncias

Compute a distância d_j entre o *input* e cada neurônio de *output* j , dado por

$$d_j = \sum_{i=0}^{n-1} (x_i(t) - w_{ij}(t))^2$$

4. Selecione a distância mínima

Designe o neurônio de *output* com o mínimo d_j como j^* .

Atualize os pesos para o neurônio j^* e seus vizinhos, definidos pela vizinhança como $N_{j^*}(t)$. Os novos pesos serão

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)(x_i(t) - w_{ij}(t))$$

para j em $N_{j^*}(t)$, $0 \leq i \leq n-1$

O termo $\eta(t)$ é um termo benefício ($0 < \eta(t) < 1$) que reduz no tempo, tão lento quanto a adaptação do peso. Note que a vizinhança $N_{j^*}(t)$ diminui com o passar do tempo, localizando assim a área de atividade máxima.

6. Repita o processo voltando ao passo 2. (Beale e Jackson, 1990, p. 111)

Como dissemos, a rede de Kohonen é capaz de reconhecer padrões complexos, como imagens obscuras ou deformadas. Por exemplo, se um padrão é apresentado à rede com algumas características a mais ou a menos do objeto representado, ela é capaz mesmo assim de classificá-lo corretamente.

Segundo os conexionistas, a vantagem das redes sobre a máquina de Turing é que aquela não precisa ter um conjunto de regras bem determinadas para produzir uma resposta. A rede é capaz de aprender a reconhecer padrões ajustando as conexões entre seus nódulos. Ao ser acessado um dado de entrada, a própria rede procura adaptar-se para produzir uma resposta correta, sem a necessidade de um programa coordenando seus passos. Nesta adaptação, seus erros são tomados como um aliado e não como um problema. Se a máquina de Turing erra uma resposta para um dado qualquer, o programador deve rever o programa da máquina e modificar as instruções de tal forma que ela produza a resposta correta. Nas redes, o erro é tomado como um auxílio para que a própria rede aprenda a reconhecer corretamente um padrão. Desse modo, não há a necessidade de qualquer mudança nas regras ou na própria rede, afirmam os conexionistas. Além disso, uma mesma rede pode resolver uma série de problemas. Para fazer uma analogia, num momento ela pode estar reconhecendo um conjunto de maçãs e a seguir pode aprender a reconhecer o conjunto dos números pares. Os cientistas da IA, porém, afirmam que a máquina de Turing Universal também pode resolver diversos problemas. Ela computa qualquer função computável.

É difícil saber qual é realmente o melhor modelo da mente, se a máquina de Turing ou a rede neural. Isso porque o pretendo objeto de pesquisa neste caso nos é desconhecido. Não sabemos de fato o que é ou como é caracterizada a mente. Por isso, como podemos ter certeza de que as máquinas simulam e explicam estados e faculdades mentais? Que os



processos pelos quais elas passam na resolução de um problema (seja de natureza lógico-matemática ou reconhecimento de padrões) são os mesmos pelos quais passa a mente humana?

Embora tenham suas características próprias, tanto IA quanto RNA procuram criar sistemas capazes de simular e explicar alguns processos cognitivos embasados no mecanicismo. O comportamento inteligente, em ambas as correntes, é entendido como um processo mecânico, cuja estrutura pode ser simulada e explicada através do uso de modelos como os expostos nos últimos dois capítulos

Tendo conhecido o modo de funcionamento de alguns modelos e como eles se “comportam”, nosso objetivo, na próxima parte, é apresentar, em primeiro lugar, uma visão geral do conceito de inteligência (capítulo 4). Em segundo lugar, analisar como cada uma das duas correntes da Ciência Cognitiva pesquisadas neste trabalho caracteriza o conceito de inteligência (capítulo 5). Levando em consideração a apresentação feita sobre os modelos da mente, podemos perceber melhor as características que um comportamento de um sistema precisa ter para ser considerado inteligente. Por fim, expomos alguns argumentos a favor e contra a possibilidade dos modelos da mente poderem ser ou não considerados inteligentes (capítulo 6).



Parte II

Cap. 4

O conceito de Inteligência

1 Apresentação

Muitas caracterizações já foram construídas para o conceito de inteligência. É desde a antiguidade que filósofos e cientistas em geral procuram definir este conceito. Sua caracterização, no senso comum, apresenta algumas vezes paradoxos e imprecisões. No plano científico, tais paradoxos e imprecisões são amenizados, mas não eliminados. Algumas vezes, as confusões, ambigüidades, imprecisões e diferenças entre as caracterizações deste conceito parecem ser acentuadas. Uma das primeiras áreas de pesquisa deste século a analisar o conceito de inteligência foi a Psicologia e seu trabalho teve algumas implicações no estudo deste objeto de pesquisa na Ciência Cognitiva.

Neste capítulo, apresentamos uma visão geral do conceito de inteligência. Em seguida, expomos a análise feita pela Psicologia Cognitiva sobre o conceito em questão.



2 O conceito de inteligência em geral

O termo 'inteligência' deriva-se do latim *inputs* (interior ou entendimento) e *legere* (ler). Etimologicamente significa ler ou compreender no interior do entendimento. Tornar consciente (daí vem consciência, conhecimento ou ciência) o que nos rodeia.

Se fizéssemos uma pesquisa popular para saber o que as pessoas entendem por inteligência, talvez encontrássemos uma resposta para cada interrogado. As diferenças entre os pesquisados ficariam ainda mais claras e evidentes se a pesquisa fosse realizada em culturas diferentes. Nos Estados Unidos, por exemplo, de acordo com Berry (1984, p. 352), a agilidade e rapidez fazem parte da cultura. Neste país, o sujeito inteligente é aquele capaz de resolver problemas rapidamente. Em outros lugares, como em Uganda, este não é um sinal de inteligência. Na tribo rural de Baganda, seus habitantes pensam na inteligência como lenta, cuidadosa, ativa e sadia. Em Zimbábue, um sujeito inteligente é a pessoa que exercita a prudência e a cautela, especialmente nas interações sociais. Em algumas culturas como a tribo africana Pulowat, seus nativos acreditam que ficamos mais inteligentes à medida que envelhecemos. Isto porque adquirimos novas experiências, conhecimentos e habilidades.

Como dissemos, algumas vezes a diferença entre a caracterização da inteligência varia entre pessoas de uma mesma comunidade, pertencentes à mesma cultura. Por exemplo, ao perguntarmos às pessoas de um mesmo local o que elas entendem por inteligência, encontramos pelo menos dois tipos de respostas: o primeiro identifica o ser inteligente com o sujeito que tem pouco conhecimento e ainda assim consegue ser criativo. Desse modo, precisa pensar muito. O segundo tipo de respostas iguala o sujeito inteligente



com o que conhece muito, como os grandes intelectuais, por exemplo, e podem ser criativos sem precisar pensar muito. Notamos aí um paradoxo nestas duas respostas. Por um lado, é inteligente o sujeito que conhece pouco e, por outro também o é aquele que conhece muito. Desse modo, a inteligência pressupõe e não pressupõe o conhecimento (Gregory, 1997, p 19).

É certo que o indivíduo pode ter muito conhecimento e mesmo assim não exibir inteligência. Pode acontecer, por exemplo, dele apenas ser capaz de repetir os conhecimentos obtidos, sem nenhuma novidade. Por outro lado, notamos que é necessária a presença de conhecimento no indivíduo inteligente. Desse modo, podemos nos perguntar então quão relevante é o grau de conhecimento para a caracterização da inteligência. Como veremos no próximo capítulo, a quantidade limitada de conhecimento impossibilita certas máquinas de realizarem comportamentos de um modo satisfatório.

As caracterizações de senso comum da inteligência, pelo fato de serem populares, não científicas, não têm a pretensão de ser rigorosas, claras e distintas. Essas características, ao contrário, são pré-requisitos na ciência para que uma teoria seja bem aceita. No plano teórico, devemos procurar explicar o que nos propomos, de um modo claro, preciso, compreensível.

O estudo científico sobre o conceito de inteligência pode ser encontrado em tempos antigos. Para Descartes (1996b), e para os cartesianos em geral, a inteligência é uma propriedade da alma. Ela, assim como as demais propriedades mentais, não pode ser explicada em termos puramente físicos. Tampouco pode ser atribuída a outras entidades não possuidoras da alma humana. Segundo Khalfá (1997, p. 7-8),

Os 'dualistas' postulavam que a inteligência era uma faculdade exclusiva dos seres regradados por uma substância imaterial, um espírito, do qual uma das mais claras manifestações no mundo era



o comando da fala - ou melhor, da linguagem, pois os papagaios falam, mas como disse Descartes, o que eles dizem não é *à propos*, (ou seja, ao mesmo tempo apropriado a circunstâncias particulares e denotando algo), a não ser por acaso. É apenas uma reação física adquirida de modificações em seu ambiente, que podem não ter nada a ver com o significado do que eles 'dizem'.

Nessa visão, o ser inteligente precisa saber o que está fazendo, ser responsável por sua ação. Isso significa ser capaz de prever as conseqüências de seu próprio curso de ação. Implica, ainda, em perceber a relevância dos princípios morais, sociais ou legais envolvidos nessas conseqüências, quer se decida ou não por realizar a ação, que não pode ser por mero acaso, instintiva ou condicionada para ser inteligente. Como afirma Ryle (1949, p. 29),

Este título [inteligência] reservamos para as pessoas responsáveis por suas ações. Ser inteligente não consiste em satisfazer critérios, mas também em aplicá-los, em regular as próprias ações e não estar simplesmente bem regulado. A ação de uma pessoa é considerada cuidadosa ou habilidosa se em sua execução é capaz de detectar e corrigir erros, de repetir e melhorar êxitos, de aproveitar o exemplo dos outros, etc. Aplicará critérios ao atuar criticamente, isto é, ao tratar de fazer as coisas corretamente.

Desse modo, um relógio bem regulado não pode ser considerado inteligente. Ele simplesmente "segue" leis físicas, não procura melhorar sua ação, não corrige erros, etc. Embora Ryle não esteja defendendo argumentos dualistas, tampouco concorda com os funcionalistas, segundo os quais algumas máquinas são capazes de satisfazer os critérios de inteligência.

Para tentar eliminar os paradoxos, incoerências e controvérsias encontrados a respeito do conceito de inteligência, surge neste século, um interesse em várias áreas de pesquisa buscando tratá-lo de um modo "racional". Porém, ainda hoje paradoxos sobre este conceito permanecem na ciência. Não há consenso sobre sua definição e muitas vezes encontramos discordâncias graves (algumas vezes até contradições) entre uma definição e outra. Um exemplo disto pode ser encontrado em uma mesma área de pesquisa. Na antropologia, por exemplo,

... pesquisadores como Levi-Strauss, por exemplo, afirmam que existe um único tipo de inteligência humana, que se expressa diferentemente em diferentes culturas. Por outro lado, Berry afirma que a inteligência difere de cultura para cultura e não existe um tipo único. (Gardner, 1998, p. 19)

Na Ciência Cognitiva, cientistas como Minsky (1967) e Putnam (1967) acreditam que comportamentos inteligentes podem ser simulados em computadores digitais. Por outro lado, conexionistas como Caudill e Butler (1992) acreditam que eles são melhor simulados por redes neurais artificiais.

Em um recente trabalho que trata exclusivamente do conceito de inteligência, organizado por Khalfa (1997) e denominado *A natureza da inteligência*, encontramos inúmeras caracterizações deste termo. A primeira delas é a apresentada pelo próprio organizador do livro. Segundo Khalfa (1997), um modo de poder identificar a inteligência é por meio da construção de ferramentas. Para saber se o sujeito é ou possui inteligência, observe sua capacidade de utilizar objetos para sua sobrevivência ou adaptação ao meio ambiente. A construção de ferramentas, segundo Khalfa (1997, p. 8), implica na

... capacidade de dotar um objeto material de um modo de existência determinado por um objetivo muito pouco relacionado com a natureza e origem do objeto (como quando alguém parte um galho para utilizá-lo como bengala).

Este modo de identificação pode dotar de inteligência outras espécies de animais além do homem. Foram feitas experiências com babuínos que, para retirar cupins de suas tocas, utilizavam galhos de árvore. Com isto, construíam uma ferramenta com o galho, dando-lhe um objetivo muito distinto de sua natureza.

A quem pode ser atribuída inteligência é uma das primeiras controvérsias a respeito desse conceito. Muitos cientistas, como a maioria dos teólogos, pensam que apenas o homem é dotado de inteligência. Por outro lado, pesquisadores como Turing (1950), Putnam (1967), Newell & Simon (1961) acreditam que tal propriedade não é



exclusivamente humana. Para estes, é possível a construção de sistemas artificiais inteligentes. Para isto, tal sistema deve ter pelo menos a capacidade de resolver problemas adequados.

Gregory (1997) afirma existirem dois tipos de inteligência: a *potencial* e a *cinética*. A primeira diz respeito à quantidade de conhecimento que a entidade possui. A segunda refere-se à capacidade de resolução de problemas, à geração de novidade apropriada. Porém, os dois tipos se completam, afirma o pesquisador em questão. Para resolver problemas mais adequada e facilmente é preciso uma certa quantidade de conhecimentos. Por outro lado, para obter novos conhecimentos, é preciso que o indivíduo seja capaz de resolver certos problemas.

De acordo com Gregory (1997), a inteligência potencial não diz respeito exclusivamente ao conhecimento adquirido intelectualmente. Ao contrário, envolve conhecimento inato, geralmente inconsciente, como por exemplo o fato de eu “saber” usar meu aparelho visual para enxergar. Gregory, a partir desta definição, atribui inteligência cinética a tesouras por exemplo, uma vez que elas “armazenam” conhecimento sobre como cortar roupas. O processo evolutivo, por outro lado, pode ter uma inteligência cinética, uma vez que resolve problemas complicados como o da manutenção das espécies no planeta. É um pouco estranho atribuir tais características a tesouras, como o permite Gregory. Isso porque ela não armazena conhecimento sem uma atribuição de sua utilidade dada pelo ser humano. Lembro-me nesse caso do filme *Os deuses devem estar loucos* onde uma tribo indígena deu toda utilidade possível a uma garrafa de coca-cola, menos a de ser uma garrafa de coca-cola. A respeito do processo de evolução, parece claro que ele não é uma entidade real, que realiza comportamentos por si só. Ao contrário, ele é uma

consequência de comportamentos realizados pelos próprios seres em evolução. Neste sentido, não parece ser apropriado atribuir inteligência a conceitos abstratos como esse, mas sim aos indivíduos agrupados sobre tal conceito. Podemos fazer uma analogia neste caso com o conceito de universidade: quando seus alunos geram alguma novidade espetacular, não dizemos que a universidade é inteligente, mas sim que seus alunos o são.

Mackintosh (1997) afirma que a inteligência consiste em compreender o mundo e não apenas em percebê-lo. Consiste ainda em prever o futuro, adequar-se à mudança das circunstâncias, resolver problemas, extrair inferências, raciocinar, etc. Por outro lado, Butterworth (1997) afirma que a inteligência é a faculdade ou capacidade de conhecer. Ela se expressa no comportamento adaptativo, em contextos particulares.

Segundo Schank & Birnbaum (1997), a inteligência pode ser aumentada de acordo com o grau de conhecimento e experiências possuídos pelo sujeito. Para eles, pode-se

... construir entidades inteligentes analisando em que consiste o comportamento inteligente, determinando as regras que governam esse comportamento e implementando tais regras em uma máquina. A inteligência, desse ponto de vista, é modificável. As entidades se tornam mais inteligentes se se puder encontrar maneiras de preenchê-las com mais conteúdo inteligente. (Schank e Birnbaum, 1997, p. 77-8)

Se continuássemos resumindo cada capítulo do livro de Khalfa (1997), em cada um deles encontraríamos uma caracterização diferente (e incompleta) do conceito. Isto mostra o quanto estamos distantes de uma boa definição de inteligência no plano científico.

Pinker (1998, p. 73) afirma que a inteligência é "... a capacidade de atingir objetivos diante de obstáculos, por meio de decisões baseadas em regras racionais (que obedecem à verdade)."

Para explicar um comportamento inteligente, Pinker (1998) compara o comportamento de Romeu frente à Julieta com o da limalha frente ao ímã: Romeu quer

Julietta como a limalha quer o ímã. Se não houvesse nenhum obstáculo, Romeu chegaria até Julietta numa linha tão reta quanto a limalha ao ímã. O comportamento inteligente em ambos poderia ser analisado quando obstáculos fossem postos entre os objetos de encontro. Se um muro fosse posto entre Romeu e Julietta, estes não ficariam com seus rostos colados um em cada lado dele, como aconteceria com a limalha e o ímã. Romeu encontraria um modo de atingir seu objetivo, que é o de ficar face a face com sua amada. Poderia fazê-lo, por exemplo, pulando o muro, e não tentando dar cabeçadas para derrubá-lo. Segundo Pinker, alcançar o objetivo, passando pelos obstáculos de um modo racional é o que caracterizaria o comportamento do indivíduo inteligente.

De acordo com esta caracterização, seria possível atribuir inteligência às máquinas e considerá-las bons modelos da mente. Isto porque, para Pinker (1998, p. 10), "... a mente é um sistema de órgãos de computação que a seleção natural projetou para resolver os problemas enfrentados por nossos ancestrais evolutivos em sua vida de coletores de alimentos."

Ainda que o computador possa ser considerado um modelo da mente, como acreditam muitos pesquisadores da IA, não podemos dizer que ele tem uma mente do mesmo modo que afirmamos no caso do ser humano. Neste, ela originou-se através de um longo processo evolutivo e não pode ser criada de uma hora para outra. Sendo assim, embora possuam as principais características para simular os órgãos de computação da mente, os órgãos de computação do computador são muito distintos dos da mente humana. Por isso, precisamos ter cuidado quando simulamos aspectos da mente como algumas de suas faculdades ou estados em computadores. Simular estados e faculdades em máquinas não significa que elas realmente os possuam (Searle, 1980).



Uma das primeiras tentativas de teorização no tocante à inteligência neste século encontra-se na Psicologia. A análise da inteligência, nesta ciência, estava relacionada com a aplicação de uma bateria de testes. Eles serviam para medir o quociente de inteligência do indivíduo.

3 O conceito de Inteligência na Psicologia Cognitiva

Muitas vezes, seja no senso comum ou na ciência, a inteligência está relacionada com a capacidade de resolução de problemas. Estes podem ser do tipo social, prático, abstrato ou de outra natureza. Um exemplo disto é a Psicologia, onde a resolução de problemas, principalmente de natureza abstrata, está intimamente ligada à inteligência. Segundo o psicólogo Sternberg (1986, p. 51), por exemplo, "...a inteligência envolve a capacidade de solucionar problemas abstratos." Tal característica é fundamentalmente encontrada em psicólogos cuja tentativa é medir a inteligência através de testes. Isto significa que a inteligência poderia ser quantificada de acordo com a capacidade do indivíduo de resolver uma série de questões.

Um dos primeiros a criar um *teste de inteligência* foi Lewis Terman (1921). Segundo Terman (1921, p. 128), "...um indivíduo é inteligente na medida em que é capaz de pensar em termos abstratos." Isto significa ser capaz de resolver tal tipo de problemas. Terman criou um teste para verificar qual o grau de capacidade do sujeito para resolvê-los. O Quociente de Inteligência (QI) deste indivíduo seria equivalente ao resultado obtido por ele no teste.

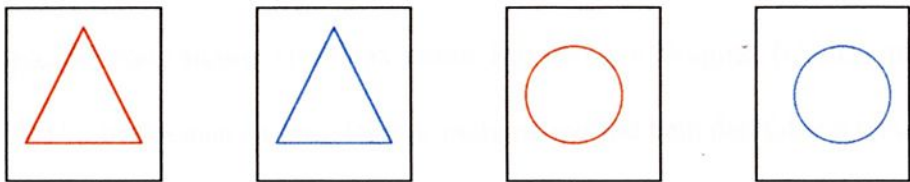


Para psicólogos como Terman (1921), Boring (1950) e Sternberg (1986), a inteligência é medida a partir de um conjunto de testes. Segundo Boring (1950), “... a inteligência é aquilo que os testes [de inteligência] testam.” Ela é resultado da aplicação destes testes segundo os quais dizemos que determinado indivíduo tem um QI de quantidade n . Este quociente pode ser medido de um modo determinado. A partir deste ponto de vista, o sujeito será mais ou menos inteligente dependendo de sua atuação nos testes.

Como é de se supor, a maioria das questões dos testes de inteligência estão relacionados à resolução de problemas do tipo lógico-matemático ou, de alguma forma, abstratos. Um exemplo de um problema deste tipo, exposto em um teste proposto por Wason (1969, p. 471), é o seguinte:

Dada a seguinte informação, resolva a seguinte questão:

Informação: Cada cartão abaixo tem um triângulo num lado e um círculo no outro.



Questão: Quais cartões você precisa virar para saber se a seguinte sentença é verdadeira: cada cartão com um triângulo vermelho de um lado tem um círculo azul no outro?

Fig.20 - Exemplo de um problema de um teste de inteligência.

Notamos, no que dissemos acima, que os psicólogos envolvidos na criação de testes de inteligência não estavam preocupados fundamentalmente com a definição do conceito de inteligência. Pretendiam, antes, criar testes para medir o grau de inteligência das pessoas.

Os testes de QI não tiveram grande aceitação pela comunidade científica na caracterização da inteligência. Uma das primeiras críticas direcionadas a eles é o fato de

ressaltarem primordialmente o aspecto lógico-matemático do sujeito. Neste sentido, pelo tipo de problemas proposto nos testes, o matemático teria muito mais possibilidade de apresentar um QI altamente elevado. Isto porque ele está familiarizado com tais problemas, muito mais do que um tenista, por exemplo. Porém, seria adequado unir numa mesma sala excelentes matemáticos, músicos, físicos, pedagogos, educadores, filósofos, políticos, tenistas e aplicar-lhes um mesmo teste para avaliar seu quociente de inteligência? Os testes de inteligência valorizavam muito o conhecimento cristalizado, ou seja, informações que já foram adquiridas pelo indivíduo, principalmente pelo matemático (Horn, 1985).

Uma crítica semelhante à apresentada acima diz respeito ao fato dos testes não revelarem com clareza a capacidade do indivíduo de resolver problemas em sua vida cotidiana (Ceci, 1990). Problemas como aprender a andar ou a consertar uma bicicleta, relacionar-se com outras pessoas, não são analisados nos testes. Eles, ao contrário, priorizam aspectos lógico-matemáticos.

De acordo com alguns cientistas como Sternberg e Wagner (citados por Gardner, 1998, p. 306) "...problemas de tipo lógico- matemático são bem definidos e tem apenas uma resposta e um método básico de solução." Eles requerem pouca informação além da que é apresentada no problema. Tampouco estão ligados à experiência cotidiana e em geral não são intrinsecamente interessantes, afirmam os pesquisadores acima. No mundo real, porém, nós poucas vezes nos deparamos com problemas deste tipo.

Os testes de QI, quando priorizam aspectos lógico-matemáticos, revelam o quociente de inteligência do indivíduo no tocante à capacidade de resolução de problemas nessa lógico-matemáticos. No que diz respeito à capacidades que envolvem características como



experiências práticas e informações imprecisas para a resolução de problemas, os testes nos auxiliam muito pouco.

Alguns dos testes de inteligência foram criados por motivos humanitários. Binet & Simon (1916), por exemplo, criaram um teste a pedido do governo Francês com o objetivo de avaliar quais crianças tinham dificuldade de aprendizagem. Com isto, poderiam direcionar uma maior atenção a elas. Ainda que alguns tivessem estes objetivos, uma das metas principais dos psicólogos em geral era a de criar um método para poder avaliar e comparar a inteligência dos indivíduos. O grande problema é que, ao comparar indivíduos de uma mesma cultura, aspectos fundamentais como biológicos são deixados de lado, segundo Ceci (1990). Ao comparar indivíduos de culturas distintas, além dos biológicos, aspectos como culturais, sociais e ambientais são desconsiderados (Ceci, 1990; Horn, 1985). É claro que cada cultura, cada povo têm suas particularidades e nem todos podem ser submetidos a um mesmo teste, com as mesmas questões.

Demos, até o momento, uma caracterização geral do conceito de inteligência. No próximo capítulo, analisaremos tal conceito a partir de um ponto de vista da Ciência Cognitiva em duas de suas correntes, quais sejam, a IA e o Conexionismo.



Cap. 5

A inteligência na Ciência Cognitiva

1 Apresentação

A Ciência Cognitiva em duas de suas principais correntes esteve até o momento mais preocupada com a simulação cognitiva do que propriamente com a simulação e estudo de estados e faculdades mentais em geral. Tanto a IA quanto o Conexionismo inclinaram-se mais ao estudo de processos implicados no raciocínio e pensamento humanos do que em desejos, crenças, sentimentos. O objetivo da maioria dos cientistas cognitivos é, em última análise, simular comportamentos que, se fossem realizados por seres humanos seriam inequivocamente considerados inteligentes. Enquanto a IA visou simular comportamentos que envolvessem a resolução de problemas, em sua grande maioria, lógico-matemáticos, o Conexionismo procurou investigar problemas do tipo reconhecimento de padrões.

As máquinas, segundo alguns de seus construtores como Turing (1950), Newell, Shaw & Simon (1958) e Minsky (1970), simulam comportamentos humanos inteligentes. Por isso, podem ser dotadas de inteligência. Por outro lado, cientistas como Hopfield (1982) acreditam que, embora simulem comportamentos humanos inteligentes, nem por isso suas



máquinas são literalmente inteligentes. Veremos, neste capítulo, o que tais cientistas entendem por inteligência.

2 O conceito de inteligência na Inteligência Artificial

Um dos objetivos principais da IA com respeito à inteligência é construir modelos computacionais capazes de simular comportamentos que, se fossem realizados por seres humanos seriam inequivocamente considerados inteligentes. Segundo Feigenbaum & Feldman (1968, p. 3), o objetivo dos cientistas da IA é "... construir programas de computador que exibem comportamentos que são chamados de inteligentes quando observados em seres humanos." Dentre estes cientistas encontramos Newell, Shaw & Simon (1958), Putnam (1967), Minsky (1967).

Para alguns pesquisadores da IA, como os citados acima, a inteligência é uma questão de aprendizado, de aquisição de memória ou conhecimento-base de uma extensão suficiente, e de desenvolver os mecanismos de recuperação necessários para usá-lo (Schank & Birnbaum, 1997). Para ser mais inteligente, um sistema deve aumentar seu conhecimento, proporcionado pela introdução de programas mais complexos. Para estes cientistas, é possível a construção de entidades inteligentes fazendo com que elas realizem comportamentos inteligentes. Segundo Schank & Birnbaum (1997, p. 78),

... nós e outros cientistas 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.

A idéia destes cientistas é que, comportamentos inteligentes podem ser explicados de um modo mecânico. Por isso, os programas da IA, acreditam seus criadores, além de

simular comportamentos humanos inteligentes, explicam os processos pelos quais o homem passa na resolução de problemas.

Na IA, um comportamento inteligente é aquele cujo resultado envolve a capacidade de resolução de um problema de um melhor modo possível. O comportamento inteligente,

... seja do homem, seja da máquina, será aquele que, dada determinada situação, precisa escolher a possibilidade correta para ela. Deste modo, para a máquina agir inteligentemente, precisa pesquisar as diversas incertezas do problema de um modo altamente seletivo, explorando caminhos relativamente férteis, com soluções, e ignorando caminhos relativamente estéreis. (Feigenbaum e Feldman, 1968, p. 6)

Pesquisadores como Newell, Shaw & Simon (1958) procuraram simular processos cognitivos segundo os quais poderia ser mais certa a possibilidade de estudo. Estes processos eram os referentes ao pensamento lógico-matemático. Segundo os pesquisadores acima referidos, o pensamento, enquanto relacionado aos processos cognitivos, poderia ser explicado mecanicamente através de programas computacionais. Críticos da IA como Penrose (1993), por exemplo, não concordam que o pensamento, em especial o matemático, possa ser simulado computacionalmente. Segundo Penrose (1993, p. 128),

O pensamento matemático não pode ser descrito computacionalmente porque nele estão contidas crenças, intuição, compreensão, sutileza, talento artístico etc... o pensamento matemático não pode ser reduzido ao cálculo cego, à pura manipulação de símbolos.

No pensamento encontramos intuição, compreensão, bom senso, elementos estes que, provavelmente não podem ser formalizados, acredita Penrose. Uma máquina, por exemplo, apenas segue regras e manipula símbolos, sem compreender o que faz. O pensamento matemático, ao contrário, requer uma boa dose de entendimento, sutileza e mesmo talento artístico. Ainda que em algumas vezes ele possa ser reduzido a um cálculo cego, onde o computador tem muito mais vantagens do que o ser humano, o pensamento em geral não pode ser computacionalmente simulado, afirma Penrose (1994).



Para Penrose (1993, p. 112), a crítica ao estudo do pensamento matemático estende-se aos outros tipos de pensamento, uma vez que

... não há nada de essencial que separe o matemático de outros tipos de pensamento, de modo que a nossa demonstração de que o entendimento matemático é algo que não pode ser simulado em termos computacionais pode ser concebida também como uma demonstração de que o próprio entendimento - um dos mais essenciais componentes da inteligência genuína - é algo que se situa além de qualquer tipo de atividade puramente computacional.

Mas, como são os modelos computacionais que, segundo cientistas da IA, afirmam manifestarem comportamentos inteligentes? Descreveremos a seguir alguns destes programas, comentando sua estrutura, funcionamento e tecendo críticas a respeito da hipótese de realmente se comportarem de modo inteligente.

A tentativa de construção de autômatos capazes de simular comportamentos humanos inteligentes é antiga. No século XVIII, Jacques de Vaucanson criou duas máquinas, uma tocadora de tambor e outra de flauta. No mesmo período, Pierre Jacquet-Droz criou um autômato semelhante a um menino capaz de escrever. No século XIX, Babbage criou um sistema mecânico para realizar cálculos. Pode-se dizer que Babbage conseguiu gastar muito dinheiro em sua máquina, sem porém, alcançar com grande sucesso seu objetivo, que era o de construir um sistema artificial capaz de realizar operações aritméticas (Hutten, 1968).

Foi neste século que a construção de máquinas capazes de realizar de um modo razoável comportamentos que antes eram de exclusividade humana concretizou-se. O melhor exemplo delas é o atual computador digital. Eles seguem instruções de programas como os criados por Newell, Shaw & Simon (1958) e Winograd (1972).

Já no início da IA, em 1956, alguns cientistas mostraram programas que haviam construído para simular aspectos da inteligência. Newell, Shaw & Simon (1958), por



exemplo, criaram um programa chamado *Logic Theorist*, capaz de demonstrar teoremas lógicos. Posteriormente, Newell & Simon (1972) criaram o *General Problem Solver* (GPS), que simulava a resolução de tipos mais complexos de problemas do que os lógicos.¹⁵

Newell e Simon estavam interessados na simulação de processos cognitivos humanos, que consiste em uma série de passos: em primeiro lugar, deve-se escolher a tarefa que se quer simular. Depois, analisar como o ser humano a realiza (observando seu comportamento, perguntando o que ele faz quando realiza a tarefa, etc.). Feito isso, começa a construção de seu modelo¹⁶. Terminada a construção, deve testar o programa e comparar os passos do modelo com os do ser humano até que eles sejam semelhantes. Isto será feito através da análise do comportamento do ser humano.

O objetivo inicial de Newell e Simon era avaliar como indivíduos em particular resolvem problemas. Cada programa, de fato, simula um indivíduo em particular, de acordo com seu protocolo (conjunto de sentenças que revelam os processos pelos quais um indivíduo passa na resolução de um problema). Depois de muitos programas construídos seria possível, segundo eles, criar uma teoria geral da solução de problemas.

O método que adotaram para a construção de seus programas consistia em apresentar o problema ao indivíduo cujo comportamento seria simulado. Após esta apresentação, o trabalho caracteriza-se pela anotação de todos os processos pelos quais o indivíduo passa na resolução deste problema. Os processos seriam identificados pelo protocolo do ser humano

¹⁵ Falaremos muito neste e nos próximos capítulos que um programa simula comportamento, computa funções, realiza comportamentos. Esta é uma característica encontrada na Ciência Cognitiva. Porém, lembramos que é necessário algo a mais do que um simples conjunto de instruções para realizar qualquer coisa. Como dissemos no segundo capítulo, é necessário ainda um agente e uma memória para isto.

em análise. Uma vez terminado o problema, o objetivo é criar um programa que execute exatamente os mesmos comportamentos do indivíduo.

Este processo adotado pelos pesquisadores acima é o seguido pelos cientistas da IA de um modo geral. O comportamento inteligente, para eles, pode ser entendido como uma sequência de processos elementares organizados de determinados modos. Segundo Newell e Simon (1972, p. 293),

... o que é mais importante não é que o programa dá uma compreensão exata do comportamento humano de solução de problemas, mas que eles finalmente revelam com grande clareza que o comportamento livre de um humano razoavelmente inteligente pode ser entendido como o produto de um complexo, mas finito e determinado conjunto de leis.

A primeira afirmação desta nota justifica-se pelas críticas dirigidas aos programas criados por Newell & Simon (1968), citados por eles mesmos. Dentre elas está exatamente o fato dos programas simplesmente manipularem símbolos e seguirem regras lógicas, deixando de lado características como as subjetivas do indivíduo. Além disso, os processos simulados são apenas aqueles comentados pelo sujeito no protocolo. Isto não garante o reconhecimento de todos os processos pelos quais o indivíduo passa ao resolver um problema. Por fim, nota-se no indivíduo a aplicação simultânea de regras (processamento em paralelo) enquanto que os programas são capazes apenas da aplicação sequencial delas (uma de cada vez).

Um sistema construído para mostrar que os programas de fato não apresentam comportamento inteligente foi o ELIZA, de Weizenbaum (1976). ELIZA foi construído para simular o comportamento de um psiquiatra rogeriano. O objetivo é fazer com que o

¹⁶ Um modelo para tais cientistas é a sequência ordenada dos processos básicos que um ser humano realizaria para fazer uma determinada tarefa. Em outras palavras, um modelo é um conjunto de instruções que simula um comportamento humano.

programa seja capaz de manter um diálogo inteligente com um suposto paciente. Isso implica em produzir respostas significativas a questões e afirmações do interlocutor.

Um exemplo de uma conversa entre ELIZA e um suposto paciente é o seguinte (as sentenças proferidas por ELIZA estão em maiúsculo)

Eu estou infeliz
 VOCÊ ACHA QUE VIR AQUI PODE AJUDÁ-LA A NÃO SER INFELIZ?
 Eu preciso de ajuda, isso é certo.
 O QUE SIGNIFICARIA PARA VOCÊ RECEBER AJUDA?
 Talvez eu pudesse aprender a me dar bem com minha mãe
 CONTE-ME MAIS SOBRE SUA FAMÍLIA.
 Minha mãe toma conta de mim.
 QUEM MAIS EM SUA FAMÍLIA TOMA CONTA DE VOCÊ
 Meu pai.
 SEU PAI?
 Você é parecida com meu pai de certas maneiras
 QUE SEMELHANÇAS VOCÊ VÊ?
 Você não é muito agressiva, mas acho que não quer que eu perceba isso.
 O QUE A FAZ PENSAR QUE EU NÃO SOU MUITO AGRESSIVA?
 (Weizenbaum, 1976, p. 3-4)

Como neste caso, ELIZA produz afirmações de tal modo que possa fazer com que o paciente os reconhecesse como significativos. Porém se analisarmos as instruções deste programa, notamos que o funcionamento de ELIZA caracteriza-se pela identificação de palavras que ela recebe. Por exemplo, uma instrução é a seguinte:

Se receber 'infeliz', randomize uma das três respostas:
 0: Você acha que vir aqui pode ajudá-lo a não ser infeliz?
 1: Fale-me mais sobre isto.
 2: Porque você está infeliz?

Isto significa que, quando o programa receber uma sentença com a palavra 'infeliz', vai selecionar uma destas três respostas, através de uma função determinada. O funcionamento de ELIZA é basicamente este.

Pelo fato de ELIZA funcionar seguindo regras como a acima citada e manipulando símbolos sem significado, muitos críticos, dentre os quais encontra-se o próprio



Weizenbaum (1976) afirmam que de fato tal programa não entende o que faz (Bodem, 1987). Porém, a compreensão do que se está realizando é um pré-requisito para um comportamento ser inteligente, afirmam alguns filósofos como Searle (1984). O programa simplesmente imprime respostas a partir de alguns chavões apresentados a ele e determinados em suas instruções. Segundo Bodem (1987, p. 156),

Eliza não tinha nenhum entendimento real de uma conversa em inglês, e nem conhecimento do conteúdo de qualquer assunto. Consequentemente, Eliza era incapaz de responder genuinamente até mesmo à mais simples das perguntas. Essa incapacidade era mascarada pela devolução de todas as perguntas com uma resposta como 'Porque você pergunta?'.

Exatamente por funcionar deste modo, em alguns diálogos ELIZA produz algumas sentenças sem sentido, idiotas. Este é o caso do seguinte diálogo:

Você acha que eu posso usar o computador hoje de manhã?
 POR QUE VOCÊ PERGUNTA?
 Eu poderia fazer algumas outras vendas *telcomp*.
 POR QUE VOCÊ NÃO TEM CERTEZA?
 Meus clientes em perspectiva nunca viram uma demonstração do sistema.
 ISSO SIGNIFICA MUITO PARA VOCÊ?
 Claro que sim.
 VOCÊ PARECE MUITO SEGURO DISSO.
 Isso é ridículo.
 O QUE QUER DIZER COM ISSO É RIDÍCULO?
 Há alguma razão para eu não poder usar o computador agora?
 CONTE-ME MAIS SOBRE SEUS CLIENTES EM PERSPECTIVA.
 (Bodem, 1987, p. 96)

Tentativas de melhora na qualidade do diálogo entre ELIZA e possíveis pacientes foram feitas aumentando o número de palavras-chave no programa. Porém, a estrutura fundamental do programa não é modificada, o que confirma a tese de Weizenbaum de que o programa não exhibe comportamentos inteligentes. Para quem observa de fora, algumas vezes os diálogos parecem ser inteligentes. Porém, se observarmos sua estrutura interna e modo de funcionamentos, notamos que de inteligente não existe muita coisa no comportamento do programa, como afirma Searle (1984).



Um terceiro programa construído para realizar comportamentos inteligentes foi criado por Bodrow, um estudante de Minsky. Tal programa, denominado Student, resolve problemas de álgebra de segundo grau. Seu funcionamento é semelhante ao de ELIZA. Determinados chavões são propostos para regular o comportamento do Student. Tais símbolos-chave são o que possibilitariam ao programa resolver os problemas.

Student é capaz de resolver problemas como o seguinte: um carro percorre 500 Km em 6 horas. Qual foi a velocidade média deste carro? O programa resolve este problema porque existe uma instrução dizendo que, 'se aparecer x KM em y horas, divida x por y '. O resultado deste cálculo será exatamente a resolução do problema.

Pelo fato da constituição fundamental do Student ser a mesma de ELIZA, as críticas referidas a este programa podem ser também aplicadas àquele. Além disso, pelo fato de apenas levar em consideração determinados símbolos, o Student resolve problemas como o seguinte: uma pedra flutua contra a corrente a 75 milhas por hora. Quantas horas demorará para percorrer 200 milhas? Student não é capaz de avaliar a natureza absurda de um problema como este. Falta-lhe o bom senso para poder fazer tal avaliação. Isto porque, como afirma Gardner (1998, p. 187), "... a resolução não implica no entendimento dos termos envolvidos no problema."

Muitos pesquisadores buscaram criar programas para realizar comportamentos inteligentes. Tais programas eram então literalmente realizados por máquinas como computadores digitais. Porém, alguns pesquisadores tentaram criar robôs que, de certa forma, podiam realizar comportamentos que envolvessem a movimentação física do sistema. Winograd (1972), por exemplo, criou um robô, denominado SHRDLU, capaz de responder perguntas formuladas em Inglês com soluções não-matemáticas.

Segundo Winograd (1972), SHRDLU percebe o que é e não é possível, como por exemplo saber que é impossível empilhar pirâmides ou estabelecer relações entre objetos num determinado ambiente. Por exemplo, no seguinte ambiente (fig. 21), SHRDLU é capaz de responder as seguintes questões:

O que a caixa contém?
 A PIRÂMIDE AZUL E O BLOCO AZUL
 O que está apoiando a pirâmide?
 A CAIXA
 Quantos blocos não estão na caixa?
 QUATRO DELES
 Algum deles é mais estreito do que aquele que lhe pedi pára apanhar?
 SIM, O CUBO VERMELHO.
 Ele está apoiado?
 SIM, PELA MESA
 A mesa pode apanhar blocos?
 NÃO.
 Uma pirâmide pode apoiar uma pirâmide?
 EU NÃO SEI.
 empilhe duas pirâmides
 EU NÃO CONSIGO. (Winograd, 1972, p. 9-10)

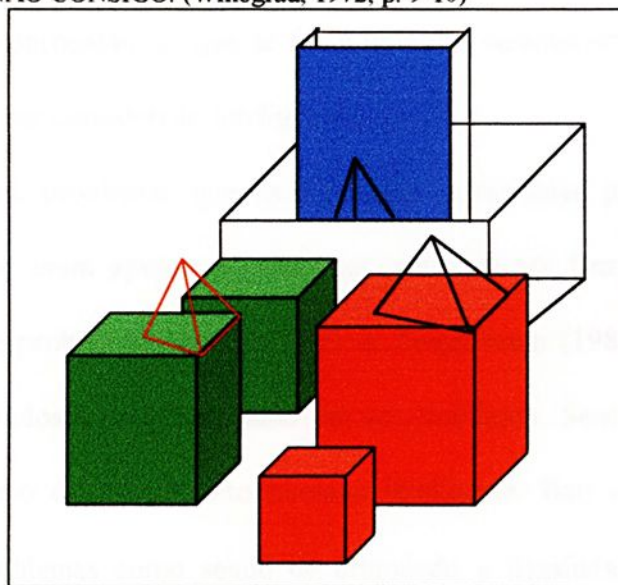


Fig. 21 - Ambiente no qual SHRDLU analisa a relação entre os objetos

O segredo, segundo Winograd (1972), é fazer com que o sistema possua o maior número de conhecimento possível sobre o ambiente. Isto significa introduzir no robô sentenças sobre este mundo. Ainda segundo tal pesquisador, o sistema, além de analisar as

relações entre os objetos, também pode reconhecer, como no caso acima, as regras semânticas para as relações entre palavras.

Segundo Winograd (1972), o conhecimento sobre o mundo permite a SHRDLU reconhecer que uma mesa apanhando um bloco é um absurdo, e, portanto, supor que a resposta é falsa. Além disso, SHRDLU pôde reconhecer, neste caso, que um bloco podia apoiar uma pirâmide sem verificar isso experimentalmente, mas não sabia se duas pirâmides podiam ser empilhadas. Quando o robô foi instruído para empilhar duas pirâmides, ele reconheceu que não era capaz de fazê-lo.

Como nos outros programas, se avaliarmos atentamente, a estrutura fundamental deste robô é determinada pela manipulação de símbolos e pelo seguimento de regras lógicas. Isso, porém, não é suficiente para a compreensão do que se está fazendo. Segundo Searle (1984), a compreensão do que se faz é uma das características fundamentais para um comportamento ser considerado inteligente.

A maioria dos problemas que as máquinas construídas pelos cientistas da IA conseguiam resolver eram apenas aqueles que poderiam ser bem definidos (Campbell, 1989). Este tipo de problemas, segundo Barr & Feigenbaum (1982) e Kohonen (1987), não são os encontrados pelo ser humano em seu cotidiano. Sendo assim, eles revelam muito pouco sobre o comportamento humano inteligente. Barr & Feigenbaum (1982) criticaram estes problemas como sendo de brinquedo e desafiaram os pesquisadores a trabalhar com problemas reais. Segundo estes cientistas,

Xadrez e lógica são problemas de brinquedo. Se vocês os resolverem, terão resolvido problemas de brinquedo... Saiam para o mundo real e resolvam problemas do mundo real. (Barr e Feigenbaum, citados em Gardner, 1998, p. 189).



Barr & Feigenbaum (1982), criaram um programa para tentar resolver problemas práticos. Seu programa, denominado Dendral, resolvia problemas específicos de espectrometria de massa.

Os pesquisadores da IA notaram que, ao construir seus programas, estes não obtinham muito sucesso devido ao pouco conhecimento possuído por eles. Por isso, tais cientistas passaram a acreditar que a quantidade de conhecimento é uma característica fundamental para resolver problemas complexos de modo mais prático. Para cientistas como Schank & Birnbaum (1997), o aumento do conhecimento produz também o aumento da inteligência. O que é preciso para fazer computadores inteligentes é dotá-los de conhecimento, afirmam. Não existe inteligência sem conhecimento real, imutável. Quanto mais conhecimento a máquina tiver, mais poderá resolver problemas e mais será inteligente. Segundo Schank & Birnbaum (1997, p. 106),

... a inteligência é uma função do conhecimento. Pode-se ter a potencialidade para a inteligência, mas sem conhecimento nada se realizará com base nesta inteligência. A aquisição do conhecimento está portanto no coração da inteligência. Na IA, a aquisição de inteligência é a questão principal. Tentamos construir máquinas que podem adquirir conhecimento por si próprias, mas na maior parte das vezes precisam que esse conhecimento venha por nossas mãos. Não há comportamento inteligente que não dependa do conhecimento.

É claro que o tipo de conhecimento a ser posto na máquina é relevante neste processo.

Novamente, segundo Schank e Birnbaum (1997, p. 106),

Não precisamos, como pesquisadores da IA, ensinar às máquinas Filosofia ou Matemática. Precisamos, por outro lado, ensinar-lhes *scripts*, ensiná-las a generalizar, a abandonar o *script*, a compreender o que é igual e o que é diferente, a caracterizar uma experiência, a lidar com uma exceção. Essas são coisas realmente necessárias para a inteligência...

Pensando em construir programas cujo funcionamento fosse orientado por uma base de dados, Barr e Feigenbaum (1982) construíram o Dendral. A estratégia neste caso é acrescentar no programa uma gama de informações muito vasta. Com isso, acreditavam tais pesquisadores, o programa poderia resolver problemas de um modo altamente inteligente.



Um dos objetivos na construção destes programas é fazer com que eles reconheçam absurdos como o exposto no Student.

Sistemas como o Dendral precisam ser especialistas. Isto significa que, para ele ter um melhor desempenho em suas atividades, precisa deter-se sobre áreas específicas. A construção de sistemas mais gerais exigiria um número muito elevado de conhecimento.

O conhecimento, ou base de dados, é formada basicamente por informações do tipo condicional, como as seguintes:

se x é rocha, então x não flutua.

se x tem 40 graus, então x tem febre.

se x tem frio quando faz calor, então x tem febre.

Segundo Barr e Feigenbaum (1982), estes sistemas possuem uma espécie de bom senso, ou seja, poderia reconhecer certos disparates. A primeira informação acima, por exemplo, não permitiria ao sistema aceitar absurdos como o do Student. As duas últimas informações serviriam para um sistema poder diagnosticar uma doença. O usuário do sistema entra com os sintomas e o sistema encontra a doença em questão.

Como visto, para os cientistas da IA a inteligência está relacionada com a capacidade e/ou habilidade de resolução de problemas. Estes são principalmente do tipo lógico-matemático ou de alguma maneira de natureza abstrata. Um problema pode ser definido como um par formado por uma situação inicial e um conjunto "solução" (possibilidades de associações à situação inicial). A situação inicial é a formulação do problema e o conjunto solução seus possíveis resultados.

A solução de um problema é encontrada de acordo com a caracterização do mesmo. Ela deve satisfazer a todas as exigências do problema. Por exemplo, suponhamos que queiramos saber se uma fórmula ϕ é um teorema (situação inicial). Suponhamos ainda que o



conjunto solução seja formado por dois elementos: 'sim' e 'não'. Para ser uma solução da situação inicial, o elemento precisa satisfazer a todas as exigências postas pelo problema (que devem ser propostas na situação inicial). No caso acima, por exemplo, se a resposta for 'sim', as exigências poderiam ser que ϕ deve ter uma demonstração lógica e ser uma sentença necessariamente verdadeira. Notamos com isso, que nem todos os elementos do conjunto solução do problema são realmente um resultado dele. Além disso, para se chegar à solução, é preciso muitas vezes passar por uma série de situações e obstáculos encontrados durante o processo de busca da resposta.

O conjunto solução do problema pode apresentar-se de modos variados. Poderia acontecer, por exemplo, dele ser vazio (pode significar que o problema é insolúvel) ou ter muitos elementos (podendo haver muitos resultados para o problema). É de se notar, porém, que mesmo este conjunto seja vazio, não significa que o problema seja insolúvel. Por outro lado, o fato dele ter muitos elementos não implica que tenha muitas soluções, ao contrário, pode ser insolúvel. O conjunto solução é uma união de possibilidades de resultados de um problema. Muitos problemas, ao serem formulados, tiveram uma infinidade de possíveis soluções, revelando-se posteriormente insolúveis. O mesmo acontece ao contrário. Vários problemas, quando formulados, não possuem nenhuma possibilidade de serem resolvidos. Ao serem analisados mais profundamente, descobriu-se não apenas um, mas vários modos de solucioná-los.

Tentar encontrar a solução ou soluções dos problemas é uma das questões centrais para a caracterização da inteligência em um sistema. Este será inteligente se conseguir resolver problemas do modo mais prático, rápido e racional possível.



Um problema assemelha-se a um quebra-cabeças, no sentido de que existem inúmeras possibilidades de montagem para ele. Basta encontrar a mais eficaz. Podemos comparar a noção de problema a uma árvore invertida. Seu topo é a situação inicial, as ramificações são as exigências e obstáculos e os galhos finais as possíveis soluções (fig. 22) da página seguinte.

Para Newell & Simon (1968), "... uma pessoa é confrontada com um problema quando ela quer alguma coisa e não sabe imediatamente que séries de ações pode realizar para alcançá-la".

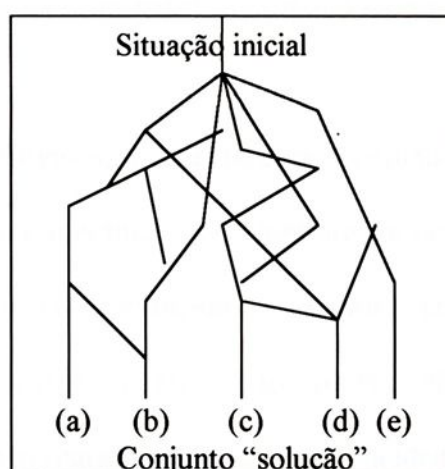


Fig. 20

Representação de um problema com cinco possíveis soluções
Os nódulos são os obstáculos e exigências do problema.

Como visto, muitos cientistas da IA concordam que os processos pelos quais um ser humano passa ao realizar um comportamento inteligente são mecânicos. Por isso, certas máquinas podem simular tais comportamentos e serem assim consideradas inteligentes. Aspectos como biológicos, sociais, culturais não são considerados nesta simulação. Além disso, características subjetivas do indivíduo são desconsideradas no estudo do comportamento inteligente.

Muitas pesquisas foram feitas para mostrar que os aspectos citados acima devem ser considerados ao tratar-se da inteligência, quer na sua aquisição, conceituação ou aplicação. Algumas delas analisavam conjuntos de indivíduos pertencentes a distintas comunidades de uma mesma região, mas que não tinham comunicação entre si (Goodenough e Terman, 1965). A diferença no grau de inteligência entre eles era significativa. Além disso, realizaram experimentos com gêmeos idênticos¹⁷ que, ao serem separados na infância e criados em sociedades distintas, mostraram que eles apresentaram capacidades distintas de resolução de problemas. Para os cientistas destas experiências, estes fatos mostram que aspectos culturais e individuais têm um papel importante na aquisição de capacidade de resolução de problemas.

Além disso, como já dissemos, o ser humano, segundo muitos cientistas não comporta-se de modo totalmente mecânico. Ao contrário, a constituição biológica é de fundamental importância. Um comportamento inteligente envolve a relação entre neurônios, a aprendizagem a partir do erro. Isso, porém, não é feito simplesmente manipulando símbolos e seguindo regras. Veremos a seguir a idéia melhor desenvolvida de alguns destes pesquisadores.

3 O conceito de inteligência no Conexionismo

Segundo conexionistas como Feldman e Ballard (1982), Hopfield (1982), Grossberg (1986), Lippmann (1987) e Caudill e Butler (1992), não é possível entendermos a

¹⁷ Gêmeos idênticos são os que provém de um só óvulo fertilizado por um só espermatozóide. O embrião, depois de começar a vida como um único indivíduo se divide e dá origem a dois indivíduos. Esses possuem exatamente a mesma disposição dos genes, são exatamente iguais em constituição genética.



inteligência humana através da utilização de máquinas do tipo Turing. Ao contrário, é necessário, para essa tarefa obter sucesso, construir sistemas mais adequados, capazes de aprender com a experiência, fazer generalizações, adquirir novos conhecimentos, analisando os erros cometidos na busca do conhecimento. Isso tudo deve ser feito levando-se em consideração aspectos físicos e biológicos do cérebro humano. Além disso, os processos pelos quais o ser humano passa na resolução de problemas não são como os apresentados pela IA, afirmam conexionistas tais como McClelland & Rumelhart (1988). Tais pesquisadores acreditam que muitas tarefas simples são realizadas ao mesmo tempo (em paralelo), e não através do processamento sequencial de uma lista de instruções. O ser humano, ao resolver problemas, não os resolve de um modo algorítmico.

Segundo conexionistas como os citados acima, as redes neurais artificiais possuem uma semelhança estrutural com os sistemas inteligentes biológicos. Ou seja, ambos são formados por partes básicas que conectam-se entre si, originando comportamentos, reconhecimento de padrões etc. Por isso, acreditam que elas são melhores para simular e explicar comportamentos humanos inteligentes do que as máquinas de Turing (Kohonen, 1987; Caudill & Butler, 1993).

Embora as redes sejam de inspiração neuronal, a diferença entre a rede artificial e a natural é muito grande. Mesmo assim, muitos conexionistas acreditam que seus modelos são melhores do que os computadores digitais para a simulação de comportamentos inteligentes. Um exemplo disso são características como aprendizagem competitiva e inibição lateral no reconhecimento de padrões. Tais características, apresentadas por redes como a de Kohonen, são também encontradas no cérebro, acreditam alguns pesquisadores como o próprio Kohonen (1987). Elas, porém, não são apresentadas pelas máquinas do tipo Turing.

Por um lado, a máquina de Turing não apresenta características que, se supõe, encontramos no cérebro, como dissemos acima. Por outro lado, afirmam conexionistas como Caudill & Butler (1993, p. 7), “o computador digital faz coisas que os sistemas biológicos não são capazes. Um destes exemplos é a computação de modo preciso e rápido de funções matemáticas.” Apenas estes dois motivos, pensam os conexionistas, nos mostram que a máquina de Turing, realizada pelo computador digital, não pode ser um bom modelo para o estudo dos processos cognitivos humanos.

As redes são construídas fundamentalmente para reconhecer padrões de um modo semelhante ao apresentado pelo homem, como mostramos no capítulo 3. De acordo com conexionistas como Caudill e Butler (1993), futuramente, além de reconhecer padrões, as redes poderão computar funções com a mesma qualidade das máquinas de Turing. Segundo estes mesmos pesquisadores, ainda não é possível construir sistemas verdadeiramente inteligentes. Porém, se continuarmos no caminho que segue o Conexionismo, a construção destas máquinas é apenas uma questão de tempo, como pensava Turing (1950).

O comportamento inteligente para os conexionistas está associado fundamentalmente à capacidade de aprender a reconhecer padrões. Como dissemos na seção 3.2 do segundo capítulo, um padrão é uma espécie de representação de algum objeto. Por exemplo, um limão pode ser representado por diversas características, como de ser fruta, ser verde, ser azedo etc. Tais características formam a idéia de limão. Reconhecemos este padrão quando somos capazes de classificar o objeto representado pelo padrão de acordo com sua classe.

Cada elemento do padrão é transformado em um determinado número. Na grande maioria da vezes, no processo de simulação das redes, os padrões transformam-se em



conjunto de números. A rede reconhecerá o padrão quando seus nódulos de saída dispararem de tal modo que o classifique corretamente.

Como dissemos, uma rede aprende a reconhecer padrões. Para os conexionistas, a aprendizagem nas redes neurais artificiais acontece fundamentalmente através do ajuste dos pesos da rede. Ela passa por um processo de treinamento, que é exatamente o ajuste dos pesos da rede. Uma vez ajustados os pesos, como afirmam Caudill & Butler (1992, p. 9), pode-se dizer que a rede aprendeu. Desse modo, a aprendizagem não tem como princípio fundamental o seguimento de regras e manipulação de símbolos. Ao contrário, ela acontece através do fortalecimento das conexões entre os neurônios da rede.

A idéia de analisar a aprendizagem através da alteração das conexões entre os neurônios teve como um de seus primeiros adeptos o pesquisador Hebb (1949), que estabeleceu uma lei, conhecida como *lei de Hebb*. Ela basicamente diz o seguinte: *quando um neurônio estimula um outro de tal modo que o faça disparar, a conexão a partir da primeira célula com a segunda é fortalecida* (Hebb, 1949):

Para entender o mecanismo de funcionamento do cérebro quando aprende e o fortalecimento da conexão entre neurônios, podemos apresentar o seguinte programa:

Para todo passo no tempo t
 Compute a atividade de A
 Compute a atividade recebida de A por B
 Compute a atividade de B
 Se a atividade de B é maior do que 0, e a atividade recebida de A por B é maior do que 0,
 então aumente a força de conexão de A para B (Caudill e Butler, 1992, p. 61)

Este programa, em última análise, quer dizer que, se o disparo de um neurônio auxilia no disparo de um segundo, cada vez mais a conexão entre o primeiro com o segundo é

fortalecida. Chegará um momento em que o simples disparo do primeiro fará com que dispare o segundo.

A grande maioria das RNA utiliza a lei de Hebb no processo de aprendizagem. A regra de aprendizagem destas redes têm como fundamento fortalecer os pesos entre os neurônios a fim de que a rede aprenda a reconhecer padrões. Na modelização de RNA, as conexões entre os núdulos são representados por valores numéricos.

Desse modo, a atividade do neurônio biológico na lei de Hebb será o cálculo da resposta do núdulo da rede neural artificial calculado pela função transferência. O aumento da conexão entre os núdulos será dado através da mudança de seus pesos.

A tentativa de assemelhar o processo de aprendizagem das redes neurais artificiais com o dos animais pode ser notada em redes como o perceptron, por exemplo. Nessa rede, se um núdulo dispara através do disparo de um outro ligado sobre ele (quando o segundo deve realmente disparar) a regra de aprendizagem fará com que o peso do núdulo sempre aumente, fortalecendo cada vez mais a conexão.

No caso da rede de Kohonen, quando um padrão incompleto lhe é apresentado, a rede é capaz mesmo assim de reconhecê-lo. Lembremos que cada elemento do padrão é assumido como sendo representado por um núdulo da camada de entrada. Se a rede foi treinada por um longo período, a conexão entre os núdulos da camada de entrada com alguns dos da camada de Kohonen é fortalecida. Se acaso algum dos núdulos de entrada não disparar, o disparo dos outros poderá fazer com que o núdulo vencedor dispare também. Isso porque a conexão entre aqueles e este está fortalecida.

A lei de Hebb apresenta algumas deficiências. Como afirmam Caudill & Butler (1992, p. 60), Hebb não diz como calcular a atividade dos neurônios, tampouco como calcular a

atividade recebida de um por outro. Por fim, não explica como aumentar a força de conexão entre eles.

Atualmente, alguns modelos conexionistas foram criados para realizar tarefas como previsão do tempo e reconhecimento de caracteres escritos à mão. Na Alemanha, uma rede foi treinada para dirigir carros a 60 milhas por hora (Davis, 1992). O problema desta rede é que, se o “cenário” em que a rede foi treinada fosse radicalmente distinto, ela poderia não ser capaz de superar os obstáculos que por ventura viessem a surgir. Por exemplo, suponhamos que a estrada onde ela aprendeu a dirigir fosse cercada por muros. Se ela enfrentasse uma estrada sem estes muros, poderia não saber se orientar. Num ser humano isto não acontece. Seu senso de localização é muito mais aperfeiçoado.

Como visto na seção antecedente, um dos objetivos dos cientistas da IA como Barr & Feigenbaum (1982), é o de construir sistemas que apresentem alguma espécie de bom senso. Isto significa que ele seria capaz de reconhecer o que é ou não absurdo, por exemplo. Um dos meios para isto é construir sistemas com uma base de dados, inserindo um conjunto de informações nele. Ainda que consigam tal aperfeiçoamento, tais sistemas sofrem de algumas críticas. Dentre elas, está a de que as informações precisam ser acessadas no sistema por um programador.

Conexionistas como Lenat, Guha, Pittman, Pratt e Shepherd (1990) tentaram criar sistemas que, segundo eles, possuem tal bom senso semelhante ao encontrado no ser humano. O sistema criado por estes cientistas, denominado CYC¹⁸, é capaz de não cometer ou de identificar absurdos.

¹⁸ Pronuncia-se *psych*, que na gíria americana significa psicanalisar (Gardner, 1998, p. 196).



O objetivo dos criadores do CYC é fazer com que a rede consiga compreender sentenças. Para isso, criam uma base de dados, como no caso do programa criado por Barr & Feigenbaum, citado anteriormente. A diferença entre CYC e Dendral, segundo Lenat et al. (1990) está em que CYC é capaz de originar novos conhecimentos por si só, não apenas pelo simples seguimento de regras lógicas e manipulação de símbolos.

Mesmo para entender sentenças muito simples, CYC precisa de um conjunto muito grande de conhecimento. Para cada tipo de sentenças, deve-se construir uma espécie de roteiro. Por exemplo, conhecimento sobre 'comprar' poderia envolver milhares de regras como as seguintes:

1. Os adultos costumam levar consigo de 10 a 100 dólares (quando estão vestidos e longe de casa).
 2. Pagamentos de menos de 10 dólares são feitos com dinheiro.
 3. Pagamentos acima de 50 dólares normalmente são feitos com cheque ou cartão de crédito.
 4. pagamentos feitos pelo correio geralmente não são feitos em dinheiro
 5. A regra 4 se sobrepõe às regras 2 e 3.
 6. Barra de chocolate custam aproximadamente um dólar.
 7. Balas e chocolates custam aproximadamente o dobro nos cinemas.
- (Lenat e colaboradores, 1990, p. 43, citados por Gardner, 1998, p. 198)

Como dissemos, o objetivo dos criadores do CYC é fazer com que ele compreenda verdadeiramente certas proposições. Porém,

... para que CYC compreenda verdadeiramente que 'Fred foi ao cinema e comprou uma barra de chocolate', CYC precisa possuir o conhecimento de senso comum de que Fred não estava dormindo quando a comprou, e provavelmente não pagou com cartão de crédito. Além disso, se a sentença fosse 'Fred está tentando conseguir 20 milhões de dólares para comprar o Crunch da Nestlé', CYC precisa entender que a pessoa provavelmente está se referindo a um negócio e não a uma barra de chocolate. (Lenat et al., p. 43, citados por Gardner, 1998, p. 197)

Atualmente, acredita-se que a solução para criar um sistema capaz de aprender com a experiência e de resolver problemas computáveis talvez seja a construção de sistemas híbridos. Estes sistemas reuniriam características das máquinas de Turing e redes neurais artificiais. Um exemplo de um sistema destes é o Soar, construído por Lavid, Newell, e Rosenboon (Newell, 1990). Newell (1990) acredita que a construção de sistemas híbridos

poderia explicar toda a amplitude da atividade intelectual humana, incluindo a resolução de problemas, tomada de decisão, memória, aprendizagem, linguagem, motivação, emoções, imaginação.

Cap. 9

Da possibilidade de uma Inteligência Artificial

9.1. Introdução

É impossível negar que a inteligência humana é uma das coisas mais maravilhosas que conhecemos. Ela é capaz de resolver problemas, tomar decisões, aprender, comunicar-se, sentir emoções, imaginar. Mas, será que ela é única? Será que não há nada de semelhante a ela em outros seres vivos ou em máquinas? Esta é a pergunta central da inteligência artificial (IA). A IA é o estudo da possibilidade de se criar máquinas que possam pensar como os seres humanos. Ela é uma das áreas mais interessantes e desafiadoras da ciência da computação. Neste capítulo, vamos explorar a possibilidade de uma inteligência artificial, discutindo os diferentes tipos de IA, os métodos utilizados para criar máquinas inteligentes e as aplicações práticas da IA.

Existem dois tipos principais de IA: a IA fraca e a IA forte. A IA fraca é aquela que é projetada para realizar uma única tarefa específica, como jogar xadrez ou reconhecer imagens. A IA forte, por outro lado, é aquela que é projetada para realizar qualquer tarefa que um ser humano possa fazer.

Cap. 6

Da possibilidade de uma Inteligência Artificial

1 Apresentação

As opiniões entre os pesquisadores das diversas ciências cognitivas sobre a hipótese da máquina ser inteligente são muito variadas. Muitos argumentos foram apresentados para defender ou refutar esta hipótese. Além da inteligência, há cientistas que acreditam na existência de sistemas artificiais com estados e faculdades mentais em geral. Segundo Searle (1984, p.37), McCarthy, por exemplo, afirma que mesmo máquinas simples como termostatos têm crenças do tipo *está muito quente* ou *está muito frio aqui*. A idéia de que máquinas como computadores digitais podem exibir inteligência e estados e faculdades mentais em geral segue a linha funcionalista: basta que o sistema exiba as relações funcionais corretas para estar em, ou possuir, um estado ou faculdade mental.

Neste capítulo, apresentamos um argumento a favor da atribuição de inteligência à máquina denominado *teste de Turing*. Foi ele o norteador de muitos estudos sobre a



inteligência na Ciência Cognitiva. Procuraremos mostrar que, embora tal teste, cujo objetivo é avaliar quem é ou não inteligente, seja embasado e defendido por funcionalistas, ele de fato parece ser fundamentalmente behaviorista.

É claro que a atribuição de inteligência a máquinas não tem a aprovação de todos os pesquisadores da Ciência Cognitiva. Uma das críticas mais conhecidas contra a possibilidade de certas máquinas serem inteligentes é denominada *argumento do quarto chinês*, apresentado na terceira seção deste capítulo. Searle, seu criador, faz uso do teste de Turing para mostrar que máquinas do tipo Turing não podem realmente ser consideradas inteligentes. Isso porque não possuem as características necessárias para serem assim caracterizadas.

2 O Teste de Turing

É impossível falar da pesquisa sobre computadores e inteligência sem lembrar o nome de Alan Turing. Foi ele um dos primeiros a sugerir que os computadores digitais podem ser considerados inteligentes (Turing, 1950). Porém, em vez de perguntar se as máquinas podem ser inteligentes, a pergunta é se elas podem pensar.

A relação entre pensamento e inteligência não é totalmente clara no artigo de Turing (1950). Algumas vezes parece que os dois termos se identificam. Em outros momentos, parece que um é pressuposto para o outro. Neste último caso, poderíamos inferir que, para ser caracterizada inteligente, é preciso que a máquina seja capaz de pensar. É através do pensamento que inferimos novos conhecimentos, entendemos a linguagem natural,



solucionamos problemas (podemos escolher a melhor possibilidade para uma determinada situação), raciocinamos logicamente etc.

Se quisermos estabelecer uma distinção entre os dois conceitos, podemos dizer que o pensamento é o processo de manipulação de símbolos e seguimento de regras. Ou seja, enquanto a máquina está realizando uma computação ela está pensando. Inteligente é o resultado bem sucedido desta computação. Sendo assim, ao ser capaz de responder a diversas perguntas relevantes, podemos dizer que ela pensa inteligentemente. Segundo Turing (1950, p.64),

...sempre que a qualquer uma destas máquinas se faz a pergunta crítica devida, e ela dá uma resposta definida, sabemos que a resposta deverá estar errada e isso nos infunde certo sentimento de superioridade. Será esse sentimento ilusório?...Em resumo, haverá talvez homens mais inteligentes do que qualquer máquina dada, mas mesmo assim poderão existir máquinas mais inteligentes, e assim por diante.

Na citação acima, parece estar claro que inteligente é aquele capaz de responder corretamente a determinadas questões. Este é o resultado final de alguma computação que, nesta última relação entre pensamento e inteligência, seria o processo de pensamento. É difícil estabelecer uma relação direta entre pensamento e inteligência porque Turing não define nem um nem outro. Como ele mesmo afirma, não é sua pretensão defini-los (Turing, 1950, p. 82).

Apresentamos a seguir o argumento de Turing sobre a possibilidade dos computadores digitais serem inteligentes. Algumas vezes, neste trabalho, falamos que o computador digital é inteligente. Outras vezes falamos que máquinas de Turing são inteligentes. Como já falamos no segundo capítulo, os computadores digitais são os sistemas que realizam as máquinas de Turing. Enquanto esta é uma máquina abstrata, ideal, aquela é uma máquina concreta, com componentes físicos e limitada, que realiza a máquina abstrata.

Turing (1950) está preocupado em querer defender a hipótese de que os computadores digitais podem ser considerados inteligentes. Porém, como dissemos, a questão que norteia seu artigo é se as máquinas podem pensar. A idéia de Turing é que, se a máquina for capaz de responder a uma série de questões relevantes, poderemos dizer que ela pensa inteligentemente.

Segundo Turing (1950, p. 50), a pergunta ‘Podem as máquinas pensar?’ é muito ampla e vaga. É preciso, antes de mais nada, delimitá-la, buscando definir o significado de máquina e pensar.

A máquina envolvida no teste é a máquina de Turing, apresentada no capítulo dois deste trabalho. Em linhas gerais, ela é um sistema formado por uma “caixa-preta” capaz de manipular símbolos que lhes são apresentados, através de uma “fita”. Tal manipulação é feita através do seguimento de certas regras lógicas. O funcionamento da máquina caracteriza-se de tal modo que é possível computar qualquer cálculo, por mais complexo que seja, decompondo-o numa série de operações extremamente simples. Tal método pode tornar-se claro na seguinte citação:

A explicação dos processos do pensamento humano é feita via processamento de informação ou nível de manipulação de símbolos. A premissa básica desta abordagem é que processos de pensamento complexos são construídos a partir de processos de manipulação de símbolos elementares. Um conjunto fundamental destes processos poderia ser o seguinte: leia um símbolo, escreva um símbolo, copie um símbolo, apague um símbolo e compare dois símbolos. Se o sistema analisa dois símbolos e os nota como idênticos, toma um curso de ação, caso contrário, toma outro curso. (Feigenbaum & Feldman, 1968, p. 272)¹⁹.

Uma vez definida a máquina, dado o seu modo de funcionamento e características, Turing precisa definir o que é pensar. De fato, não cria uma definição tal como é comum se fazer, do tipo “pensar é x”. Ao invés disto, constrói um jogo para servir de teste segundo o

qual seria possível determinar se a máquina pode ou não possuir esta faculdade. Tal teste, denominado *jogo da imitação*, é jogado por três pessoas,

...um homem (A), uma mulher (B) e um interrogador (C), que pode ser de qualquer dos sexos. O interrogador permanece num quarto separado dos outros dois. O objetivo do jogo para o interrogador é determinar, em relação aos outros dois, qual o homem e qual a mulher. Ele os conhece por rótulos X e Y e no fim do jogo dirá ou “X é A e Y é B” ou “X é B e Y é A”. Ao interrogador é permitido fazer perguntas a A e B tais como: C: Por favor, X, diga-me qual o comprimento de seu cabelo. Suponhamos que X seja A, então A pode responder.

O objetivo de A no jogo é tentar induzir a C a uma identificação errada. Sua resposta, portanto, poderia ser: “Meu cabelo é curto, e os fios mais longos têm cerca de 20 centímetros de comprimento”.

Para que tons de vozes não ajudem o interrogador, as respostas devem ser escritas, ou ainda melhor, digitadas. O arranjo ideal é um telegravador com comunicação entre dois quartos. As perguntas e respostas, de um modo alternado, podem ser repetidas por um intermediário.

O objetivo do jogo para a segunda jogadora (B) é ajudar o interrogador. Sua melhor estratégia será provavelmente dar respostas verdadeiras. Ela pode acrescentar frase como “Eu sou a mulher, não ouça ele” em suas respostas, mas isto é inútil, porque o homem pode dar respostas semelhantes. (Turing, 1950, p. 50).

As perguntas neste jogo podem ser variadas. Vão desde questões sobre arte até a resolução de problemas lógico-matemáticos. De acordo com Turing (1950, p. 51), se a máquina substituir o jogador A e conseguir enganar C, ela é capaz de pensar.

Uma vez melhor caracterizadas a noção de máquina e pensar, Turing substitui a questão original ‘*Podem as máquinas pensar?*’ pelas seguintes:

O que acontecerá quando a máquina tomar o lugar de A neste jogo? O interrogador decidirá erroneamente com a mesma frequência, quando o jogo é jogado dessa forma, do que quando o fazia ao tempo em que o jogo era jogado por um homem e uma mulher? (Turing, 1950, p. 51)

Por fim, as questões seguintes: Podem as máquinas pensar? Existem computadores digitais imagináveis que pudessem ter bom desempenho no jogo da imitação? Existe uma máquina de estado discreto capaz de bom desempenho? podem ser substituídas pela seguinte:

¹⁹ Podemos encontrar uma estreita semelhança deste método com o método cartesiano de análise-síntese, que visa separar problemas em partes simples e depois sintetizá-las, chegando assim à sua resolução ou a algum conhecimento desejado.

Fixemos nossa atenção num computador digital particular C. É verdade que, modificando-se este computador para obter uma memória adequada, aumentando-lhe convenientemente a velocidade de ação, e provendo-o de um programa apropriado, C pode ser preparado para desempenhar satisfatoriamente o papel de A no jogo da imitação, sendo o papel de B desempenhado por um homem? (Turing, 1950, p. 60)

A idéia é que a entidade submetida ao teste seja capaz de realizar muitos e variados comportamentos essencialmente intelectuais. Para pensar, a máquina precisa resolver problemas intelectuais, emitir opiniões sobre arte, participar em jogos etc. Para isso, deve possuir uma memória satisfatória, ser capaz de manipular símbolos rapidamente e conter os programas computacionais corretos. Para Turing (1950), estas três características são o que possibilitam que uma máquina real como o computador digital, possa apresentar um bom desempenho em seu teste. Turing acreditava que até o final deste século, tal máquina já existiria. Atualmente muitos computadores já foram construídos para tentar passar no teste. Porém, ainda não temos computadores que possam entender de um modo razoável a linguagem natural e reconhecer padrões, por exemplo. Por isso tais máquinas ainda não estariam totalmente preparadas para ter um bom desempenho no teste. Portanto, ainda não podem ser realmente consideradas inteligentes, se seguirmos o argumento de Turing.

Embora tenha se tentado criar computadores digitais que tivessem bom desempenho no teste de Turing, deve-se notar que não é este o objetivo fundamental do teste. O importante é a caracterização do conceito de inteligência feito através de um “exame”. O fundamental encontra-se na tentativa de criar um critério segundo o qual poderíamos atribuir inteligência a sistemas artificiais como o computador digital, por exemplo. Se existe ou não tal máquina concretamente, esse é outro problema, talvez de engenharia. O fato é que, se este sistema for capaz de responder às questões do teste de modo apropriado, seria inteligente.



Como dito acima, o teste de Turing caracteriza-se como um “exame” segundo o qual um sistema capaz de responder certas questões é inteligente. Podemos notar neste ponto uma certa semelhança entre este teste e os testes de QI apresentados no capítulo anterior. Porém, tal semelhança não deve ser bem vista na Ciência Cognitiva. Como apresentamos no primeiro capítulo, existe uma tentativa de refutação do behaviorismo por parte de pesquisadores como Putnam (1967), por exemplo. Porém, o teste de Turing parece ser fundamentalmente comportamental. Isso porque, para o interrogador do teste, responsável pela atribuição ou não de inteligência à máquina, o modo de funcionamento dela não é relevante. O que importa realmente é que a máquina comporte-se do mesmo modo que o homem.

É estranho que pesquisadores defensores do funcionalismo aceitem um teste aparentemente com essência comportamental. O teste deveria tomar como relevante a estrutura funcional dos sistemas envolvidos. Porém, pelo contrário, a “caixa-preta” parece ser quase totalmente ignorada. O que é realmente importante são as entradas e saídas dos participantes do teste.

Turing (1950, p. 76) confirma que seu teste tende a ser fundamentalmente comportamental. Para ele, não importa saber as características físicas ou estéticas da máquina para saber se ela é ou não inteligente. Tampouco é preciso resolver problemas como o mistério da consciência para avaliar a inteligência da máquina. O importante é que ela apresente um comportamento semelhante ao do homem. A falta de consciência na máquina é um dos pontos principais contra a tese de que ela pode ser inteligente, como veremos na próxima seção.



Além de parecer mais comportamental do que funcional, podemos citar algumas outras críticas ao teste idealizado por Turing, ainda que ele acredite que questões sobre o teste não sejam pertinentes. Dentre elas, podemos nos perguntar quais características devem apresentar o interrogador e o segundo jogador do teste. Qual o grau de inteligência que eles devem apresentar? Quantos interrogadores a máquina deve enganar para ser considerada inteligente? O jogo testa realmente a capacidade da máquina pensar ou do interrogador de poder distinguir uma máquina de um ser humano?

Turing, de fato, parece pressupor que tanto o interrogador quanto o jogador B devem ser inteligentes. Porém, sem apresentar uma definição de inteligência para estabelecer os critérios de escolha deles. Por isso mesmo, podemos nos perguntar que garantia temos de que ambos interrogador e jogador B têm os requisitos suficientes para podermos nos assegurar de que a máquina seria considerada inteligente, caso passasse no teste? Talvez não seria melhor se ambos tivessem passado anteriormente por ele? Porém, se assim fosse, o teste de Turing geraria uma circularidade com respeito ao interrogador e ao segundo jogador.

A circularidade, no caso do interrogador, se configuraria no seguinte sentido: se ele está tentando distinguir o jogador A do jogador B, precisaria ser capaz de entender e saber responder às questões que ele mesmo apresenta aos outros dois jogadores. Deveria, portanto, ser inteligente. Sendo assim, precisaria ter passado no teste. Caso contrário, a confiabilidade do jogo ficaria comprometida. Como poderíamos aceitar o resultado do teste se ele foi aplicado por um indivíduo do qual não sabemos se está preparado para distinguir os jogadores? Mas, para este interrogador ter passado no teste, alguém deve ter assumido o papel de interrogador para ele. Para este novo interrogador, o raciocínio é o mesmo que



para o anterior. Desse modo, chegará algum momento em que não haverá mais interrogador. Sendo assim, temos duas possibilidades: a primeira é aceitar um primeiro interrogador necessariamente inteligente. Se não aceitamos esta possibilidade, a única restante é submeter-lhe ao teste. Porém, para isto, devemos tomar como interrogador alguém que, direta ou indiretamente o tinha como interrogador quando foi submetido ao teste. Portanto, um seria o interrogador do outro.

Como dissemos, Turing não está interessado no tipo de questões apresentadas acima. Ele pressupõe que o segundo jogador e o interrogador são inteligentes. Porém, não fornece os critérios para a escolha de tais indivíduos. Por isso, acredito que as críticas à caracterização do teste são pertinentes.

Na próxima seção, apresentaremos as críticas não ao teste propriamente dito, mas à hipótese de que as máquinas que apresentam bom desempenho nele podem ser realmente consideradas inteligentes.

Quais são de fato as hipóteses e conclusões do argumento de Turing? Em primeiro lugar, Turing toma como hipótese que um ser humano e uma máquina de Turing são semelhantes no sentido de que funcionam do mesmo modo. Como consequência desta hipótese, podemos obter que eles comportam-se igualmente. O comportamento aqui são as respostas da máquina e do homem referentes às perguntas que dizem respeito ao intelecto. Uma segunda hipótese é que, se o comportamento do homem é inteligente, o da máquina também o será. Isto é claro, uma vez que o comportamento de ambos são semelhantes, para não dizer idênticos. Como terceira hipótese, Turing assume que o comportamento do homem é realmente inteligente. Portanto, das duas últimas hipóteses, deduz-se que o comportamento da máquina é inteligente, sendo assim, ela própria inteligente.



Como dissemos, o fundamento do teste de Turing está na tentativa de criar um critério segundo o qual poderíamos atribuir inteligência a certas máquinas. Turing (1950, p. 61) acreditava que até o final deste século, haveriam máquinas reais capazes de passar no teste. Qual seria então o caminho para a construção de tais máquinas reais?

Segundo Turing (1950), o caminho para a construção de máquinas reais para passar no teste é ensiná-las como se estivéssemos ensinando uma criança do primário. Ao invés de procurar simular uma mente adulta, que tem muitos conhecimentos, experiências e programas, Turing acredita que o segredo está em se tentar simular uma mente infantil. Tal mente teria um conjunto limitado de informações, regras e programas. Por isso, haveria mais possibilidade de simulá-la.

O processo de aprendizagem da máquina seria realizado principalmente através da introdução nesta máquina de novas regras, informações e programas. Para Turing (1950), este é o caminho necessário para se construir uma máquina inteligente: inserir programas relevantes na máquina e aumentar sua quantidade de memória.

Pelo fato de ser programada e ter seus passos todos determinados, a máquina de Turing é criticada por não possuir criatividade (Bodem, 1999). Muito antes de Turing tal crítica já havia sido feita à máquina analítica de Babbage, que realizava cálculos matemáticos. Para refutar esta crítica, Turing (1950, p. 71) afirma que “as máquinas lhe pegam de surpresa com muita frequência.” Muitas vezes ela produz respostas pelas quais ele não esperava. Por isso, nota uma certa criatividade na máquina. Ela cria algo novo, que não é previsto pelo programador.

É claro que o fato da máquina produzir um comportamento que foge da previsão do programador pode ser explicado pelo fato dele não ter todo o programa da máquina em

mente. Ou seja, se ele conseguisse ter toda a estrutura do programa em vista, poderia determinar, em qualquer ocasião, qual o comportamento futuro da máquina.

Como dissemos, as respostas que poderiam pegar Turing de surpresa caracterizavam-se por fugirem de suas previsões. Elas poderiam tanto estar certas quanto erradas. Esta última afirmação refuta, segundo Turing (1950), a crítica de que a máquina não erra. Ao contrário, elas erram com muita frequência.

Um dos motivos pelos quais a máquina produz erro é porque ela tem presente em si a regra de indução. A inferência indutiva caracteriza-se pela generalização de um conhecimento limitado sobre um conjunto de objetos ou de fatos. Por exemplo, suponhamos que analisamos um número limitado de cavalos e constatamos que todos eles possuem coração. A partir daí concluímos que os cavalos em geral têm coração. Porém, não temos nenhuma garantia de que a conclusão segue-se de nossas constatações. Isto porque faltam cavalos a ser examinados e, por isso, nossa conclusão pode estar errada. Sendo assim, uma vez que a máquina de Turing segue tal regra, Turing (1950) assume que ela é passível de erro.

Ainda porque a máquina de Turing possui a regra de indução, Turing (1950) acredita que ela pode aprender com a experiência. Um exemplo disso seria, por exemplo, constatar que, ao deparar-se um grande número de vezes com determinada resolução de problemas de um certo tipo, a máquina constata que, cada vez que aparecer problema semelhante, deve seguir o mesmo curso pelo qual resolveu os outros.

Dentre as críticas à hipótese de que máquinas tipo Turing aprendem com experiência podemos citar aquela feita por conexionistas como McClelland & Rumelhart (1988), por exemplo. Para eles, a aprendizagem é, dentre outras coisas, o resultado de ajustes das partes

básicas de um sistema, adequando as conexões entre elas. Por outro lado, a aprendizagem envolve compreensão do que se faz e não é feita simplesmente via manipulação de símbolos e seguimento de regras (Searle, 1984).

Um simples exemplo sobre o modo de comportamento de uma máquina considerado inteligente é o seguinte: suponhamos que as respostas a serem dadas pelos participantes do jogo dizem respeito à relação de parentesco entre eles e determinadas pessoas. A questão a ser respondida é, por exemplo, ‘Paulo é seu tio?’. Para responder a essa pergunta, tanto a máquina quanto o ser humano do jogo têm um conjunto de informações armazenadas em sua memória de longo prazo (onde estão estocadas informações que ficam ‘memorizadas’ por um longo tempo). Existe ainda uma memória de curto prazo, na qual armazenam-se informações por pouco tempo e onde estão os processos que devem ser feitos para a resolução de um problema. Assim, por exemplo, suponha que, para responder à pergunta ‘Paulo é meu tio?’ os participantes do teste tenham as seguintes informações:

<i>Memória de Longo Prazo</i>	<i>Memória de Curto Prazo</i>
Informações	Processos
João é meu pai Maria é minha mãe Pedro é meu irmão Júnior é meu primo Marta é minha prima Paula é irmã de Maria Paulo é irmão de João	Paulo é meu tio?

Para realizar esta operação, ou para resolver este problema, os participante precisam saber de outras informações ou seguir outras regras, tais como as seguintes:

<i>X é meu pai</i>	<i>X é irmão de Y</i>	<i>X é meu pai</i>
<i>Y é ‘irmão’ de X</i>	<i>Logo, X é masculino</i>	<i>Y é irmão de X</i>
<i>Logo, Y é meu ‘tio’</i>		<i>Logo, X é masculino</i>
		<i>Logo, Y é meu tio</i>

(os termos que aparecem entre aspas são termos sem gênero)

O participante do teste, além das regras, precisa saber quem são seu pais, os irmãos deles e seu sexo, para poder determinar se Paulo é ou não seu tio. Desta forma, o conjunto de informações neste passo deveria ser o seguinte:

<i>Memória de Longo Prazo</i>	<i>Memória de Curto Prazo</i>	
	Informações	Processos
João é meu pai Maria é minha mãe Pedro é meu irmão Júnior é meu primo Marta é minha prima Paula é irmã de Maria Paulo é irmão de João		Paulo é meu tio? Encontre meus pais Encontre os irmãos de meus pais

Para encontrar as soluções destes subproblemas postos como novos processos, basta que o participante procure pelas informações contidas na memória de longo prazo. Os subproblemas surgiram da utilização da regra seguinte: se x é meu tio, então x é irmão de meu 'pai'. Se o conseqüente desta sentença for falso, a resposta à questão será negativa. Portanto, o jogador precisa saber quem são seus pais e se Paulo é irmão de algum deles para poder responder corretamente à questão. Encontradas as informações necessárias, o novo quadro de informações assim ficaria:

<i>Memória de Longo Prazo</i>	<i>Memória de Curto Prazo</i>	
	Informações	Processos
João é meu pai Maria é minha mãe Pedro é meu irmão Júnior é meu primo Marta é minha prima Paula é irmã de Maria Paulo é irmão de João	João é meu pai Maria é minha mãe Paula é irmã de Maria Paulo é irmão de João	Paulo é meu tio?

A partir destas informações, seguindo as regras dadas acima, a resposta correta a ser dada à questão inicial será positiva. Como veremos na seção seguinte, exatamente por simplesmente manipular símbolos e seguir regras lógicas, a máquina não entende o que faz. Por isso, comportamentos como o acima descrito não podem ser considerados inteligentes (Searle, 1984)

Hoje podemos saber se o sonho de Turing foi ou não alcançado ou se é possível ou não o seu alcance. Temos máquinas campeãs em jogos como xadrez, máquinas que demonstram teoremas lógico-matemáticos, diagnosticam doenças, compõem música. Porém, ainda não temos máquinas que dominam a linguagem natural e reconhecem padrões do modo como o ser humano o faz, por exemplo. É questão de saber se é possível a criação de máquinas capazes de alcançar este nível de atuação ou não para considerá-la inteligente. Isso por que tais atividades estão presentes no teste de Turing.

Se, por um lado, as máquinas têm seus defensores, seguros de que elas terão bom desempenho no teste de Turing, por outro têm também seus críticos. Alguns destes acreditam que nem passando neste teste as máquinas podem ser caracterizadas inteligentes. Isso porque lhes faltam características fundamentais para que a elas seja atribuída esta propriedade. Vejamos a seguir quem são e o que dizem alguns destes críticos.

3 O Argumento do Quarto Chinês

Se é impossível falar de computadores e inteligência sem lembrar o nome de Turing, também o é sem lembrar o de Searle. O argumento criado por este filósofo para refutar a idéia de que é possível a atribuição de inteligência (e estados e faculdades mentais em geral) às máquinas do tipo Turing foi um dos mais populares da história da Ciência Cognitiva.



Atribuir inteligência a um sistema, segundo a maioria dos cientistas da IA significa ele ser capaz de realizar certos comportamentos, realizados através da manipulação de símbolos e seguimento de regras lógicas (Newell, Shaw & Simon, 1958, Putnam, 1967). De acordo com Searle (1984, p.37),

O colega de Simon, Alan Newell diz 'já descobrimos' (notem que Newell diz 'descobrimos', não 'supusemos' ou 'consideramos a possibilidade', mas *descobrimos*) que a inteligência é justamente uma questão de manipulação de símbolos físicos; não tem nenhuma ligação essencial com qualquer tipo de material ou umidade biológica ou física. Antes, qualquer sistema capaz de manipular símbolos físicos de modo correto é capaz de inteligência no mesmo sentido literal que a inteligência humana dos seres humanos. Simon e Newell sublinham, pela sua honra, que não existe nada de metafórico nestas pretensões; proferem-nas de um modo inteiramente literal... Marvin Minsky do MIT diz que a próxima geração de computadores será tão inteligente que 'teremos muita sorte se eles permitirem manter-nos em casa como animais de estimação domésticos.'

O objetivo de Searle (1984) é mostrar que a concepção acima é equivocada. Para isso, baseia-se no fato de que o computador digital funciona apenas sintaticamente. Segundo tal filósofo, isso não é suficiente para esta máquina compreender o que faz. Porém, a compreensão é condição necessária para a inteligência. Além disso, características fundamentais da mente humana como consciência e intencionalidade são causadas pelo cérebro e são propriedades dele. Sistemas puramente formais não são capazes de originar ou possuir consciência e intencionalidade, afirma Searle (1998). Uma máquina do tipo Turing, conclui, não pode ser uma mente porque esta possui mais do que uma estrutura formal, possui um conteúdo (Searle, 1984, p. 40)

Searle recria o teste de Turing para mostrar que os computadores digitais não pensam e não são inteligentes. Substitui a máquina implicada no teste por um ser humano funcionando do mesmo modo que ela, ou seja, manipulando símbolos e seguindo regras lógicas. Se Searle conseguir provar que o comportamento deste ser humano não é inteligente, tampouco o comportamento da máquina o será, mesmo passando no teste de

Turing. Para Searle, a questão fundamental não está no fato da máquina de Turing ser ou não capaz de responder às questões do interrogador daquele teste. O problema encontra-se na falta de características fundamentais desta máquina, tais como a consciência, para a caracterização do pensamento ou da inteligência.

O argumento do quarto chinês é assim exposto por Searle (1984, p. 40):

Imaginemos que alguém está fechado num quarto e que neste quarto há vários cestos cheios de símbolos chineses. Imaginemos que alguém, como eu, não compreende uma palavra de chinês, mas que lhe é fornecido um livro de regras em inglês para manipular os símbolos chineses. As regras especificam as manipulações dos símbolos de um modo puramente formal em termos da sua sintaxe e não da sua semântica. Assim a regra poderá dizer: 'tire do cesto número um um símbolo esticado e ponha-o junto de um símbolo encolhido do cesto número dois.' Suponhamos agora que alguns outros símbolos chineses são introduzidos no quarto e que esse alguém recebe mais regras para passar símbolos chineses para o exterior do quarto. Suponhamos que, sem ele saber, os símbolos introduzidos no quarto se chamam 'perguntas' feitas pelas pessoas que se encontram fora do quarto e que os símbolos mandados para fora do quarto se chamam 'respostas às perguntas'. Suponhamos, além disso, que os programadores são tão bons para escrever programas e que alguém é igualmente tão bom em manipular os símbolos que muito depressa as suas respostas são indistinguíveis das de um falante chinês nativo. Lá está ele fechado no quarto manipulando os símbolos chineses e passando para fora símbolos chineses em resposta aos símbolos chineses que são introduzidos.

Searle afirma então que o comportamento do indivíduo do quarto com o do comportamento de um nativo chinês são praticamente indistinguíveis. Porém, o indivíduo do quarto não entende uma só palavra do chinês. Ora, é certo que o comportamento deste indivíduo não pode ser considerado inteligente, diz Searle, porque lhe falta a semântica, característica fundamental para um comportamento deste tipo ser inteligente. Para Searle (1984, p. 45),

pensar é mais do que apenas uma questão de eu manipular símbolos sem significado; implica conteúdos semânticos significativos. Estes conteúdos semânticos são aquilo que nós indicamos por 'significado'.

O argumento de Searle é proposto tendo em vista um programa de computador criado por Schank que, segundo o autor, é capaz de entender histórias em inglês. De acordo com o argumento de Searle, a máquina que realiza tal programa está na mesma situação do

indivíduo do quarto chinês. De fato, não entende nada do que faz. Isto porque simplesmente manipula símbolos formais sem significado de acordo com as instruções fornecidas (Searle, 1998, p. 81).

Embora o argumento de Searle esteja direcionado para a compreensão de uma língua, de fato ele pode ser aplicado aos estados e faculdades mentais em geral. Mesmo que um computador consiga simular uma dor, por exemplo, ele de fato não tem esta dor. Isto porque, dentre outras coisas, não possui intencionalidade e consciência, que não podem ser apenas originadas pela manipulação de símbolos, afirma Searle (1998, p. 82).

Como dissemos na terceira seção do primeiro capítulo, de acordo com Searle (1984), a mente é causada pelo cérebro. Ou seja, estados e faculdades mentais, além da consciência e intencionalidade, emergem da interação entre neurônios. Os cientistas da IA não levam em consideração as propriedades biológicas do cérebro ao construir suas máquinas, afirma Searle (1984). Ao contrário, elas apenas funcionam manipulando símbolos e seguindo regras lógicas. Por isso, não podem ser mentes. Sendo assim, jamais poderão possuir estados ou faculdades mentais.

Uma vez que estamos tratando de um argumento, é natural que possamos distinguir suas hipóteses e conclusões, principalmente no que diz respeito à inteligência. A primeira hipótese do argumento de Searle é que o funcionamento do homem e da máquina de Turing é o mesmo. Ou seja, ambos apenas manipulam símbolos e seguem regras lógicas. Uma segunda hipótese é que, se o homem não compreende chinês, a máquina tampouco o compreenderá. Esta hipótese é assegurada pela anterior. Isto porque a máquina não faz nada mais do que o homem. Se este não consegue fazer algo, aquela tampouco o conseguirá. Uma terceira hipótese diz respeito ao fato de que o homem (do quarto) não compreende

chinês. Das duas últimas hipóteses, podemos concluir que a máquina não compreende chinês. Daí segue-se que seu comportamento não pode ser considerado inteligente, uma vez que a compreensão do que se faz é pré-requisito para a inteligência.

Porém, sabemos que o ser humano, em muitas ocasiões (talvez na maioria delas) compreende o que está fazendo. Daí vem duas outras conclusões: o homem não funciona apenas sintaticamente e ele é distinto da máquina de Turing. Ora, sabemos que a máquina de Turing funciona sintaticamente. A partir daí deduzimos que, funcionando desse modo apenas, não é capaz de compreender o que faz. Se o homem funcionasse apenas sintaticamente, tampouco compreenderia o que estaria realizando. Porém, o homem compreende o que faz. Logo, não pode funcionar apenas sintaticamente. Portanto, é diferente da máquina de Turing. O salto que fizemos da compreensão do chinês para a compreensão de outros comportamentos é justificável pelo fato do caso do chinês ser apenas um exemplo de um comportamento. Em última análise, tudo o que a máquina faz segue o mesmo processo da compreensão de uma língua.

Citamos a seguir um exemplo de uma aplicação do argumento de Searle. Um indivíduo precisa responder questões na língua “asturttes”, inventada por nós. A este indivíduo são dados três conjuntos de caracteres asturtteses: 1) *script* (conjunto qualquer de informações, geralmente associado à história a ser compreendida). Neste conjunto estão as regras para o indivíduo poder manipular os símbolos asturtteses; 2) conjunto de símbolos que constituem a história; 3) conjunto de símbolos que constituem as perguntas sobre a história. Os três conjuntos poderiam ser os seguintes:

- 1) símbolos em asturttês representando informações variadas, principalmente sobre a história contada,

Regra i) se $\frac{\bar{U}}{\bar{U}}$ então $\bar{U}$

Regra ii) $\frac{\bar{U}}{\bar{U}}$ se e somente se $\bar{U}$

2) $\bar{O} \bar{A} \bar{A} \bar{E} \bar{Q} \bar{Z} (\bar{A} \bar{O} \bar{O} \bar{D} \bar{I} \bar{E}$

3) $\bar{U}?$

Um indivíduo capaz de manipular símbolos e seguir as regras acima poderia tranqüilamente responder afirmativamente à pergunta feita no terceiro conjunto de símbolos. O símbolo da pergunta não aparece na história. Portanto, há de se supor que o indivíduo fez uma inferência do que ela diz, o que pode pressupor que ele está entendendo o que lê. É claro que mesmo neste caso não há garantia de seu entendimento sobre o que está “lendo”. Faça você esta experiência e responda-me o que entendeu da história. Tudo não passa de uma seqüência ininteligível de símbolos para o leitor.

De fato, o exemplo acima revela um caso muito simples e localizado do comportamento de um sistema. O que Turing buscava realmente era pensar num conjunto de regras tão poderoso a tal ponto que um sistema fosse capaz de apresentar um bom desempenho em qualquer situação a que viesse a se deparar. Neste caso, teria condições de apresentar um bom desempenho em qualquer situação do teste de Turing e não apenas em casos isolados. O problema que nos parece é que, sendo regras muito poderosas ou não, a essência do funcionamento do sistema continuaria sendo a mesma. Como no exemplo da compreensão de uma língua, apresentado há pouco, para quem recebe as respostas do lado de fora do quarto, a impressão é a de que o indivíduo está entendendo perfeitamente a

história. Porém, se observarmos atentamente o modo de funcionamento do sistema, veremos que não existe nada além da simples manipulação de símbolos insignificantes, afirma Searle (1984). Ele não entende o que está fazendo.

O mesmo ocorrido com o indivíduo do quarto chinês acontece com a máquina no teste proposto por Turing, afirma Searle (1984). Vemos apenas as aparências do comportamento. Olhamos de fora o que a máquina está realizando. É apenas um teste comportamental, no qual o objetivo é fazer com que a máquina obtenha os mesmos *inputs-outputs* do homem no jogo. Ter o mesmo comportamento, porém, não significa ter a compreensão do que se está fazendo. Sem ter compreensão ou entendimento do que se faz, não é possível atribuir inteligência a tal comportamento de um indivíduo.

Mas, suponhamos que um ser humano ou mesmo um robô super estruturado (à imagem e semelhança de um homem), armazenasse todas as regras e fosse capaz de manipular símbolos muito rapidamente. Tal máquina seria capaz de conversar com um nativo da língua em questão. Isso devido ao fato de poder manipular de imediato os símbolos recebidos e produzir respostas adequadas a eles. Desse modo, o nativo do lugar em que tal língua é a falada poderia lhe fazer perguntas sobre o que havia falado. O sistema manipulador de símbolos as responderia com uma grande margem de acerto. Isto significaria então que tal sistema estaria realmente entendendo a língua em questão. De acordo com Searle (1984), mesmo agindo desse modo, o sistema não entenderia o que faz, pois continuaria manipulando símbolos e seguindo regras lógicas. Para entender uma língua qualquer, por exemplo, precisaria ultrapassar o limite sintático: deveria também funcionar semanticamente (atribuir significado aos símbolos, saber do que se está falando) e pragmaticamente (entender o significado não literal das palavras, ou seja, seu significado



metafórico, conotativo, etc.). Não basta, segundo Searle (1984), observar apenas o comportamento externo do sistema. É preciso, fundamentalmente, analisar seu funcionamento interno para atribuir-lhe inteligência.

Como dissemos, Searle afirma que a máquina não entende o que faz exatamente pelo fato de funcionar apenas sintaticamente. Isso, juntamente com a falta de propriedade biológicas do cérebro humano, leva-o a afirmar que a máquina não tem consciência nem intencionalidade. A falta destas características, pré-requisitos para o sistema ser inteligente, não é um problema tecnológico, mas sim de princípios, afirma este filósofo. Para Searle (1984), máquinas do tipo Turing, tais como construídas, jamais poderão ser conscientes ou ter intencionalidade.

Porém, está Searle afirmando que nenhuma máquina pode ser considerada inteligente? Esta questão é claramente respondida de modo negativo por tal filósofo. Para Searle (1984, p. 44) "... nós somos máquinas biológicas." Qualquer outro sistema capaz de realizar certos tipos de operações poderia ser considerada inteligente. Porém, tais operações poderiam ser realizadas por uma máquina apenas se fossem levados em consideração aspectos biológicos do cérebro humano.

A crítica de Searle é direcionada essencialmente aos computadores digitais (realizadores dos programas da IA) que não precisam apresentar aspectos biológicos ou físicos do cérebro humano. Para Searle (1997), ainda não conseguimos criar uma máquina artificial inteligente porque não conhecemos o cérebro satisfatoriamente. Quando os neurocientistas descobrirem as capacidades cerebrais que causam a mente, poderemos criar sistemas artificiais capazes de exibir estados e faculdades mentais e comportamentos realmente inteligentes.



Existem muitas variações do argumento do quarto chinês apresentadas pelos cientistas para refutar a idéia de que o computador possa ser considerado inteligente. Uma delas é a seguinte: suponhamos que seja dado para um aluno do primário um cálculo complicado de matemática para resolver. Este aluno sabe muito pouco de matemática, mas é capaz de seguir as operações básicas que solucionarão o cálculo. Seguindo instrução por instrução, o estudante chega ao fim da computação e resolve o problema corretamente. Então nos perguntamos: poderíamos dizer que ele entende ou sabe o que fez? Há alguma inteligência em seu comportamento? A resposta razoável a estas questões deveria ser não. Isso porque ele apenas está seguindo instrução por instrução para a resolução do cálculo. Não tem uma idéia clara do que faz no todo. Agora, suponhamos que substituíssemos o aluno pela máquina. O comportamento de ambos é exatamente o mesmo, qual seja, manipular símbolos e seguir regras lógicas, de um modo algorítmico. Poderíamos dizer que ela entende o que está fazendo ou que seu comportamento é inteligente? Se a resposta a estas questões é negativa para o aluno, certamente também o será para a máquina. O grande problema da máquina neste sentido é que ela simplesmente manipula símbolos e segue regras lógicas (se é que realmente fazem pelo menos isto), sem saber ou entender o que faz.

Button e seus colegas (1997, p. 178-79) constróem um exemplo semelhante ao do teste de Turing, utilizando o quarteto Kronos e um aparelho de CD tocando uma mesma música. Sem ver nenhum deles, uma pessoa é posta para diferenciar qual é um e qual é outro. É muito provável que ele não consiga diferenciá-los, uma vez que o som é praticamente o mesmo. Porém, parece óbvio que este sujeito não atribuiria inteligência ao aparelho de CD apenas porque está reproduzindo uma música tocada pelo quarteto. O aparelho não tem a criatividade, a capacidade de relação entre os componentes do grupo

que, cada um com sua tarefa, juntos formam uma bela melodia. Além disso, para chegar ao estágio de perfeita harmonia, o quarteto teve muitos ensaios, erros e reparos destes erros. O aparelho, por sua vez, não é capaz de nada disto. Assim como o teste do aparelho de CD, diz Button e seus colegas, o teste de Turing esconde a máquina num quarto para que o interrogador não possa vê-lo, podendo assim enganá-lo com mais facilidade. Este é um dos motivos pelos quais o teste de Turing não é bom para testar a inteligência da máquina.

É preciso ter cuidado com algumas analogias tais como a citada no parágrafo anterior. É claro que não podemos comparar um aparelho de CD com um computador. Sabemos que este é infinitamente mais complexo e capaz do que aquele.

A crítica, também feita por Searle de que as máquinas ficam escondidas tem fundamento para os simpatizantes do teste. Isso porque, se visse a máquina, haveria uma predisposição do interrogador, movido por um sentimento de superioridade e de portador único da capacidade de inteligência, de negar que a ela pensa, mesmo antes de aplicar o teste.

Se o interrogador vir a máquina, o teste se desconfigura, pois não haverá mais o que adivinhar. O teste então seria feito com interrogador e máquina frente a frente, como uma espécie de entrevista para aceitar ou não a máquina no grupo dos inteligentes. É pelo fato de que o interrogador esteja vendo o interrogado que aquele poderia não caracterizar o comportamento da máquina como inteligente, simplesmente pelo fato de saber que trata-se de uma máquina.

Searle faz uso muitas vezes do bom senso para refutar os pressupostos da tese de que máquinas do tipo Turing são ou podem ser inteligentes. Isso faz com que muitos cientistas tomem suas críticas à Inteligência Artificial como muito limitada e pouco científica. Searle



concorda em algumas vezes com isto. Porém, ainda assim, acredita que qualquer pessoa de bom senso, ao analisar o comportamento e funcionamento de uma máquina de Turing não o classificaria como inteligente.

Alguns cientistas tentaram criar argumentos mais “científicos”, utilizando a matemática, lógica e computabilidade para refutar a idéia de que o pensamento e inteligência possam ser analisadas mecanicamente e aplicados à máquina. Um destes cientistas é o matemático Penrose (1993, 1994). Para refutar a idéia de mecanização do pensamento, Penrose tomou como ponto de partida o teorema de Gödel. Segundo este teorema, não é possível provar, para qualquer teoria, mesmo a mais simples que envolva a aritmética, que ela seja completa. Ou seja, existem verdades nesta teoria que não podem ser demonstradas pela simples manipulação de símbolos. Sendo assim, nem todas as verdades podem ser demonstradas mecanicamente. Isso significa que precisamos algo a mais do que a pura sintaxe quando resolvemos problemas deste tipo. Ora, mas se o comportamento inteligente consiste na resolução de problemas como estes, então o comportamento inteligente não pode ser examinado de modo puramente sintático.²⁰

Penrose não afirma que nenhum comportamento possa ser analisado desse modo. Porém, existem aqueles que exigem algo a mais do que a pura computação. Ao existir um por cento de comportamentos que não podem ser simulados computacionalmente, isto já é suficiente para mostrar que não somos máquinas de Turing, segundo Penrose (1994). Penrose afirma que trata de habilidades matemáticas porque é o que ele mais conhece. Porém, segundo Searle (1998, p. 86), em conversas informais, Penrose afirma que

²⁰ Este argumento é exposto aqui muito rapidamente. No cap. 2, argumento semelhante é melhor desenvolvido, utilizando o problema de Hilbert para mostrar o aparente fracasso da possibilidade de mecanização do pensamento.



argumentos semelhantes podem ser apresentados sobre música, arte ou outras atividades humanas conscientes.

A clássica defesa da crítica acima é encontrada em Turing (1950). Segundo ele, o teorema de Gödel não traz nenhuma ou traz poucas limitações para a proposta de mecanização do pensamento. Se a máquina não é capaz de demonstrar algumas verdades, isto não é problema, porque o ser humano sofre da mesma dificuldade. Não é necessário que a máquina demonstre toda verdade para garantir que o pensamento humano dá-se através da computação de programas. Ao contrário, tal limitação apenas ajuda a corroborar a idéia de que o pensamento pode ser entendido mecanicamente.

Como dito, a crítica feita por Searle e Penrose à hipótese de que as máquinas podem ser caracterizadas inteligentes refere-se fundamentalmente às máquinas do tipo Turing. O argumento do quarto chinês não diz respeito às redes neurais artificiais. Isso porque o funcionamento delas é distinto do da máquina de Turing. Em trabalhos recentes, Searle (1998) acredita que estas redes podem possuir alguma possibilidade de apresentar semelhanças com a mente humana.

Uma das críticas feitas à hipótese de que os modelos conexionistas podem ser considerados inteligentes é feita por Dreyfus (1993, p. xxxviii), segundo o qual "... a rede exibirá a inteligência nela embutida pelo projetista para aquele contexto, mas não terá o senso comum que lhe permitiria adaptar-se a outros contextos."

Dreyfus chega a esta conclusão através da análise de experiências feitas com redes para reconhecimento de padrões. Um destes trabalhos consistia em fazer com que uma rede neural artificial reconhecesse a presença de tanques de guerra numa floresta. Segundo Dreyfus (1993, p. xxxvi, citado por Button et al. p. 155), o exército tirou um

certo número de fotografias de uma floresta se os tanques e, em seguida, alguns dias mais tarde, com os tanques aparecendo claramente por trás das árvores, e treinaram uma rede para que ela distinguísse os dois tipos de fotos. Os resultados foram impressionantes, e o exército ficou ainda mais impressionado quando se soube que a rede podia generalizar seu conhecimento para fotos que não haviam feito parte do conjunto de treinamento. Só para se ter certeza de que a rede estava de fato reconhecendo tanques parcialmente ocultos, no entanto, os pesquisadores tiraram mais fotos na mesma floresta e as mostraram à rede treinada

Porém, neste novo lote de fotos, a rede não obteve sucesso. Não conseguiu distinguir fotos com tanques parcialmente escondidos atrás de árvores e fotos sem nada atrás delas. O que descobriu-se depois é que o primeiro lote de fotos havia sido tirado em dias distintos, um ensolarado e outro não. Desse modo, a rede não aprendeu a reconhecer a existência de tanques na floresta, mas sim, aprendera a reconhecer florestas com e sem sombras. Dreyfus afirma que

..os projetistas de redes não mais podiam permitir que seus sistemas fossem 'treinados' sem pré-especificar, e portanto, restringir, a classe de generalizações apropriadas permitida para o problema (ou 'espaço de hipótese'). A arquitetura das redes, é assim, planejada para transformar *inputs* em *outputs* 'somente das maneiras que estão no espaço de hipótese' (1993, p. xxxviii, citados por Button e al., p. 155)

Porem, é claro que um ser humano, por exemplo, seria capaz de reconhecer a existência de tanques em muitos contextos distintos. Não há necessidade de restringir o cenário a um específico, ou sempre determiná-lo.



Conclusão

Notamos neste trabalho que o comportamento inteligente, para a maioria dos conexionistas, resume-se no reconhecimento de padrões. No caso das redes neurais artificiais, tal reconhecimento é produto, dentre outras coisas, da adequação das conexões entre seus nódulos. Ao ser apresentado um determinado padrão à rede, pretende-se que seus nódulos sejam relacionados de tal forma a classificá-lo corretamente. Se o fizer, realizou um comportamento inteligente.

Para a IA, o comportamento, seja do homem ou da máquina, pode ser entendido como uma seqüência de operações básicas. Qualquer cálculo ou resolução de problemas pode ser subdividido em simples instruções, exigindo um pequeno número de habilidades para serem realizadas. Este comportamento é mecânico, caracterizado pelo seguimento de regras e manipulação de símbolos. Se um sistema terminar a realização de um programa e resolver problemas de modo adequado, pode-se dizer que ele realizou um comportamento inteligente.

Vimos ainda que um dos argumentos principais para a atribuição de inteligência a uma máquina de Turing é o famoso teste proposto por este matemático. Nesse teste, tal máquina



será considerada inteligente se for capaz de responder a uma série de questões dirigidas a ela.

Pensamos ser necessária a análise de algo a mais do que as respostas às questões feitas ao sistema submetido ao teste para dotá-lo de inteligência. É preciso, além disto, levar em consideração o seu modo de funcionamento. Concordamos com o argumento de Searle de que não é possível um sistema exibir um comportamento inteligente simplesmente manipulando símbolos sem significado. A compreensão do que se está fazendo é fundamental para um indivíduo apresentar inteligência em seus atos.

O tipo de comportamento apresentado pela máquina de Turing poderia ser comparado ao de alguém que traduz textos do francês para o português, por exemplo, sem prestar atenção no que faz. Seu comportamento resume-se em olhar o dicionário francês-português e traduzir as palavras. Porém, ele não entende o que traduz, embora possa ter pelo menos consciência de estar fazendo uma tradução. O comportamento deste indivíduo caracteriza-se pela manipulação símbolos e seguimento de regras do tipo '*encontre a palavra y no dicionário francês*' ou '*se y é x, então traduza y por x*'. Poderíamos considerar este comportamento inteligente? Acredito que não, pois uma tradução exige mais do que a substituição de palavras. É fundamental a manutenção do significado do texto a ser traduzido. Uma vez que a máquina de Turing funciona como o indivíduo acima, penso que ela não exibe inteligência.

Poderíamos adotar uma postura do tipo *como se* frente ao comportamento da máquina. Desse modo, diríamos que ela apresenta um comportamento como se fosse inteligente. De fato, acreditamos que tal postura é viável, mas perigosa porque podemos passar do plano metafórico ao literal sem dar-nos conta disso. Podemos, de repente, passar a



atribuir genuína inteligência ou estados e faculdades mentais às máquinas sem apresentar uma boa justificção desta atribuição. Talvez isso possa ter acontecido na Ciência Cognitiva, uma vez que muitos de seus pesquisadores analisam o comportamento da máquina como se ele fosse inteligente. Porém, ao notar que os resultados finais do comportamento do sistema artificial comparado ao do ser humano são semelhantes, passam do plano metafórico para o literal. Sendo assim, deixam de afirmar que ela exibe um comportamento como se fosse inteligente, atestando que ela apresenta um comportamento inteligente do mesmo modo do ser humano. Novamente, acreditamos que o problema está no fato de terem esquecido de analisar o funcionamento interno dos sistemas.

Muitas são as caracterizações do conceito de inteligência formulados por diversos pesquisadores de variadas áreas de pesquisas. Porém, cada uma destas visões sobre esse conceito é fragmentada. Cada pesquisador procura apresentar a sua versão, tomando algum aspecto da inteligência. Por isso, vemos a necessidade de uma reunião das diversas caracterizações, tentando relacioná-las o melhor modo possível, buscando uma caracterização um mais razoável do conceito de inteligência.

Julgamos que a interdisciplinaridade na Ciência Cognitiva deve servir também para criar uma unidade com respeito às *opiniões* sobre os diversos aspectos de estudo desta ciência. Talvez seja do filósofo o papel de buscar unir as diversas caracterizações propostas sobre o conceito de inteligência. O trabalho em recortes, como muitas vezes é realizado, ainda não revelou o que é realmente a inteligência. Talvez a postura relacional entre as diversas caracterizações poderia nos proporcionar um caminho melhor em direção à compreensão da natureza da inteligência.



O trabalho na Ciência Cognitiva deixou de lado no estudo da inteligência, e da mente em geral, aspectos do tipo social, cultural e biológico. No caso da inteligência, por exemplo, não é possível querer analisá-la do mesmo modo em culturas distintas. É óbvio que cada cultura tem suas particularidades. Obviamente também será que seus indivíduos se apresentam de maneiras distintas de outras culturas frente ao meio ambiente. Sendo assim, os critérios para a análise da inteligência em uma cultura devem ser distintos dos adotados em outra.

Outro aspecto que acreditamos deva ser levado em consideração é o social. O ser humano é um ser social. Ele aprende, se desenvolve, vive de acordo com a sociedade na qual faz parte. Sendo que os modelos mecanicistas da mente não se comunicam, não vivem em sociedade, falta um componente essencial para a caracterização da inteligência humana na ciência cognitiva.

Não temos a pretensão de exibir uma conclusão crítica a respeito da pesquisa sobre a mente na Ciência Cognitiva. Porém, confessamos que não é possível ainda notar um grande avanço no tocante ao conhecimento da mente.

Nossa posição é a de que, principalmente a respeito da IA, as máquinas de fato não exibem comportamento genuinamente inteligente. Basta analisar o funcionamento de sistemas como o ELIZA, por exemplo, apresentado no quinto capítulo desta dissertação, que foi criado exatamente para mostrar que não existe inteligência nestes sistemas. Do ponto de vista comportamental imediato, parece haver inteligência no comportamento deste sistema. Porém, se travarmos uma série de diálogos com ELIZA, por mais complexo que seja o programa, notaremos que as situações em algum momento tornar-se-ão repetitivas.



Um psiquiatra humano, ao contrário do psiquiatra artificial, é capaz de adequar-se a inúmeras situações de modos distintos, sem seguir regras de modo mecânico.

Não sabemos exatamente o quanto podemos dizer que as máquinas nos revelam algo sobre a inteligência humana. Se não a conhecemos precisamente, como podemos aceitar ou negar que sistemas artificiais nos mostrem sua natureza? Se quisermos descobrir qual é realmente a natureza da inteligência humana, precisamos tomar, literalmente, o próprio homem como objeto de estudo.

É inegável que a inovação tecnológica produzida pela criação de modelos da mente é explícita. Porém, desde um ponto de vista filosófico, é difícil quantificar o progresso no estudo sobre a inteligência e sobre a mente em geral. Se progrediremos!?... Talvez!



Referências bibliográficas

- ALEKSANDER, I. (Ed). *Neural computing architectures: the design of brain-like machines*. Oxford: North Oxford Academic Publishers, 1989.
- BAHM, A. Meanings of intelligence. *Philosophical Studies*, v. 14, p. 151-5, 1968.
- BARR, A., FEIGENBAUM, E. *The handbook of artificial intelligence*. Stanford, CA: Hueristech Press/William kaufman, 1982. 3 v.
- BEALE, R, JACKSON, T. *Neural computing: an introduction*. Bristol: Institute of physics publishing, 1990.
- BERRY, J. Toward a universal psychology of cognitive competence. *International journal of psychology*, v. 19, p. 335-61, 1984
- BINET, A., SIMON, T. *The development of intelligence on children*. New York, 1916.
- BLOCK, N. Troubles with functionalism. In: *Minnesota Studies in the Philosophy of Science*, v. 11, p. 261-325, 1978.
- _____. (Org.). *Readings on philosophy of psychology*. Cambridge: Harvard University Press, 1980.
- BODEN, M. *Artificial intelligence and natural man*. 2. ed. New York: Basic Books, 1987.
- _____. (Ed) *The philosophy of artificial intelligence*. Oxford: Oxford University Press, 1990.
- _____. (Org.) *Dimensões da criatividade*. Trad: Pedro Theobald. Porto Alegre: Artes Médicas Sul, 1999;
- BORING, E. *A history of experimental psychology*. New York: Appleton-Centery-Crofts, 1950.
- BUTTERWORTH, G. Inteligência infantil. In: KHALFA, J. (Org.) *A natureza da inteligência*. Trad: Rouanet, L. São Paulo: UNESP, 1997.
- BUTTON, G, et al. *Computadores, mentes e conduta*. Trad: Ferreira, R. São Paulo: UNESP, 1997.
- CAMPBELL, J. *The improbable machine: what the upheavals in artificial intelligence research reveal about how the mind really works*. New York: Simon & Schuster, 1989.
- CAUDILL, M., BUTLER, C. *Understanding neural networks: computer explorations*. Massachusetts: MIT Press, 1992. 2 v.
- _____. *Naturally intelligent systems*. Massachusetts: MIT Press, 1993.
- CECI, S. *On intelligence.... more or less: a big-ecological treatise on intellectual development*. Englewood Cliffs, NJ: Prentice Hall, 1990.
- COPELAND, B. J. *Artificial intelligence: a philosophical introduction*. Cambridge: Blackwell, 1995.
- CHARNIAK E. *Naturally intelligent systems*. Massachusetts: MIT Press, 1993.
- CHARNIAK, E., McDERMOTT, D. *Introduction to artificial intelligence*. Addison-Wesley, 1985.
- CHURCHLAND, P. Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, v. 78, p. 67-90, 1981.
- _____. *Matter and consciousness*. Cambridge: MIT Press, 1984.
- CUTLAND, N. *Computability: an introduction to recursive functions*. Cambridge: Cambridge University Press, 1992.



- DAVIS H. *Computability and unsolvability*. New York: McGraw-Hill, 1958.
- DENNETT, D. *Brainstorms: philosophical essays on mind and psychology*. Cambridge: Cambridge Massachusetts. Press, 1978.
- DESCARTES, R. *As paixões da alma*. Trad: Gilles-Gaston Granger. São Paulo: Nova Cultural, 1996a. p. 129-240 (Os Pensadores).
- _____. *Meditações*. Trad: Gilles-Gaston Granger. São Paulo: Nova Cultural, 1996b. p. 241-338 (Os Pensadores).
- DREYFUS, H. *What computers still can't do*. New York: Harper & Row, 1993.
- DUPUY, J.P. *Nas Origens das ciências cognitivas*. Trad: Ferreira, R., São Paulo: UNESP, 1996.
- ECKARDT, B. von *What is cognitive science?* Massachusetts: MIT Press, 1995.
- FEIGENBAUM, E., FELDMAN, J. *Computers and thought*. New York: McGraw-Hill Book Company, 1968.
- FELDMAN, J., BALLARD, D. Connectionist models and their properties. In: *Cognitive Science*, v. 6, p. 205-254, 1982.
- FELDMAN, J. et al. *Parallel distributed processing*. 3. Ed. Massachusetts: MIT. Press, 1987, 2v.
- FODOR, J. *Psychological explanation*. Massachusetts: MIT Press, 1968.
- FREGE, G. *The foundations of arithmetic*. Trad: Austin, J. Oxford: Oxford University Press, 1950.
- GARDNER, H. *A Nova ciência da mente*. Trad: São Paulo: USP, 1995.
- GARDNER, H., KORNHABER, M., WAKE, W. *Inteligência: múltiplas perspectivas*. Trad: Maria A. V. Veronese. Porto Alegre: ArtMed, 1998.
- GOUDENOUGH, T. *La inteligencia del niño pequeño*. Ed. Paidós, 1965.
- GREGORY, R. Vendo a inteligência. In: KHALFA, J. (Org.) *A natureza da inteligência*. Trad: Rouanet, L. São Paulo: UNESP, 1997.
- GESEHWIND, N. *Selected papers*. Boston: Reidel, 1974.
- GROSSBERG, S. *The adaptive brain I: cognition, learning, reinforcement and rhythm, and the adaptive brain II: vision, speech, language, and motor control*. Amsterdam: Elsevier/North-Holland, 1986.
- GUILFORD, J. P. *The nature of intelligence*. Maidenhead: Mc-Graw Hill, 1967.
- GUNDERSON, K. *Mentality and machines*. New York: Anchor Books, 1971.
- HAMPSHIRE, S. Critical notice of Ryle, the concept of mind. *Mind*, v.59, n. 234, p.237-55, 1950.
- HEBB, D. *Organization of behavior*. New York: John Wiley, 1949.
- HECHT-NIELSEN, R. *Neurocomputing*. Addison-Wesley Publishing Company, 1991.
- HILBERT, D., ACKERMANN, W. *Elementos de lógica teórica*. Madrid: Editorial Tecno, 1962.
- HOPFIELD, J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.* v. 79, p. 2554-8, 1982.
- HORN, J. M. Remodeling old models of intelligence. In: WOLMAN, B.(Ed). *Handbook of intelligence*. New York: Wiley, 1985.
- HUSH, R., HORNE, B. Progress in supervised neural networks. *IEEE Signal Processing Magazine*, Jan. 1993.
- KHALFA, J. (ED) *A natureza da inteligência*. Trad: Rouanet, L., São Paulo: UNESP, 1997.
- KLEENE, S.C. *Introduction to metamathematics*. Princeton: Van Nostrand, 1952.
- _____. Representations of events in nerve nets and finite automata. In: SHANNON, C., MCCARTHY, J. (Orgs) *Automata Studies*. 2 ed. Princeton: Princeton University Press, 1956.
- KOHONEN, T. *Self organization and associative memory*. Springer-Verlag, 1987.
- LASHLEY, K. In search of the engram. *Symposium of the society for experimental biology*. p. 454-482, 1950.
- _____. Cerebral organization and behavior. In: SOLOMON, H., COBB, PENFIELD, S. (Orgs.). *The brain and human behavior*. Baltimore: Williams and Wilkins Press, 1956.

- LENAT, D. et al CYC: toward programs with common sense. *Communications of the association for computing machinery*, v. 33 n.8, p. 30-49, 1990.
- LENNENBERG, E. *Biological foundations of language*. New York: Wiley, 1967.
- LEWIS, D. Psychological and theoretical identification. *Australian Journal of Philosophy*, v. 50, p.249-58, 1972.
- LIPPMANN, R. An introduction to computing with neural nets. *IEEE ASSP Magazine*, p. 4-23, 1987.
- McCLELLAND, J., RUMELHART, D. *Explorations in parallel distributed processing*. Cambridge: MIT Press, 1988.
- McCULLOCH W., PITTS, W. A logical calculus of the ideas immanent in nervous activity. In *Bulletin of Mathematical Biophysics*, v. 5, p. 115-33, 1943.
- _____. *Embodiments of mind (collected Papers)*. Cambridge: MIT Press, 1965.
- McCURDOCK. *Machines who thinks*. S. Francisco: W.H. Freeman, 1979.
- MCKINTOSH, N. Inteligência em evolução. In: KHALFA, J. (Org.) *A natureza da inteligência*. Trad: Rouanet, L. São Paulo: UNESP, 1997.
- MINSKY, M. L. *Computation: finite an infinite machines*. New Jersey: Prentice-Hall Inc. Englewood Cliffs, 1967.
- _____. Matter, mind and models. In: MINSKY, M. (Org.) *Semantic information processing*. Cambridge: MIT Press, 1968.
- MINSKY, M. L., PAPERT, S. *Perceptrons*. Cambridge: MIT Press, 1988.
- MUÑOZ, J. *Dialogo sobre inteligencia artificial*. Logos, Mexico, 1988.
- NEWELL, A. *United theories of cognition*. Harvard: Harvard University Press, 1990.
- NEWELL, A., SHAW, S., SIMON, H. Elements of a theory of GPS. *Psychological Review*, v. 65, 1958
- NEWELL, A., SIMON, H. The simulation of human Through. *Current Trends in Psychological Theory*, n. 1, p. 152-79, 1961.
- _____. GPS: a program that simulates human thought. In: FEINGEMBAUM. E, FELDMAN J. (Eds.). *Computers and through*. New York: McGraw-Hill, p. 279-93, 1968.
- _____. *Human problem solving*. Prentice-Hall, 1972.
- PENROSE, R. *A mente nova do rei: computadores, mentes e as leis de fisica*. Trad: Waltensir Dutra, 2. ed. Rio de Janeiro, 1993.
- _____. *Shadows of the mind: a search for the mystery science of consciousness*. Oxford: Oxford University Press, 1994
- _____. Inteligência matemática. In: KHALFA, J. (Org.) *A natureza da inteligência*. Trad: São Paulo: UNESP, 1997.
- PINKER, S. *Como a mente funciona*. Trad: Laura T. Mota, São Paulo: Companhia das letras, 1998
- PLACE, U. T. Is Consciousness a Brain Process? *British Journal of Psychology*, v. 47, p. 44-50, 1956.
- POST, E. Formal reductions of the general combinatorial decision problem. *American Journal of Mathematics*. v. 65, p. 197-268, 1943.
- PUTNAM, H. *Brains and behavior*. In: BUTLER, R.(Org.) *Analytical philosophy*. Oxford: Basil Blackwell, 1963.
- _____. The mental lives of some machines. In: CANTANEDA, H.(Org.) *Intentionality, Minds and Perception*. Detroit, MI: Wayne State University Press, 1967.
- _____. Philosophy and our mental life. In: PUTNAM, H. *Mind, Language and Reality: Cambridge University Press*, (1973) 1975. (Philosophical papers, v. 2)
- PYLYSHYN, Z. *Computation and cognition: toward a foundation for cognitive science*. Cambridge: Cambridge Massachusetts Press, 1984.
- ROGERS, H. *Theory of recursive functions and effective computability*. McGraw-Hill, 1947.

- RORTY, R. Mind-body identity, privacy and categories. In: *Review of metaphysics* v. 29, n. 1, p. 24-54, 1965.
- ROSENBLATT, F. *Principles of neurodynamics*. Spartan Books, 1962.
- RUSSELL, B., WITEHEAD, A. *Principia mathematica*. Cambridge: Cambridge University Press, 1925.
- RYLE, G. *The concept of mind*. New York: Barnes and Noble, 1949.
- SEARLE, J. *Mente, cérebro e ciência*. Trad: Morão, A. Edições 70, 1984.
- _____. *A redescoberta da mente*. Trad: Ferreira, E. São Paulo: Martins fontes, 1997.
- _____. *O mistério da consciência*. Trad: André Pinheiro Uema e Vladimir Safatle, Rio de Janeiro: Paz e Terra, 1998.
- SCHANK, R., ABELSON, R. *Scripts, plans, goals understanding*. Erlbaum, Hillsdale, 1977.
- SCHANK, R., BIRNBAUM, L. Aumentando a inteligência. In: KHALFA, J. (Org.) *A natureza da inteligência*. Trad: Rouanet, L. São Paulo: UNESP, 1997.
- SHANKER, S. Turing and origins of AI. *Philosophical Mathematics*. v. 3, p. 52-85, 1995.
- SHEPHERD, G. *Neurobiology*. Oxford: Oxford University Press, 1983.
- SHEPHERSON, J., STURGIS, H.E. Computability of recursive functions. *Journal Association Comp. Mach.* v. 10, p. 217-55, 1963.
- SIMON, H. *The new science of management decision*. New York, 1960.
- SMART, J. Sensations and brains process. *Philosophical review*. v. 68: p. 141-56, 1959.
- SMOLENSKY, P. Connectionist AI, symbolic AI and the brain. *Artificial Intelligence Review*. v.1, p. 95-109, 1987.
- STERNBERG, R. *What is Intelligence?* Norwood, NJ: Ablex, 1986.
- STERNBERG, R., WAGNER, R. The geocentric view of intelligence and job performance is wrong. *Current directions in psychological science*. v. 2, n. 1, p. 1-5, 1993.
- STICH, S.P. *From folk psychology to cognitive science: the case against belief*. Cambridge, MA: MIT Press, 1983.
- TERMAN, L. Intelligence and his measurement. *Journal of Educational Psychology*. v. 12, n. 3, p. 127-33, 1921.
- THAGARD, P. *Mente: introdução à Ciência Cognitiva*. Trad: Maria R. Hofmeister. Porto Alegre: ArtMed, 1998.
- THORNDIKE, E. *Educational psychology*. New York: Teachers college, Columbia University, 1913.
- TURING, A. On the computable numbers with an application to the Entscheidungs problem. In: DAVIS, M. *The undecidable*. New York: Raven Press, 1936.
- _____. Computing machinery and intelligence. In: FEIGENBAUM, E., FELDMAN, J. (Eds.) *Computers and thought*. New York: McGraw-Hill, p. 11-35, (1950) 1968.
- WAGMAN, M. *AI and human cognition*. New York: Praeger, 1991.
- WASSERMAN, P. *Neural computing: theory and practice*. New York: Van Nostrand Reinhold, 1989.
- WASON, P. C. Regression in reasoning. *British Journal of Psychology* v. 60 n. 4, p. 471- 80, 1969.
- WATSON, J.B. Psychology as a behaviorist views it. *Psychological Preview*. v. 20, p.158-77, 1913.
- _____. *Behaviorism*. Nova York: Norton Press, 1925.
- WEIZENBAUM, J. *Computer power and human reason: from judgment to calculation*. San Francisco: Freeman, 1976.
- WINOGRAD, T. *Understanding natural language*. New York: Academic Press, 1972.



 **ILEUS**
ENCADERNADORA
Fone (014) 4.3.66.66 - Marília-SP

