3. A ciência cognitiva em sua fase inicial: contexto epistêmico

Marcos Antonio Alves
Alan Rafael Valente

SciELO Books / SciELO Livros / SciELO Libros

ALVES, M. A., and VALENTE, A. R. A ciência cognitiva em sua fase inicial: contexto epistêmico. In: *O estatuto científico da ciência cognitiva em sua fase inicial*: uma análise a partir da Estrutura das revoluções científicas de Thomas Kuhn [online]. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2021, pp. 89-128. ISBN: 978-65-5954-052-5. Available from: http://books.scielo.org/id/w2nq4/pdf/alves-9786559540525-06.pdf. https://doi.org/10.36311/2021.978-65-5954-052-5.

3

A ciência cognitiva em sua fase inicial: contexto epistêmico

Apresentação

Neste terceiro capítulo do livro expomos as bases teóricas e epistemológicas que habitavam a ciência cognitiva em sua fase inicial. Para isso, dividimos o capítulo em quatro seções. Na primeira seção apresentamos os primeiros anos da ciência cognitiva, após o fim da cibernética, tomando como referência o Relatório do Estado de Arte solicitado pela Fundação Sloan. Na segunda seção, a fim de evidenciar as suas bases metodológicas, expomos alguns conceitos básicos dessa área, principalmente os de modelo e representação, buscando, como em todo o texto, expor razões para o estabelecimento de seu estatuto científico. Baseados neste intuito, nas duas **últimas** seções apresentamos duas das principais perspectivas daquele momento que poderiam se configurar como paradigmas dominantes. A primeira delas é o cognitivismo, tratado na terceira seção. Também chamada de Inteligência Artificial, em seu sentido mais forte, essa vertente conjectura que a cognição funciona de maneira idêntica ou semelhante aos procedimentos computacionais, podendo ser modelada por máquinas do tipo Turing. Na quarta seção examinamos as bases de outra grande concepção, o conexionismo, também chamado de Redes Neurais Artificiais. Essa vertente compreende que a cognição é resultado do processamento, distribuído e em paralelo, feito pelas várias unidades simples que compõem uma rede neural. Seriam essas posições

duas candidatas a paradigmas, seriam teorias distintas de um paradigma apenas ou nenhuma dessas alternativas? Esperamos estabelecer as bases desta discussão que será realizada nas considerações finais.

3.1 A ciência cognitiva após a Cibernética

No capítulo anterior procuramos mostrar alguns elementos históricos que propiciaram o surgimento da ciência cognitiva, visando refletir sobre seu estatuto científico nesta fase. Entretanto, como também procuramos mostrar, ela ainda não havia se estabelecido como uma área de pesquisa formalmente constituída. Ainda que em gérmen, os pesquisadores trabalhando em torno do problema da inteligência ou dos processos cognitivos, mais genericamente falando, foram dando forma a essa nova área de pesquisa incipiente.

Existe um consenso na história da ciência cognitiva de que esta foi reconhecida oficialmente em 1956. De acordo com Miller (1979), a comunidade de cientistas cognitivos ganhou vida no Simpósio sobre Teoria da Informação realizado no MIT, de 10 a 12 de setembro de 1956. O segundo dia desse encontro se destaca, para Miller (1979), graças à exposição de alguns trabalhos de grande impacto. Um deles, apresentado por Allen Newell e Herbert Simon, intitulado "Logic Theory Machine", focalizava a primeira prova completa de um teorema executado em uma máquina computadora. O segundo trabalho de grande impacto foi apresentado pelo linguista Noam Chomsky, intitulado "Three Models of Language". Neste texto, Chomsky mostra que um modelo de linguagem derivado da teoria da informação, proposto por Shannon e Weaver (1949), não poderia, de forma alguma, ser aplicado com êxito à "linguagem natural". Assinala Gardner (1996, p. 44) sobre as considerações de Miller:

> Saí do simpósio com uma forte convicção, mais intuitiva que racional, de que a psicologia experimental humana, a linguística teórica e a simulação computacional de processos cognitivos eram todas partes de um todo maior, e de que o futuro veria uma crescente elaboração e coordenação de seus interesses comuns [...] Eu venho trabalhando por uma ciência cognitiva há aproximadamente vinte anos, tendo começado antes de saber como chamá-la.

Como já assinalamos no capítulo anterior, em meados da década de 1950 os neurocientistas estavam começando a registrar impulsos de neurônios individuais do sistema nervoso. No MIT, a equipe de pesquisa de McCulloch, dirigida pelos neurofisiologistas Jerome Lettvin e Humberto Maturana, havia feito um registro do funcionamento da retina de uma rã. Eles haviam conseguido mostrar que os neurônios são sensíveis a formas extremamente específicas de informação, como, por exemplo, pequenos pontos escuros semelhantes a insetos, os quais se moviam através de seu campo perceptivo.

Dentre outras descobertas nos ramos da antropologia e da neurociência, ainda em 1956, um grupo de cientistas, com formação em matemática e lógica e interessados nos problemas dos computadores, reuniu-se no Dartmouth College para discutir seus trabalhos. Nessa faculdade, estava concentrada a maior parte dos cientistas desenvolvendo suas investigações com o que viria a ser denominado "Inteligência Artificial" (IA). Nesse grupo estavam, inclusive, os que geralmente são considerados os pais fundadores da IA: John McCarthy, Marvin Minsky, Allen Newell, Noam Chomsky e Herbert Simon. Durante o encontro, foram examinadas ideias para programas que iriam solucionar problemas, reconhecer padrões, raciocinar logicamente, tendo sido determinadas as principais questões a serem debatidas nos anos seguintes. A despeito de não ter emergido qualquer síntese dessas discussões, os participantes pareciam ter estabelecido uma espécie de grupo permanente ou comunidade científica, centrada nos *campi* do MIT, de Stanford e de Carnegie-Mellon, diz Gardner (1996). Proeminente crítico desta abordagem, Varela (1991, p. 29, grifo do autor) afirma que

> A principal ideia que viria a impor-se no decorrer desta conferência foi o facto de a inteligência (inclusive a inteligência humana) se aproximar de tal forma daquilo que, intrinsecamente, é um computador e que a cognição pode ser *definida* pela *computação* de representações simbólicas.

A década de 1960 foi caracterizada pelo afloramento das sementes plantadas nos anos 1950 em discussões como as realizadas nas Conferências Macy. Fontes governamentais e privadas forneceram apoio financeiro significativo para o desenvolvimento da ciência cognitiva. Duas das figuras

principais para a consolidação dessa nova área, Jerome Bruner e George Miller, fundaram, em Harvard, o Centro de Estudos Cognitivos. Embora os projetos e produtos reais desse centro provavelmente não tenham sido indispensáveis para a vida desta área de pesquisa, durante aquele período praticamente não havia nenhuma pessoa jovem trabalhando nesse campo que não tivesse sido influenciada pela presença desse centro, pelas ideias que eram debatidas por lá e pela forma como elas eram implementadas em pesquisas subsequentes.

De acordo com Gardner (1996), George Miller, seu colega neurocientista Karl Pribram e Eugene Galanter abriram a década com um livro de grande impacto na psicologia e áreas relacionadas, intitulado *Plans and the structure of behavior*. Nele, os autores apresentam um enfoque cibernético do comportamento, em termos de ações, *feedback* e reajustes das ações conforme a retroalimentação. Em linhas gerais, esses três cientistas propunham legitimar, na prática, o abandono da discussão de estímulos e respostas, em favor de modelos mais abertos, capazes de interações propositais. No decorrer da mesma década, surgiam inúmeros "exemplares" decorrentes dos últimos desenvolvimentos. Estes recentes livros e outras publicações serviram como base para a formação dos novos cientistas cognitivos que, por sua vez, consolidavam a nova comunidade científica.

Segundo Gardner (1996), no final de 1969, já era possível pensar em uma abordagem de ciência cognitiva como um todo. Quando o nível de atividade em um campo chega a esse ponto, com uma comunidade aparentemente entusiasmada em torno dos avanços iminentes, geralmente se tem a consolidação de algum tipo de organização.

Em 1970, estava ocorrendo uma série de eventos que, para a ciência cognitiva, ocasionaram grandes avanços, através de uma fundação privada sediada em Nova York – a Fundação Alfred P. Sloan. Ela financiava o que ela mesma denominava "Programas Particulares", nos quais investia um montante considerável de dinheiro, em uma área, por um período de alguns anos, esperando com isso estimular algum progresso significativo no desenvolvimento da área de pesquisa.

No início de 1970, um Programa Particular havia sido implementado nas neurociências. Após esse evento, a Fundação Sloan estava à procura de um campo análogo, de preferência dentro das ciências

naturais, no qual pudesse investir uma quantidade semelhante de recursos. Em 1975, a fundação estava estudando o apoio a programas de vários campos e o Programa Particular na incipiente ciência cognitiva era o principal deles. No ano seguinte, foram promovidas algumas reuniões, nas quais os principais nomes que viriam a ser denominados cientistas cognitivos, oriundos de diferentes ciências, que também viriam a ser chamadas cognitivas, expuseram suas ideias. Embora a ciência cognitiva não estivesse tão madura quanto os neurocientistas da época acreditavam, apontavam os dirigentes da Fundação Sloan:

> [...] ainda assim, há muitas indicações, confirmadas pelas autoridades envolvidas nas investigações iniciais, de que muitas áreas das ciências cognitivas estão convergindo, e, além disto, há uma necessidade igualmente importante de desenvolver linhas de comunicação de uma área a outra, a fim de que instrumentos e técnicas de pesquisa possam ser compartilhados na construção de um corpo de conhecimento teórico [...] (GARDNER, 1996, p. 50).

Existe uma semelhança entre os estímulos fornecidos pela Fundação Macy à geração da cibernética e a iniciativa que a Fundação Sloan teve, com a ciência cognitiva. Depois de deliberado, a Fundação Sloan decidiu dar início a um programa de cinco a sete anos, envolvendo o investimento de até 15 milhões de dólares, o qual acabou sendo elevado para 20 milhões de dólares. Com base nesses financiamentos, dentro de pouco tempo o periódico *Cognitive Science* foi fundado, tendo seu primeiro número publicado em janeiro de 1977; logo em seguida, em 1979, uma sociedade de mesmo nome foi criada. A sociedade promoveu o seu primeiro encontro anual, em La Jolla, Califórnia, em agosto de 1979.

Com o passar do tempo, programas, cursos e boletins informativos estavam espalhados por todo o mundo. No entanto, havia discordâncias sobre o que tratava o campo da ciência cognitiva, quem a atendia, quem a ameaçava de ruir e em que direção ela deveria seguir. Com base nessas controvérsias, a fundação solicitou um relatório, em 1978, com o objetivo de explicitar tais questões. Esse Relatório do Estado de Arte [*State of the Art Report*] foi redigido pelos principais estudiosos do campo, com a colaboração de vários conselheiros. Os autores elaboraram uma figura com

as inter-relações entre os seis campos constituintes, chamada de hexágono cognitivo, que especificava o rol de pelo menos parte das ciências cognitivas:

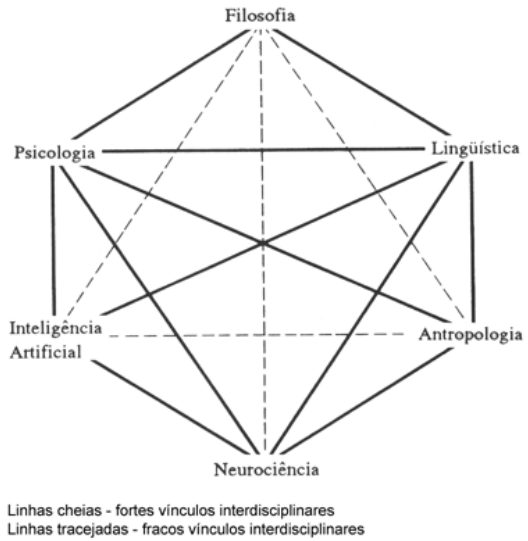

Figura 1. Hexágono cognitivo
Fonte: GARDNER, 1996, p. 52.

Foi realizado um esforço para indicar as conexões entre os campos que atuavam na ciência cognitiva e as possíveis conexões entre eles, expressos, na figura acima, nas linhas cheias e pontilhadas. A imagem e o relatório pretendiam disponibilizar um exame das principais linhas de pesquisa da ciência cognitiva. No entanto, a comunidade, de modo geral, teve uma recepção extremamente negativa do relatório. Segundo Gardner (1996), a reação negativa foi resultado do fato de que cada leitor teve uma interpretação do documento sob o prisma de suas próprias disciplinas. Naquele período, aparentemente, dada a inexistência de um paradigma de pesquisa estabelecido, os cientistas cognitivos tendiam às suas próprias inspirações, baseados em suas próprias convicções, formação acadêmica, intuições oriundas de suas experiências e suas práticas em suas áreas de pesquisa. Em virtude desses fatores, provavelmente, não era possível, em 1978, escrever um documento que obtivesse total apoio, ou mesmo uma concordância substancial dos principais integrantes do grupo

de pesquisadores, o que significaria a constituição de uma comunidade científica.

Em uma primeira análise, parece que, no princípio da história dessa área, muitas disciplinas compunham a base teórica daquilo que até o momento era chamado de ciência cognitiva. Esse é mais um elemento a corroborar a nossa hipótese exposta no final do capítulo anterior de que aquilo que chamamos de ciência cognitiva é decorrente do movimento da cibernética e que a ciência cognitiva não dispõe de um paradigma bem estabelecido em seus momentos iniciais, o que a faz começar como a grande maioria das ciências, por um estágio, na melhor das hipóteses, de pré-ciência. Nestes primeiros momentos, a ciência cognitiva dispunha de uma dinâmica própria que visava, de alguma forma, as aspirações de cada pesquisador ou grupo de pesquisadores. Havia certa discordância entre os primeiros membros oriunda da própria metodologia presente desde o período das Conferências Macy.

Partimos do princípio de que as disciplinas que constituem o momento de transição entre a cibernética e a ciência cognitiva não refletem necessariamente um modelo bem fixado de uma comunidade científica e, por sua vez, esse primeiro momento revela o ambiente das ciências cognitivas, que era um conjunto de disciplinas unidas, com suas especialidades e com o objetivo de investigar as questões atreladas à cognição.

Boa parte dos cientistas cognitivos visava o progresso da comunidade científica a partir das especificidades de suas áreas de pesquisas. Assim, por exemplo, os neurocientistas estavam muito mais preocupados com as estruturas dos neurônios e com os processos e relações neuronais do que com os aspectos fundamentalmente abstratos, subjacentes ao exame da cognição. O desenvolvimento disforme da pretensa comunidade científica suscitou a produção de exemplares que careciam de coerência com as demais estruturas desenvolvidas simultaneamente. Por sua vez, havia exemplares que diziam tratar da ciência cognitiva, mas não abarcavam as elucidações propostas pela antropologia, os estudos desenvolvidos pela linguística ou mesmo compartilhavam dos mesmos pressupostos, sejam metafísicos, sejam filosóficos. Nem todos, por exemplo, concordavam com a perspectiva funcionalista da mente como abordagem subjacente ao

tratamento dos processos cognitivos, que envolvia aspectos internalistas e representacionistas, conforme tratamos a seguir

3.2 Modelos e representações na ciência cognitiva

Como já enunciado no segundo capítulo, uma das características metodológicas centrais da ciência cognitiva na sua fase inicial era o uso de modelos na tentativa de oferecer abordagens explicativas de processos cognitivos. Ademais, boa parte, senão a totalidade dos pesquisadores, defendia a ideia de que os processos cognitivos pressupõem ou envolvem representações mentais.

A concepção e o uso de modelos aparecem com certa constância na atividade científica. Muitos campos, como, por exemplo, a física, a biologia e a astronomia, representam suas entidades ou fenômenos de investigação a partir de esquemas capazes de instaurar uma relação de equivalência ou de isomorfismo com os seus objetos de estudo.

Conforme assinala Dupuy (1996, p. 23): "O modelo científico é uma imitação humana da natureza que o cientista logo toma como 'modelo' – no sentido comum – desta." Os modelos permitem, até certo ponto, com o uso de ferramentas matemáticas, um controle explicativo e preditivo capaz de sugerir novas experiências e formular hipóteses inéditas sobre um dado problema, um fenômeno, uma entidade. Continua Dupuy (1996, p. 27): "Conhecer é produzir um modelo do fenômeno e efetuar sobre ele manipulações ordenadas. Todo conhecimento é reprodução, representação, repetição e simulação."

Os modelos, apesar de imitar o modelado, com frequência apresentam uma performance própria com uma dinâmica "desligada" e baseada na realidade. Em geral, eles costumam ser mais controláveis e fáceis de se manipular do que os fenômenos do mundo real. Há, por isso, um cuidado para evitar que o modelo se torne o objeto exclusivo de análise dos cientistas, podendo se desconectar da relação epistemológica com o fenômeno do mundo investigado. Talvez seja este exatamente um dos pontos da atual ciência cognitiva que, aos poucos, de modo consciente ou não, foi alterando seus objetivos centrais. Atualmente, parece-nos que essa área de pesquisa perdeu um pouco o objetivo epistemológico referente à explicação de processos cognitivos para a criação de autômatos

capazes de reproduzir, simular e emular processos cognitivos. O resultado são os androides ou robôs humanoides construídos por empresas de tecnologias, em poucos anos indistinguíveis do ser humano, pelos menos em sua aparência, física e psicológica. Mas essa é outra história, para outro momento.

Embora os modelos sejam utilizados com frequência na atividade científica, é necessário estar atento no tocante a saber quando efetivamente eles oferecem abordagens explicativas de um fenômeno. Segundo Pessoa Junior (2016, p. 102):

> Pode-se dizer que um modelo matemático ou computacional "captura" a realidade? Tomemos como exemplo a simulação computacional de um furacão. O que está sendo capturado são as relações entre as partes da atmosfera, ou seja, a estrutura dinâmica da realidade. Com a informação armazenada no computador, podem-se prever aproximadamente os efeitos causais de situações reais e também de situações contrafactuais, como por exemplo, o que aconteceria com um avião que entrasse em um furacão.
>
> [...]
>
> O modelo captura apenas a organização ou estrutura do sistema, e não os elementos em si, não a materialidade do sistema representado. Todo modelo tem sua materialidade própria, sejam os dispositivos e circuitos elétricos de um computador, sejam as estruturas celulares em um encéfalo (segundo a concepção materialista do problema mente-corpo). Mas se a materialidade do sistema que representa for distinta da do objeto que é representado, esta distinção estabelece um limite "qualitativo" para a modelagem.

Os modelos também podem apresentar questões ou dificuldades associadas ao significado, os quais dizem respeito à forma como poderiam surgir e em que consistiriam os conteúdos ou significados de sistemas representacionais. Enfim, como tais sistemas poderiam referir-se às coisas do mundo, ou melhor, como sistemas poderiam exibir intencionalidade intrínseca (SEARLE, 1980).

Os dois modelos mais utilizados na ciência cognitiva para simular e explicar os processos cognitivos são as máquinas do tipo Turing e as Redes Neurais Artificiais. Trataremos de ambos mais tarde neste capítulo,

ao falarmos do cognitivismo e conexionismo. Antes disso, enunciamos rapidamente um elemento fundamental no tratamento dos processos cognitivos neste momento: as representações.

Na ciência cognitiva, a noção de representação está relacionada ao binômio e disputa entre o internalismo e o externalismo. De modo geral, segundo o internalismo, a mente ou, para evitar debates e comprometimentos ontológicos no momento, os processos cognitivos estariam, de alguma forma, dentro do indivíduo, enquanto, no externalismo, tais processos se encontrariam, de algum modo, fora dele. Para os internalistas mais radicais, tais como os defensores de abordagens dualistas como a cartesiana, a mente existe independente do mundo externo e ela pode ser compreendida sem qualquer necessidade de recorrermos ao mundo externo.

Na visão dos internalistas radicais, elementos como dor, sentimentos e crenças seriam, em menor ou maior grau, independentes do corpo. Nesse sentido, a mente seria capaz de instanciar estados como uma dor no braço sem necessariamente ter um braço, tal como ocorre em alguns casos documentados pela medicina de "membros fantasmas".

Putnam (1975), embora não seja um externalista radical, fornece um exemplo contrário aos internalistas radicais, pretendendo mostrar, com um exemplo fictício, chamado "Argumento da Terra Gêmea", que os estados mentais não determinam, por exemplo, a referência ou os conteúdos do pensamento. Com este experimento mental, o seu autor procura mostrar que o conteúdo é externo à mente. Na Terra Gêmea, existe uma cópia idêntica a todos os indivíduos de nosso planeta. Nossos clones agem, falam, pensam, acreditam e fazem tudo da mesma maneira que nós. Todas as nossas propriedades internas são idênticas. Nesse exemplo, a palavra "água" é usada para fazer referência à substância idêntica, em todos os sentidos, com aquela que possuímos na Terra original. Entretanto, existe uma particularidade da Terra Gêmea, no sentido de que o líquido chamado "água" não é descrito pela fórmula H^2O, mas por um aglomerado de elementos que o autor descreve como XYZ.

Putnam (1975) procura mostrar, com esse exemplo, que os conceitos não são suficientes para determinar a referência do conteúdo e, embora nossos estados mentais sejam iguais, nós nos referimos a coisas distintas. Podemos notar que, quando eu e minha cópia falamos sobre "água", cada um de nós faz referência a um composto químico diferente.

Na concepção de Frawley (2000), que busca oferecer uma concepção intermediária entre os extremos nesse caso, a divisão feita entre internalismo e externalismo serve apenas para organizar a preferência explanatória de cada pesquisador. Frawley (2000, p. 15, grifo do autor) diz:

> [...] o internalismo e externalismo se referem às ideologias científicas que colocam o maior peso da explicação em unidades seja interna, seja externa ao pensamento. O *internalismo* não pode ser igualado ao nativismo, que é apenas um tipo de explicação internalista, e, também não é o mesmo que computacionalismo, embora tais explicações da mente geralmente enfatizem os fatos dentro do limite mente-mundo. O *externalismo* privilegia o externo em relação ao interno e ocorre em muitas formas – empiricismo, behaviorismo ou alguma escola de pensamento dessa natureza que considera fatores externos à mente. Mas é possível ser um externalista *e* um computacionalista, como atesta qualquer uma das muitas novas teorias da mente contextualizada.

Boa parte dos internalistas acredita na existência de um elemento mediador, chamado representação mental, entre o sujeito e o objeto de conhecimento. Para outros, a relação com o mundo ocorre de maneira direta. Essas duas vertentes são denominadas, respectivamente, representacionistas e não representacionistas.

Para os representacionistas, a representação funciona como uma ponte entre a mente e o mundo. "Uma representação é uma versão modificada do mundo", define Charniak (1993, p. 8, tradução nossa). O que caracteriza uma representação é sua propriedade de ser algo que pode se colocar no lugar de "outra coisa" (como um mapa, por exemplo). A principal característica definidora da ciência cognitiva seria, por conseguinte, a ideia de simuladores e/ou reprodutores de propriedades mentais, e, para que isso se efetive, as noções de representação e computação são fundamentais, durante o processo de modelagem computacional, como salienta Fodor (1980, p. 31, tradução nossa): "Sem representação, não há computação; sem computação, não há modelagem." Nesse sentido, mais do que uma escolha metafísica, a defesa de um internalismo representacionista também significava, de alguma forma, uma necessidade metodológica para o

tratamento de processos cognitivos para uma ala, muito significativa, da ciência cognitiva em sua fase inicial.

No contexto da ciência cognitiva, aos sistemas a que se atribuem representações (os quais podem ser artefatos tanto do cognitivismo quanto do conexionismo), pode-se enfatizar que as representações seriam dotadas de um conteúdo cuja natureza explicativa e ontológica varia dependendo da abordagem. Mais do que isso, a representação, sobretudo, guia o comportamento ou a atividade de um sistema. Haselager (2004, 106) assevera que os adeptos do representacionismo entendem que

> As duas características mais importantes das representações são que elas se colocam no lugar de algo e que o sistema usa as representações com o objetivo de guiar seu comportamento. De acordo com a ciência cognitiva tradicional, então, as representações desempenham um duplo papel: carregam um conteúdo e causam o comportamento. Mesmo se a ciência cognitiva clássica e o conexionismo discordam a respeito do formato das representações, eles têm esse pressuposto em comum.

O uso das representações implica um compromisso coletivo da comunidade científica da ciência cognitiva. Do ponto de vista da matriz disciplinar, o uso das representações é compartilhado pela comunidade de pesquisadores, todavia, a sua aplicação pode ser relativa a cada perspectiva presente no paradigma geral.

As representações são concebidas como um modo abstrato de reter conhecimentos sobre o mundo, por meio de símbolos, esquemas, imagens, ideias. A representação mental é uma espécie de imagem mental de algum objeto ou fenômeno do mundo. Os representacionistas afirmam que podemos reconhecer os objetos do mundo porque deles possuímos algumas representações em nossa mente.

A natureza das representações também é cercada de problemas. Seriam estas baseadas em símbolos bem estruturados ligados a uma rígida articulação sintática e semântica (cognitivismo)? Ou seriam elas representações distribuídas, fundadas em pesos ajustados mediante treinamento (conexionismo)? Seriam ambas? De mais algum tipo além dessas? Sem pretendermos dar conta desse problema da ciência

cognitiva, podemos sublinhar, segundo Thagard (1998), apenas que as estruturas representacionais da Inteligência Artificial, bem como das redes conexionistas, são complementares, em vez de competitivas.

A operacionalidade, por sua vez, sofre da dificuldade em operacionalização do próprio conceito de representação na ciência cognitiva. Ou seja, em qualquer ciência os conceitos devem ser aplicáveis, e em ciência cognitiva, não parece ser claro a quais sistemas deveria se aplicar a noção de representação.

Além da circunstância de que o próprio observador pode gerar problemas, esse problema é ilustrado, segundo Haselager (2004), pelo seguinte exemplo: ao observar a complexidade do caminho de uma formiga na areia da praia, um cientista cognitivo poderia ser tentado a considerar a complexidade da trilha como efeito de representações mentais "na cabeça" da formiga. Imagina-se que o cientista observa representações, quando possivelmente isso é desnecessário.

A relação entre mediação e local de ação da mente cria uma série de possibilidades. Supondo-se a possibilidade de um paradigma da ciência cognitiva, existe uma tendência de que seus adeptos são representacionistas, ou seja, eles acreditam na existência de algo mediador entre o programa e seu ambiente externo. Para eles, a representação consiste em um conjunto de símbolos adquiridos, a partir de alguma mediação, como sensores, os quais são capazes de expressar algo que está acontecendo com eles mesmos ou com o ambiente externo.

Procuramos mostrar nesta seção que os primeiros cientistas cognitivos em geral concordavam com o pressuposto metodológico do uso de modelos para o estudo de processos cognitivos. Também aceitam, de modo semelhante, que a representação exerce um papel essencial tanto do ponto de vista metodológico quanto do ponto de vista epistemológico e até ontológico desses fenômenos. No entanto, as semelhanças, se é que existem realmente, terminam por aqui. As estranhezas podem ser explicitadas na descrição de duas das mais fortes vertentes da ciência cognitiva em sua fase inicial: o cognitivismo e o conexionismo, dos quais tratamos a seguir.

3.3 Cognitivismo

No capítulo anterior mostramos a existência de duas frentes ideológicas na constituição da ciência cognitiva, oriundas de duas obras capitais publicadas 1943: *Behavior, Purpose and Teleology*, de Norbert Wiener, Arturo Rosenblueth e Julian Bigelow, e *A logical calculus of the ideas immanent in nervous activity*, de Warren McCulloch e Walter Pitts. Essas duas obras são o gérmen do que posteriormente seriam as duas principais vertentes da ciência cognitiva em sua fase inicial: o cognitivismo e o conexionismo, dos quais tratamos no restante deste capítulo. Seguindo a abordagem de Kuhn, avaliamos em que medida a existência desses dois programas de pesquisa indicaria que a ciência cognitiva nesse estágio se configuraria como pré-ciência ou se, de alguma forma, poderia ser considerada ciência normal, avaliando se algum desses projetos, uma vez podendo ser considerados paradigmas, estivesse em evidência, ou seja, fosse dominante.

No decorrer de 1956 ocorreu o afloramento da primeira grande corrente da ciência cognitiva: o cognitivismo. Também denominado Inteligência Artificial, funcionalismo lógico-computacional, o cognitivismo, durante os primeiros momentos da ciência cognitiva, contou com o apoio de instituições de prestígio, revistas científicas e com a aplicação do programa em tecnologias.

Segundo Gardner (1996, 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 maneia 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 e McDermott (1985, p. 113, tradução nossa), "IA é o estudo de faculdades mentais por meio do uso de modelos

computacionais. Seu objeto é a mente, entendida como um sistema processador de informação". Já para Dreyfus (1993, p. 39, tradução nossa), um dos críticos dessa 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".

Varela, Thompson e Rosch (1991), outros críticos dessa corrente, assinalam que, nesse momento, o cognitivismo podia ser considerado como o centro ou núcleo da ciência cognitiva. Na visão de Varela, Thompson e Rosch (1991, p. 30, grifo do autor), a ferramenta mais importante do cognitivismo é a metáfora do computador digital:

> [...] cognitivismo consiste na hipótese de que a cognição – incluindo a cognição humana – é a manipulação de símbolos nos moldes daquilo que é executado pelos computadores digitais. Por outras palavras, a cognição é uma *representação mental*: a mente é definida como operando em termos de manipulação de símbolos que representam características do mundo ou representam o mundo como sendo de um determinado modo.

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 eles 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 primeira, 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 essa versão, a mente está para o cérebro assim como o *software* está para o *hardware* do computador. Como afirmam Newell e Simon (1972, p. 19, tradução nossa), "[...] o homem é um sistema de processamento de informação, pelo menos quando está resolvendo problemas", e um computador pode perfeitamente simular tal sistema.

A IA em sua versão 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.

A noção do funcionamento mental comparado ao funcionamento computacional forneceu um meio muito poderoso para se abordar a cognição. O argumento cognitivista concebe que o comportamento inteligente pressupõe a capacidade de representar o mundo de alguma forma. Dessa maneira, o comportamento cognitivo só pode ser explicitado partindo-se do princípio de que os sistemas atuam representando as mais diversas situações do mundo, conforme já discutido no capítulo anterior.

Os cognitivistas não defendem que, se tivéssemos de abrir a cabeça de alguém e olhar o cérebro, encontraríamos símbolos sendo manipulados. Embora seja fisicamente realizado, o nível simbólico não é redutível ao nível físico. Desse modo, os mesmos símbolos podem ser implementados em numerosas formas físicas capazes de instanciá-los. O cognitivismo defende um nível simbólico irredutível e distinto de sistemas físicos na explicação da cognição. Uma vez que os símbolos são elementos semânticos, os cognitivistas supõem um terceiro nível que assinala o semântico e o representacional. Desse modo, um dos problemas principais do cognitivismo está na correlação entre os estados intencionais, como as crenças e os desejos, com as mudanças físicas de um agente. Em linhas gerais, se os estados intencionais possuem propriedades causais sobre os estados físicos, é importante mostrar em que medida esses estados são capazes de determinar o comportamento de um sistema. A noção de computação simbólica, na perspectiva cognitivista, pressupõe que os símbolos têm uma realidade simultaneamente física e semântica.

Essa vertente tende a ser representacionista, visto que seus adeptos pressupõem a existência de algum elemento mediador entre o programa e o ambiente externo. Seja por intermédio de sensores, seja por outros mecanismos, quando a máquina recebe algum símbolo, seja por meio do ambiente, seja por si mesma, é capaz de o reconhecer e fazer referência à sua representação.

Os modelos do cognitivismo para a explicação dos processos cognitivos são as máquinas do tipo Turing. Inicialmente, Turing estava voltado a resolver problemas de computabilidade, mais especificamente, de definir funções computáveis. Para resolver essa e outras questões em torno delas, propôs, em 1936, uma definição de algoritmo. Posteriormente, tal sistema formal foi utilizado tanto como modelo para o estudo de processos cognitivos quanto serviu de base para a construção dos atuais computadores digitais.

As máquinas de Turing são geralmente distinguidas entre finitas e infinitas. Ambas possuem as mesmas especificações técnicas, diferenciando-se basicamente apenas entre suas capacidades de armazenamento de dados. As máquinas finitas são aquelas que apresentam memória limitada, enquanto as infinitas possuem memória potencialmente infinita, sendo capazes de computar funções mais complexas do que as outras, tais como operações aritméticas como a soma de dois números naturais quaisquer.

Turing (1936, p. 231, tradução nossa) assim apresenta a máquina:

> 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 q1, q2, ..., qR 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 Ø(r) que está "na máquina", chamado "quadrado lido". O símbolo sobre ele é denominado o "símbolo lido", 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. Seu comportamento possível em qualquer momento é determinado pela m–configuração qn e o símbolo lido 124 Ø(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. Pode ainda mudar o quadrado que está sendo lido, mas somente deslocando-se um lugar à direita ou à 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 ela é uma máquina computadora.

A máquina de Turing é constituída basicamente por um conjunto finito *S* de símbolos, um conjunto finito *Q* de estados internos, uma fita de memória potencialmente infinita, um conjunto *P* finito de instruções e um agente que realiza as instruções. Dessa maneira, o funcionamento da máquina é determinado conforme as instruções compostas no seu interior, visando realizar alguma função computável. A estrutura das instruções é a seguinte: “”. “” representa o estado atual da máquina, “” representa o símbolo lido da máquina “” representa o novo símbolo da máquina, “D”, “E” e “I” representam a direção da cabeça de leitura da máquina (direita, esquerda ou imóvel), “” representa o novo estado da máquina.

Como descreve Turing (1950), para cada estado da máquina existem instruções fundamentais para iniciar a computação das funções. Dessa maneira, é fundamental que a memória da máquina seja potencialmente infinita. A figura abaixo ilustra a estrutura da máquina de Turing.

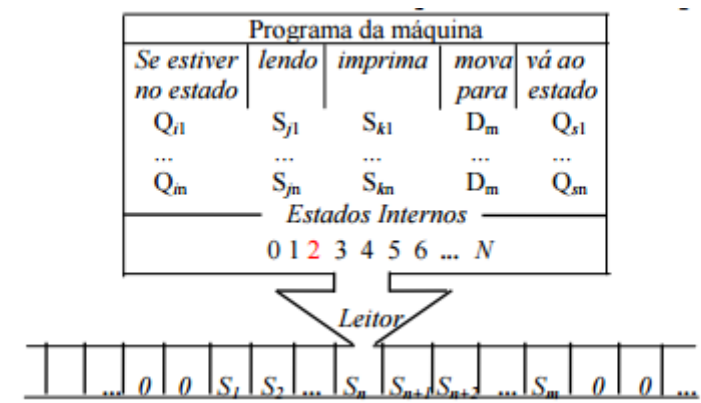

Figura 2. Estrutura geral da máquina de Turing
Fonte: ALVES, 1999, p. 61.

A máquina possui uma fita dividida em quadrados, cada um capaz de armazenar um símbolo. Em cada momento a máquina manipula apenas um único quadrado, podendo apagá-lo, inserir um novo símbolo ou deixá-lo inalterado. Conforme as suas configurações, a máquina pode se "recordar" de um símbolo anteriormente lido. O seu movimento é determinado pelas configurações expressas pelos estados internos e o símbolo lido. O par constituído do símbolo e do estado interno é nomeado configuração.

Abaixo, apresentamos um exemplo particular de uma máquina de Turing, denominada somadora unária. Ela soma dois números M e N na notação unária, ou seja, com base 1, no qual 0 = 0, 1 = 1, 2 = 11, 3 = 111 etc.

```
0-0-0-D-2     1-1-1-D-1     2-A-0-D-2
0-1-0-D-1   1-A-1-D-2
```

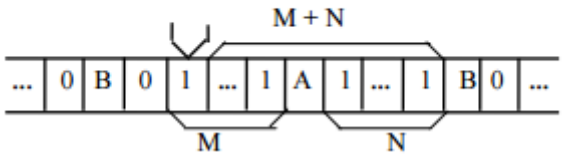

Figura 3. Soma entre M + N, conforme o modelo da máquina de Turing.
Fonte: ALVES, 1999, p. 64.

Conforme Alves (1999), a computação da soma inicia-se a partir do estado 0, lendo o primeiro símbolo à esquerda de M. Caso M seja 0, a máquina move o leitor para a direita e vai ao estado 2, como indicado pela primeira instrução da esquerda para a direita do programa acima. Após isso, substitui A por 0, move o leitor para a direita, como dito na quinta instrução. O leitor vai ficar no estado dois lendo 1 ou 0 e nenhuma das instruções de seu programa possui 0–0 ou 0–1 como configuração. 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, indo ao estado 1, como manda a segunda instrução. Assim, o leitor busca a letra A, seguindo a terceira instrução. Substitui-se esta letra pelo número 1. Isso é o que manda a quarta instrução. A resposta será a sequência de números 1's que restou na fita.

Já com o objetivo de tratar da inteligência e não mais propriamente de questões de computabilidade, Turing (1950) formula as primeiras discussões envolvendo as aplicações da informação a uma teoria

adequada que pretende analisar a inteligência e o pensamento, segundo uma concepção mecanicista formal. A sua proposta consiste na defesa do pressuposto de que "[...] pensar é calcular" (TURING, 1950, p. 436, tradução nossa). Essa tese parte do pressuposto de que, se aceitarmos que a atividade de pensar é, preponderantemente, constituída pela manipulação eficiente de informação/dados, por meio de regras lógicas, poderíamos sustentar que computadores também seriam capazes de pensar, se fossem capazes de tal manipulação. A tese de que "pensar é calcular" está pautada em um famoso jogo proposto por ele, chamado "jogo de imitação", ou teste de Turing. O jogo, assevera Turing (1950, p. 433, tradução nossa):

> 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 é o homem e qual é 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".

O objetivo do homem é, justamente, tentar induzir o interrogador a fazer uma identificação errada. Já o objetivo da mulher é ajudar o interrogador a fazer a escolha certa. Sua melhor estratégia é possivelmente responder a todas as perguntas de maneira correta. Como ressalta Turing (1950, p. 433, tradução nossa), ela poderia dizer coisas como: "Eu sou a mulher, não o escute!". Mas isso seria inútil, porque o homem também poderia dar uma resposta semelhante.

O objetivo do jogo é avaliar em que medida, através de perguntas e respostas, ao substituir um dos interlocutores por uma máquina, ela seria capaz de enganar um interrogador. Ao final, se a máquina fosse capaz de responder às perguntas do interrogador e conseguisse confundi-lo, de sorte que ele julgue que as respostas foram dadas por um ser humano, então ela passaria no teste, ou seja, ela teria capacidade de pensar. Turing (1950) destaca que poderíamos substituir a pergunta "Podem as máquinas pensar?" por "Existem computadores digitais imagináveis com bom desempenho no jogo de imitação?". Segundo sua concepção, ambas as perguntas equivalem à ideia de que um computador digital, programado de forma adequada, pode ser preparado para desempenhar satisfatoriamente o papel do jogador A no jogo de imitação.

A máquina de Turing é um modelo abstrato mecânico capaz de manipular símbolos segundo regras específicas. Em cada instante há apenas um símbolo sendo manipulado. A máquina, a partir da sua configuração algorítmica, pode recuperar símbolos lidos anteriormente. Seus procedimentos possíveis são resultantes da relação entre os símbolos lidos e suas configurações. Ela computa funções, de modo que, fornecidos os elementos do domínio de uma função desejada, ela encontrará uma resposta adequada, se bem construída, de acordo com os padrões preestabelecidos no momento de sua configuração.

Para Putnam (1960), o desenvolvimento das noções da máquina de Turing e a invenção do computador ajudaram a resolver ou a dissolver o clássico problema da relação mente-corpo. O respectivo problema clássico faz referência ao fato de que os fenômenos mentais parecem ser qualitativa e substancialmente diferentes dos corpos que parecem instanciá-los.

Na concepção funcionalista, subjacente à proposta de boa parte dos cognitivistas, os diferentes programas, em computadores iguais ou diferentes, poderiam ser executados em diferentes estruturas adequadamente construídas para processar informações. Os funcionalistas entendem que a mente é um sistema processador de informações que manipula os símbolos de entrada (*inputs*) do sistema gerando uma saída (*output*), seguindo certas regras lógicas. Assim, o conjunto de operações lógicas (*software*) poderia ser descrito independentemente do *hardware* específico no qual ocasionalmente houvesse sido instanciado. A analogia dos computadores com o sistema humano sugere que o seu cérebro corresponde ao *hardware*, enquanto os padrões de processamento de informações, ou seja, o pensamento, fazem referência ao *software*. Além disso, os seres humanos, não menos que os computadores, armazenam programas, e as mesmas linguagens simbólicas podem ser invocadas para descrever programas de ambas as entidades.

Um dos objetivos principais da IA é construir modelos computacionais capazes de simular comportamentos que, se fossem realizados por seres humanos, seriam inequivocamente considerados inteligentes. Segundo Feigenbaum e Feldman (1968, p. 3, tradução nossa), 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 e Simon (1958), Putnam (1967) e 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 esses cientistas, é possível a construção de entidades inteligentes fazendo com que elas realizem comportamentos inteligentes. Segundo Schank e 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 ideia desses 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 do melhor modo possível. O comportamento inteligente, dizem Feigenbaum e Feldman (1968, p. 6, tradução nossa),

> [...] 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.

Pesquisadores como Newell, Shaw e Simon (1958) procuraram simular processos cognitivos segundo os quais poderia ser mais certa a possibilidade de estudo. Esses 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 (1993).

A crítica ao estudo do pensamento matemático estende-se aos outros tipos de pensamento, segundo Penrose (1993, p. 112), 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.

Outra crítica bastante devastadora, pelo menos em termos epistemológicos, ao projeto cognitivista foi feita por John Searle. O argumento criado por este filósofo para refutar a ideia da possibilidade de 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). Searle (1984, p. 37) ironiza:

> 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, um sistema biológico com certas características físicas que permitem a emergência de fenômenos mentais, de processos cognitivos. Sistemas puramente formais não são capazes de originar ou possuir consciência e intencionalidade. Uma máquina do tipo Turing, conclui Searle (1984, p. 40), "[...] não pode ser uma mente porque esta possui mais do que uma estrutura formal, possui um conteúdo".

Searle (1984) 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 (1984), 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 então afirma que o comportamento do indivíduo do quarto e o de um nativo chinês são praticamente indistinguíveis. Porém, o indivíduo do quarto não entende uma só palavra do chinês. Assim, o comportamento deste indivíduo não pode ser considerado inteligente 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".

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 consiga simular uma dor, por exemplo, um computador de fato não tem essa 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).

Na concepção de Searle (1984), processos cognitivos são causados pelo cérebro, 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 efetivamente apresentar processos cognitivos e não podem ser considerados bons modelos para a sua explicação. A seguir apresentamos a vertente conexionista, alternativa ao cognitivismo, que procura considerar os aspectos físicos e biológicos de um sistema no tratamento explicativo dos processos cognitivos.

3.4 Conexionismo

O conexionismo é a outra grande vertente da ciência cognitiva em seus primeiros passos. É também denominado de Redes Neurais Artificiais, PDP (processamento distribuído em paralelo) ou funcionalismo neurocomputacional, por muitos de seus expoentes, além de McCulloch (1965), já exposto anteriormente. Alguns dos principais nomes ligados a essa área de estudos são Rosenblatt (1962), Hopfield (1982), Kohonen (1987), Rumelhart e McClelland (1986) e Caudill e Butler (1992).

Essa perspectiva funda sua abordagem epistemológica na ideia de que o cérebro é o sistema em que ocorrem ou são possíveis a emergência de processos cognitivos. O cérebro humano é formado por mais de uma centena de bilhão de elementos computadores chamados neurônios. Essa 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 acima citados têm por fim criar sistemas inspirados no cérebro humano para simular e explicar, dentre outras coisas, comportamentos humanos inteligentes. Porém, pesquisadores como Caudill e Butler (1992, v. 1, p. 4) e McCulloch e Pitts (1943, p. 117) reconhecem que as redes neurais artificiais (RNAs) são apenas uma aproximação muito limitada do cérebro humano. Um simples exemplo disso é o fato dos conexionistas chamarem as partes constituintes das RNAs 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 disso, ela é entendida como um conjunto de neurônios relacionados entre si, produzindo estados mentais, originando conhecimento, aprendizagem, comportamento inteligente.

A abordagem conexionista dos processos cognitivos, segundo seus defensores, proporciona um novo modo de pensar sobre percepção, memória, aprendizagem, pensamento e sobre os mecanismos computacionais básicos para o processamento inteligente de informações em geral.

Como já dissemos, as bases do conexionismo foram estabelecidas por McCulloch (1965). Em linhas gerais, a sua ideia 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, conforme ilustra Alves (1999).

Segundo McCulloch e 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 ideia 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 (1996, 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 sinergético 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.

Os conexionistas propuseram essa nova ideia sobre a representação e a computação, inspirados na analogia da estimulação dos neurônios e da ativação difusa. Ao passo que a proposta cognitivista tem como base um processamento de informação em série, a proposta conexionista trabalha com um processamento de informação em paralelo e distribuído, possibilitando fazer mais de uma operação ao mesmo tempo.

A perspectiva conexionista consiste na ideia de que a cognição resulta do processamento coletivo feito pelas várias unidades simples que compõem uma rede neural. Destacam Rumelhart e McClelland (1986, p. 10, tradução nossa):

> Esses modelos [conexionistas] assumem que o processamento de informação ocorre pela interação de um grande número de elementos processadores simples chamados de unidades, cada um enviando sinais excitatórios e inibitórios para os outros.

Como os cognitivistas, grande parte dos conexionistas tendem a ser representacionistas e internalistas. Diferentemente dos cognitivistas, porém, os conexionistas não entendem a representação como um conjunto de símbolos, mas como um padrão de conectividade, ou seja, a partir da relação entre as partes básicas da rede. Assinala Gardner (1996, p. 414):

> Em vez de operações seriais ou computações sobre símbolos ou cadeias de símbolos, em vez de "executivos", "intérpretes" e "unidades centrais de controle", a abordagem PDP [processamento distribuído em paralelo, ou conexionismo] 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".

Segundo O'Reilly e Munakata (2000), o conexionismo consiste na ideia de que, para explicar a cognição, não basta apenas reduzi-la a elementos mais simples, como neurônios ou interconexões decorrentes de parâmetros excitatórios e inibitórios. É preciso explicar como a combinação destes elementos é capaz de produzir os processos cognitivos. Tal qual as engrenagens que interagem no interior de uma máquina, para se compreender o seu funcionamento é importante especificar como elas interagem para produzir fenômenos mais gerais. Assim, para se explicar processamento cognitivo, por exemplo, existe uma necessidade de se entender a forma como ocorre a interação entre o grande número de "elementos simples", chamados de neurônios, com a sua imensa quantidade de conexões, chamadas de conexões sinápticas. As RNAs são marcadas como uma das primeiras vertentes a produzir modelos do sistema nervoso com grau de precisão suficiente para poder se observar o comportamento emergente dos neurônios trabalhando em paralelo.

Segundo Kovács (2006), a origem da teoria das RNAs está atrelada aos modelos matemáticos e aos modelos da engenharia, os quais tomam como base os neurônios biológicos. Ao longo da história de

pesquisa dessa base biológica, observou-se a existência de manifestação elétrica entre os chamados neurônios biológicos. Nas últimas décadas, em decorrência do trabalho de vários pesquisadores, passou-se a compreender os neurônios biológicos como elementos processadores fundamentais do sistema nervoso, compostos de um grande número de entradas, chamadas de conexões sinápticas.

Grosso modo, nos sistemas biológicos de neurônios, os sinais que chegam são pulsos elétricos, denominados impulsos nervosos, cujas sinapses correspondem a regiões eletroquímicas entre neurônios, por onde existe a troca de estímulos por meio de substâncias conhecidas como neurotransmissores. O resultado dessa transferência de estímulos, dependendo do tipo de neurotransmissor, é classificado como uma conexão sináptica excitatória ou inibitória.

As bases para o conexionismo no período de investigação da ciência cognitiva aqui considerado advém da concepção de McCulloch e Pitts (1943), para os quais o sistema nervoso é composto por uma rede de neurônios formados, dentre outras coisas, por um soma e um axônio. O soma do neurônio consiste no seu corpo celular, e o axônio é o cilindro-eixo. As denominadas sinapses são as conexões entre um axônio e a soma de outro neurônio.

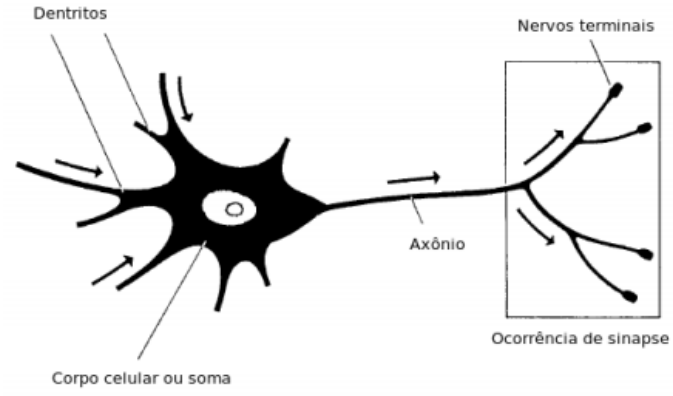

Figura 4. Modelo abstrato do neurônio biológico
Fonte: Adaptado de ARBIB, 2002, p. 4.

Cada neurônio pode receber inúmeros *inputs* de outros neurônios ou do ambiente externo. Entretanto, eles podem produzir apenas uma resposta que é transmitida a outros neurônios ou para o ambiente externo. A resposta do neurônio é enviada pelo axônio através das terminações axonais. As sinapses podem ser divididas entre inibitórias e excitatórias. As sinapses excitatórias auxiliam no disparo, enquanto as inibitórias buscam diminuir a possibilidade do disparo ou mesmo inibi-los totalmente. As conexões entre neurônios no cérebro são feitas a partir da transmissão de substâncias químicas. Os neurônios sempre possuem uma espécie de limiar, cujo estímulo precisa ultrapassar para dar início a um impulso.

A expressão redes neurais artificiais parte da motivação de criar modelos capazes de simular as capacidades do cérebro humano de reconhecer, processar e generalizar dados e padrões. Em geral, esses modelos são utilizados em ambientes nos quais o fluxo de informação muda constantemente. Uma rede neural é um sistema computacional constituído a partir de centenas de unidades básicas que simulam as funções dos neurônios. Esses elementos são interligados, trabalhando em paralelo, para desempenhar uma determinada tarefa. Essas redes de neurônios são responsáveis pelo que chamamos de pensamento, emoção, cognição.

Dentre os primeiros projetos de redes neurais artificiais, McCulloch e Pitts (1943), baseados na ideia do potencial inibitório e excitatório dos neurônios, interpretaram que o seu funcionamento ocorria de maneira semelhante a um circuito binário. Nesse modelo, as conexões entre os neurônios, também chamados de nódulos, são realizadas por meio da transmissão de substâncias eletroquímicas que disparam informações para outros nódulos com base em seu limiar. Estes estímulos, na forma de sinapses, podem ser inibitórios ou excitatórios. Nas RNAs, o limiar e as sinapses são representados a partir de valores numéricos, em vez de substâncias eletroquímicas.

Para McCulloch e Pitts (1943), o disparo dos nódulos pode ocorrer conforme apenas dois estados possíveis, ativado ou desativado. Nas RNAs, o disparo ocorre quando certo valor numérico é atingido pelo nódulo e este valor é determinado por uma fórmula lógica, cujo disparo ocorre apenas se sua fórmula correspondente for verdadeira. O modelo geral dos nódulos parte de algumas pressuposições: $N_i(t)$ significa que um nódulo c_i dispara no tempo *t*. N_i é a *ação* de um neurônio c_i – o tempo é

discreto e pode ser representado por números naturais. Ele é determinado pelas sinapses entre os nódulos: cada sinapse representa um tempo. A *solução* de uma rede $\tilde{N}$ é um conjunto de sentenças que regulam o disparo dos nódulos dessa rede. Apresentadas essas pressuposições, podemos ter as seguintes relações:

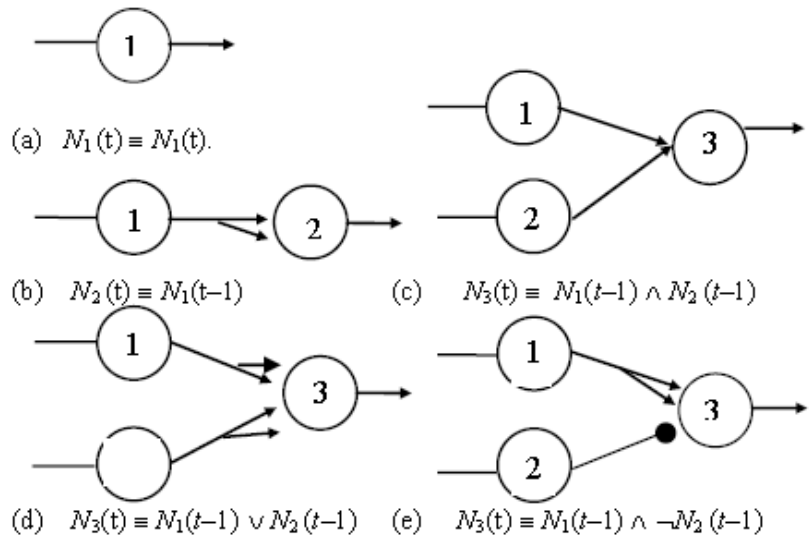

Figura 5. Modelos de redes neurais básicas do sistema apresentado por McCulloch e Pitts.
Fonte: ALVES, 1999, p. 98.

O número no interior de cada nódulo representa a ordem de sua execução, enquanto seu limiar é sempre considerado como sendo 1. Na rede (b) ocorre o que chamamos de demora sináptica. Essas demoras ocorrem, pois a velocidade do impulso produzido pelo disparo dos neurônios pode variar de acordo com o número de sinapses. Assim, expressões com n demoras sinápticas, denotadas por "$t - n$", exigem n nódulos à esquerda do nódulo calculado para tornar o seu disparo equivalente ao valor de verdade de sua proposição lógica.

A ideia principal defendida pelos pensadores conexionistas, em especial McCulloch (1965), é a de que as RNAs são capazes de simular atividades mentais e nervosas, como, por exemplo, a sensação de frio ou calor. Nos sistemas formais, as RNAs utilizam-se de receptores capazes

de medir a temperatura. Sendo estes receptores adequados e capazes de transmitir a informação para outros nódulos com a capacidade de implicar a sensação de calor ou frio a partir de seu limiar, estipula-se que essas sensações podem ser simuladas por essas redes.

Embora o modelo de McCulloch-Pitts tenha uma construção diferente da máquina de Turing, podemos notar algumas semelhanças entre ambos os modelos. Para computar uma função, a rede deve ser construída de tal maneira que, apresentada alguma entrada, a resposta deve ser correta, caso contrário, deve-se produzir uma rede nova. A máquina de Turing e o modelo McCulloch-Pitts não apresentam regras de aprendizagem e, de certa forma, funcionam de maneira algorítmica.

Na década de 1950 surgiram novas RNAs mais sofisticadas. Uma das grandes vantagens dessas redes é que elas podem reconhecer padrões, mesmo aqueles defeituosos, com algum elemento faltando. Como os nódulos estão interligados com inúmeras camadas de outros nódulos, considera-se que cada um deles também receba parte do padrão total de ativação. Essencialmente, o reconhecimento de padrões para os conexionistas consiste em colocar valor de conexões sinápticas por meio da memorização, pela inserção de informações, por exemplos, por analogia ou por exploração e descoberta. Entretanto, não existe uma definição formal de RNA como existe para a máquina de Turing. Cada rede tem sua particularidade e cada definição geral pode variar de modelo para modelo.

Um dos objetivos fundamentais dos conexionistas posteriores ao modelo de McCulloch-Pitts (1943) era fazer com que as RNAs reconhecessem padrões. Existem dois tipos de padrão: o primeiro é caracterizado, segundo Caudill e Butler (1992), como uma propriedade emergente da dinâmica da interação entre os nódulos de uma rede. Os do segundo tipo são aqueles que as redes são capazes de reconhecer, que são um conjunto de elementos que representam algum objeto. Na simulação de uma rede, um padrão é considerado como um conjunto de valores distribuídos metricamente. Dessa forma, caso desejemos apresentar para uma rede uma letra do alfabeto, a respectiva letra seria representada como um conjunto de valores numéricos e cada nódulo trataria de representar um desses valores do padrão de entrada.

Nem sempre a rede reconhece os padrões na primeira tentativa. A aprendizagem geralmente ocorre a partir da presença ou da ausência

de um elemento capaz de testar o reconhecimento de padrões da rede. Uma realimentação explícita significa que em certos intervalos de tempo um sistema assinala os erros e acertos. No caso em que a realimentação não é explícita, a aprendizagem ocorre sem a presença de um sistema externo. Costuma-se chamar esses dois casos de ensino supervisionado e não supervisionado.

No caso da aprendizagem supervisionada, o sistema externo indica explicitamente o que é considerado um comportamento bom ou um ruim. Por exemplo, imaginemos um caso em que seja desejado o reconhecimento entre os números 1 e 2. Apresentam-se para eles letras sucessivamente aos nódulos de entrada. Observa-se qual dos nódulos de saída parece estar mais excitado. Se for o que convencionou a representar o número que foi apresentado, nada deve ser corrigido, caso contrário deve-se mudar os valores das conexões sinápticas a fim de se chegar ao resultado desejado. No caso da aprendizagem não supervisionada, em vez de informar se a resposta dos nódulos foi correta ou não, usa-se um esquema capaz de induzir a rede a responder de maneira semelhante às regularidades apresentadas. A aprendizagem não supervisionada é marcada pela presença de redes auto-organizadas, ou seja, que possuem a capacidade de criar padrões de comportamento não previsíveis e decentralizados e, em alguns casos, em constante adaptação. Segundo Debrun (2009, p. 54), "[...] uma organização ou 'forma' é auto-organizada quando se produz a si própria".

Aos componentes gerais das RNAs podemos elencar a presença de unidades de processamento (nódulos ou neurônios artificiais), estados de ativação destas unidades (limiar), funções de saída que determinam a resposta de cada nódulo, padrões de conectividade que definem a conexão entre os nódulos, regras de propagação, algum dispositivo capaz de representar o meio e os objetos, e, por fim, de regras que permitam a aprendizagem. Embora apresentemos estas sete características gerais, destacamos que não há uma definição precisa sobre as RNAs, pois cada uma possui as suas próprias particularidades, podendo ou não apresentar grande parte dessas características.

Dentre as redes posteriores à rede construída por McCulloch e Pitts, destacamos o modelo conexionista *perceptron.* Criado na década de 1950 por Rosenblatt (1962), seu criador afirmava que o *perceptron* não devia ser comparado ao modelo de McCulloch-Pitts, pois esse modelo

apresentava um funcionamento quase algorítmico. Além disso, não apresentavam uma regra de treinamento para a aprendizagem da rede.

O *perceptron* foi criado fundamentalmente com o objetivo de modelar a percepção visual. O objetivo para essa rede consiste em classificar padrões em duas classes distintas, A ou B. Os nódulos do *perceptron* funcionam de maneira semelhante ao modelo proposto por McCulloch-Pitts. Se o padrão pertence à classe A, o *perceptron* deve disparar. Se o padrão pertence à classe B, ele não deve disparar A.

Segundo Caudill e Butler (1992), a função transferência do *perceptron* é formada por dois passos: o primeiro consiste no cálculo de entrada total encontrado pela função:

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

Onde são vetores peso e *input*. 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,* segundo Caudill e Butler (1992), é o cálculo da resposta do nódulo, encontrada pela seguinte função:

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

Ou seja, o nódulo dispara quando a sua entrada total é maior ou igual ao seu limiar. Caso contrário, não dispara.

O seu treinamento também segue uma respectiva regra:

$$w_{novo} = w_{velho} + \beta_{yx}$$

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

β

Podemos notar, na função acima, que a mudança de peso necessita da resposta desejada do nódulo. Dizemos, assim, que a sua aprendizagem é supervisionada. O treinamento deve ser feito de maneira organizada e ordenada para que uma rede aprenda a reconhecer um determinado grupo de padrões; para isso, será necessário um algoritmo de treinamento. Segundo Caudill e Butler (1992, p. 29, tradução nossa), no caso do *perceptron*, ele é o seguinte:

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.

Após efetuar os cálculos para todos os padrões, deve-se analisar se a rede foi capaz de classificar corretamente cada um deles. Se foi capaz de classificar, ela aprende, caso contrário, deve-se recomeçar a fazer os cálculos até que a rede seja capaz de reconhecê-los.

Podemos observar que o conexionismo, semelhante ao cognitivismo, adota como base uma perspectiva funcionalista sobre o estudo da cognição. Dentre seus pressupostos, consta a relevância do uso de modelos, mas, diferentemente do cognitivismo, o conexionismo julga ser importante a construção de modelos análogos à arquitetura dos neurônios biológicos e que a cognição é resultado do processamento em paralelo e distribuído efetuado por vários nódulos. Entre seus quebra-cabeças está a pretensão em criar modelos capazes de simular a cognição, como, por exemplo, as capacidades de reconhecer, processar e generalizar dados e padrões. Nesse sentido, em termos kuhnianos, podemos dizer que a perspectiva conexionista, embora possua as suas diferenças em relação ao cognitivismo, parece não possuir uma diferença propriamente paradigmática.

O comportamento inteligente para os conexionistas está associado fundamentalmente à capacidade de aprender a reconhecer padrões, 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 ideia 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 das 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.

Para os conexionistas, a aprendizagem nas redes neurais artificiais acontece fundamentalmente através do ajuste dos pesos da rede, ou seja, por meio de um processo de treinamento. Uma vez ajustados os pesos, como afirmam Caudill e Butler (1992, p. 9, tradução nossa), "[...] pode-se dizer que a rede aprendeu." Desse modo, a aprendizagem não tem como princípio fundamental o seguimento de regras e manipulação de símbolos. Diferentemente, ela acontece através do fortalecimento das conexões entre nódulos.

A ideia 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 outro de tal modo que o faça disparar, a conexão a partir da primeira célula com a segunda é fortalecida (Hebb, 1949).

Por ser capaz de realizar comportamentos inteligentes, via reconhecimento de padrões, o projeto conexionista começou com muita euforia na década de 1950. Tal euforia, porém, durou pouco tempo. Minsky e Papert (1969) mostraram que o *perceptron* não podia resolver problemas não linearmente separáveis. Um dos exemplos desses problemas é o do ou-exclusivo da lógica proposicional. Esses dois pesquisadores previram ainda que nem *perceptrons* mais sofisticados com camadas intermediárias seriam capazes de resolvê-los. Isso porque não haveria um bom modo de saber quais são as respostas desejadas para os nódulos das camadas intermediárias da rede. Além de não poder resolver problemas linearmente inseparáveis, Minsky e Papert (1969) mostraram que o *perceptron* não é capaz de separar mais do que duas classes distintas.

Outra forte crítica direcionada à hipótese de que os modelos conexionistas podem ser considerados inteligentes é feita por Dreyfus (1993, p. xxxviii, tradução nossa), para 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". O autor chega a essa conclusão por meio da análise de experiências feitas com redes para reconhecimento de padrões. Um desses trabalhos consistia em fazer com que uma rede neural artificial reconhecesse a presença de tanques de guerra numa floresta. Segundo Dreyfus (1993, p. xxxvi, tradução nossa), o exército tirou um

> [...] certo número de fotografias de uma floresta sem 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 distinguisse 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 se descobriu 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 (1993, p. xxxviii, tradução nossa) 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".

Entretanto, 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.

As dificuldades no projeto conexionista fizeram com que ele tivesse seu progresso abalado por um longo tempo. Abalado, mas não eliminado. Alguns pesquisadores, como Hopfield (1982), Kohonen (1987), Rumelhart e McClelland (1986), continuaram trabalhando firmes em seus propósitos conexionistas. Na década de 1980, criaram redes capazes de resolver problemas não linearmente separáveis e reconhecer objetos, dividindo-os em mais do que apenas duas classes distintas. Eles concordavam com o argumento de Minsky e Papert (1969) de que as redes de multicamadas não eram capazes de resolver esses problemas. Isso se a regra de treinamento fosse a mesma do *perceptron* simples. Porém, se essa regra fosse modificada ou aperfeiçoada, tal argumento seria refutado. Foi o que fizeram os conexionistas: criaram redes cuja regra de treinamento é uma extensão da regra do *perceptron* simples, denominada *backpropagation*. Dentre as redes com tal regra, encontramos o *perceptron* multicamadas e as redes de Kohonen. Os sucessos, alcances e limites dessas novas redes e seu poder explicativo, no entanto, fogem ao escopo deste trabalho, embora sejam ilustrativos da tentativa dos adeptos de uma matriz curricular em

tentar firme e insistentemente resolver seus quebra-cabeças e em não abandonar seu paradigma.

Como procuramos mostrar, o cognitivismo e o conexionismo eram as duas principais vertentes da jovem ciência cognitiva. Seriam essas posições duas candidatas a paradigmas, seriam teorias distintas de um paradigma apenas ou nenhuma dessas alternativas? Sobre isso tratamos a seguir, a fim de encerrar nossa análise neste trabalho.