PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ

ESCOLA DE EDUCAÇÃO E HUMANIDADES

PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU EM FILOSOFIA

SILVIA CENZOLLO PELOI

UMA ANÁLISE DA CAPACIDADE DA CIÊNCIA COGNITIVA EM EXPLICAR O PROCESSO DE DECISÃO SOCIAL

CURITIBA

2015

SILVIA CENZOLLO PELOI

UMA ANÁLISE DA CAPACIDADE DA CIÊNCIA COGNITIVA EM EXPLICAR O PROCESSO DE DECISÃO SOCIAL

Dissertação do curso de Mestrado em Filosofia, vinculado ao Programa de Pós-Graduação Stricto Sensu da Pontifícia Universidade Católica do Paraná.

Orientador: Professor Doutor Kleber Bez Birolo Candiotto

CURITIBA

2015

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central

P392a
2015

Peloi, Silvia Cenzollo
Uma análise da capacidade da ciência congnitiva em explicar o processo de decisão social / Silvia Cenzollo Peloi ; orientador, Kleber Bez Birolo Candiotto. – 2015.
198 f. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2015
Bibliografia: f. 95-98

1. Ciência cognitiva. 2. Neurociência. 3. Altruísmo. 4. Emoções. 5. Filosofia. I. Candiotto, Kleber Bez Birolo. II. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Filosofia. III. Título.

CDD 20. ed. – 100

PUCPR
GRUPO MARISTA

Pontifícia Universidade Católica do Paraná
Escola de Educação e Humanidades
Programa de Pós-Graduação em Filosofia

ATA Nº. 142/PPGF – DEFESA DE DISSERTAÇÃO

Aos vinte e sete dias do mês de agosto de dois mil e quinze, às onze horas na sala de defesa de dissertações da Escola de Educação e Humanidades desta Universidade realizou-se a sessão pública de defesa da dissertação da mestranda **Silvia Cenzollo Peloi** intitulada: UMA ANÁLISE DA CAPACIDADE DA CIÊNCIA COGNITIVA EM EXPLICAR O PROCESSO DE DECISÃO SOCIAL. A Banca Examinadora foi composta pelos professores: Dr. Kleber Bez Birolo Candiotto, Dr. Léo Peruzzo Junior e Dr. João de Fernandes Teixeira. Após a instalação dos trabalhos pelo presidente da banca, professor Kleber Bez Birolo Candiotto, a candidata fez uma exposição sumária da dissertação, em seguida procedeu-se à argüição pelos membros da banca e à defesa da candidata. Encerrada essa fase, os examinadores, em reunião reservada, apresentaram suas avaliações, tendo considerado a cardidata Aprovada em sua defesa de dissertação conforme as notas e o conceito registrados abaixo. Após a proclamação dos resultados, o presidente da banca Arnildo a candidata o título de Mestre em Filosofia. Encerrados os trabalhos às 12 h 45 min. lavrou-se a presente ata que segue assinada pelos membros da Banca Examinadora.

MEMBROS DA BANCA	ASSINATURA	NOTA
Prof. Dr. Kleber Bez Birolo Candiotto		9,5
Prof. Dr. Léo Peruzzo Junior		9,5
Prof. Dr. João de Fernandes Teixeira		9,5
MÉDIA FINAL 9,5	**CONCEITO**	A

CIENTE

Prof. Dr. Ericson Sávio Falabretti
Coordenador do Programa de Pós-Graduação
em Filosofia - *Stricto Sensu*

Para meus pais e meu marido Eduardo.

AGRADECIMENTOS

Agradeço a meu professor orientador, doutor Kleber Bez Birolo Candiotto, pelo auxílio indispensável ao desenvolvimento deste trabalho.

Agradeço a meu marido Eduardo Brindizi Simões Silveira pelo incentivo durante todo o percurso.

Agradeço a minha irmã Cláudia Cenzollo Peloi pelos importantes comentários, que certamente contribuíram para o resultado final.

RESUMO

Esta dissertação tem como objetivo principal analisar as insuficiências do modelo proposto pela ciência cognitiva para explicar a decisão social, entendida como a decisão que afeta tanto a vida daquele que decide quanto as relações que estabelece com outros indivíduos. Para tanto, este estudo partiu do pressuposto de que dois elementos da ciência cognitiva poderiam representar um obstáculo para explicar as decisões sociais: um seria a crença de que o computador é o modelo mais viável para a compreensão da mente humana, e outro a opção metodológica de excluir as emoções dos estudos. Com o objetivo de analisar a viabilidade da metáfora computacional para explicar a decisão social, abordou-se a crítica de John Searle ao modelo computacional proposto pela ciência cognitiva. Para avaliar os problemas que poderiam surgir da opção metodológica de excluir as emoções dos estudos, analisou-se a importância das emoções no processo de decisão social. Para isso, inicialmente foi utilizado o embasamento teórico fornecido por estudos neurocientíficos, sobretudo a hipótese do marcador somático de António Damásio. Em seguida, apresentou-se a psicologia evolucionista, especificamente a teoria do altruísmo recíproco de Robert Trivers, para a análise da relação entre emoção e decisão social. Esta dissertação conclui apresentando uma crítica metodológica à ciência cognitiva, entendendo que a característica metodológica de excluir as emoções dos estudos representa o principal problema da ciência cognitiva para explicar a decisão social.

Palavras-chave: Ciência cognitiva. Decisão social. Intencionalidade. Emoção. Neurociência. Altruísmo Recíproco.

ABSTRACT

This essay aims to analyze the inadequacy of cognitive science in explaining social decision-making, understood as the decision that affects both the life of him who decides and the relationships it has with other people. Therefore, this dissertation started from the assumption that two cognitive science elements could represent an obstacle to explain social decisions: one, the belief that the computer is the best model for understanding the human mind, other, the methodological option not to study the emotions. Aiming to analyze computational metaphor viability to explain social decision, we addressed to criticism of John Searle to computational model as proposed by cognitive science. To test the problems that could arise from methodological option to drop the emotions of studies, we examined emotions importance in social decision-making, especially the somatic marker hypothesis, created by António Damásio. Then presented to evolutionary psychology, specifically Robert Trivers' theory of reciprocal altruism, to analyze the relationship between emotion and social decision-making. This paper concludes by presenting a methodological critique of cognitive science, understanding that the methodological feature to exclude the emotions of studies is the main problem of cognitive science to explain social decision.

Keywords: Cognitive science. Social decision-making. Intentionality. Emotion. Neuroscience. Reciprocal Altruism.

SUMÁRIO

1 INTRODUÇÃO

O objetivo principal desta dissertação é analisar as insuficiências do modelo proposto pela ciência cognitiva para explicar a decisão social, entendida como a decisão que envolve o ambiente social, afetando tanto a vida daquele que decide quanto as relações que estabelece com outros indivíduos. A decisão social é dotada de complexidade, e envolve questões como retribuir um favor, prestar auxílio, mudar de emprego ou em quem votar.

Para tanto, este estudo analisará se, dentre os elementos que compõem a ciência cognitiva, existe algum que não se coadune com o processo de decisão social. Caso a ciência cognitiva possua uma crença central ou uma característica metodológica incompatível com a decisão social, a capacidade de explicá-la seria posta em dúvida.

Dentre os elementos que compõem a ciência cognitiva estão o uso do computador como metáfora para o funcionamento dos processos mentais e a exclusão das emoções dos estudos. A metáfora computacional é uma crença central da ciência cognitiva, enquanto a exclusão das emoções é uma característica metodológica (GARDNER, 2003, p. 20). Esta dissertação analisará teorias que indicam a incompatibilidade da crença central e da característica metodológica com a explicação do processo de decisão social.

Para o desenvolvimento do estudo, a primeira seção apresentará o conceito, percurso histórico e fundamentos metodológicos da ciência cognitiva. Nela serão analisadas as crenças centrais da ciência cognitiva, que incluem o uso de representações e do computador, entendido como o modelo mais viável de como a mente funciona. Também serão estudadas as características metodológicas da ciência cognitiva, que incluem a importância de estudos interdisciplinares, a busca de respostas para questões filosóficas clássicas e a exclusão de elementos complicadores para o funcionamento cognitivo, como emoção, contexto ou cultura.

Na segunda seção será abordada a crítica tecida por John Searle à metáfora computacional, crença central da ciência cognitiva, com o intuito de verificar se os argumentos dessa crítica refutam a capacidade da ciência cognitiva em explicar a decisão social. Para tanto, inicialmente serão apresentados os conceitos de intencionalidade e atos de fala de Searle, para, em seguida, apresentar a crítica que ele desenvolveu a respeito da analogia entre a cognição humana e o computador

digital. Searle buscou fornecer uma explicação exclusivamente biológica para os estados mentais, e formulou, na obra *Rationality in Action*, a própria teoria da racionalidade – que também será apresentada na segunda seção – sem se utilizar do modelo computacional cognitivista. A seção apresentará ainda as críticas aos conceitos de intencionalidade e atos de fala, que colocam em dúvida a validade dos argumentos apresentados por Searle contra a ciência cognitiva.

A terceira seção abordará outra linha de argumentação que pode levar ao questionamento da capacidade da ciência cognitiva de explicar a decisão social. Uma vez que a ciência cognitiva escolheu como pressuposto metodológico a exclusão das emoções de seus estudos, será analisado se essa exclusão pode ter prejudicado os resultados das por ela conclusões apresentadas. Para sopesar a relevância das emoções nas decisões sociais, serão apresentados estudos neurocientíficos, dentre os quais se destaca a hipótese do marcador-somático, desenvolvida pelo neurocientista António Damásio no livro *O erro de Descartes*. Com o objetivo de introduzir adequadamente a neurociência, a seção se inicia analisando como o estudo do cérebro influenciou a filosofia a partir da década de 1990.

A quarta seção segue a análise da relação entre emoções e processo de decisão social. Nessa seção, será apresentada teoria da seleção natural, elaborada por Charles Darwin na obra *A origem das espécies.* Será mostrado o percurso que levou à introdução da teoria biológica da seleção natural nas ciências humanas. Em seguida, será analisada a teoria do altruísmo recíproco, desenvolvida por Robert Trivers no artigo *The evolution of reciprocal altruism*. Ao argumentar a respeito do altruísmo recíproco, Trivers desenvolve argumentos a respeito do processamento das decisões sociais, e da relação delas com as emoções.

Caso o posicionamento de John Searle se mostre plausível, ou seja, se não for possível atribuir intencionalidade ao computador da mesma maneira que se atribui à mente humana, a crença da ciência cognitiva de que o computador é o modelo mais viável para o estudo da cognição se mostrará um obstáculo para explicar o processo de decisão social. Mas se a presença das emoções se revelar uma parte integrante da decisão social, a característica metodológica da ciência cognitiva que exclui as emoções de seus estudos, será a principal limitação da ciência cognitiva para elucidar a decisão social.

Dessa maneira, a presente dissertação abordará a possibilidade de que tanto a crença central de que o computador é o melhor modelo para explicar a decisão

social quanto a opção metodológica de excluir as emoções possam ter sido limitações quanto à possibilidade da ciência cognitiva de explicar a decisão social.

2 A CIÊNCIA COGNITIVA E O PROBLEMA DA DECISÃO SOCIAL

O presente estudo tem como objetivo central analisar se o modelo proposto pela ciência cognitiva é suficiente para explicar a decisão social.

Para exata compreensão do tipo de explicação que uma teoria fornece, é possível identificar duas modalidades de explicação:

a) explicação semântica: para esclarecer o significado de palavras ou outros símbolos. Fornece um conjunto de palavras com significado equivalente ou semelhante ao que está sendo explicado. A inteligibilidade da explicação semântica varia de pessoa para pessoa, pois a explicação que é suficiente para um pode não ser para outro, e o significado só foi esclarecido caso aquele para quem a é dirigida explicação a compreenda (CANDIOTTO & BASTOS, 2011, p. 20);
b) explicação científica: “é um enunciado que se propõe a explicar certo acontecimento ainda que ninguém aceite a explicação” (CANDIOTTO & BASTOS, 2011, p. 20). O objetivo de uma explicação cientifica não é esclarecer o significado, mas sim apresentar um significado verdadeiro. Na explicação semântica, o postulado é claro desde que a pessoa o compreenda, na explicação científica, será verdadeiro caso a pessoa nele acredite. Uma teoria fornece uma explicação científica, e o “êxito de uma teoria é medido pela capacidade que ela tem de fazer previsões e o *quantum* de explicações que ela pode fornecer” (CANDIOTTO & BASTOS, 2011, p. 21).

A análise realizada no presente estudo, no sentido de apreciar a capacidade da ciência cognitiva de explicar a decisão social, deve ser vista sob o prisma da explicação científica.

Antes de analisar a ciência cognitiva, neste momento será feita uma exposição do que este estudo entende como decisão social.

2.1 A DECISÃO SOCIAL

Este estudo adotou a definição de decisão social fornecida por António Damásio (1996, p. 199). A decisão social é aquela que envolve o ambiente social, ou seja, que afeta tanto a nossa vida quanto a dos demais.

Damásio apresenta as modalidades de decisões que entende inerentes aos seres humanos. A decisão, para ele, está ligada à racionalidade, uma vez que a finalidade do raciocínio seria a tomada de decisões: não há que se falar em uso do raciocínio em uma situação na qual não há nada para decidir. Portanto, os seres humanos raciocinam para decidir a respeito de algo (DAMÁSIO, 1996, p. 197).

A necessidade de decisão implica a existência de duas ou mais ações possíveis. Supondo que mais de uma ação seja possível, cada uma delas deve apresentar uma consequência futura distinta. Logo, o uso do raciocínio no processo de tomada de decisões pressupõe a existência de mais de um curso de ação possível, e que cada opção apresente consequências distintas.

Nem todas as decisões que tomamos são dotadas de complexidade. Há decisões relacionadas aos apetites, tais como fome ou sede (DAMÁSIO, 1996, p. 198). Sentimos um impulso para satisfazer tais necessidades. Essas decisões pressupõem a existência de um mecanismo corporal que nos faça sentir o desejo dos elementos que necessitamos.

Também não possui complexidade a decisão relacionada a atos reflexos (DAMÁSIO, 1996, p. 199), como na situação de uma brusca freada no trânsito. Embora nem sempre seja a ação apropriada, o ato reflexo é instantâneo e não nos permite pensar sobre o assunto. É uma resposta imediata a uma circunstância alheia a nossa vontade.

Há, finalmente, outro grupo de decisões, essas dotadas de complexidade, onde o uso do raciocínio é fundamental. Esse terceiro grupo se subdivide em dois (DAMÁSIO, 1996, p. 199).

O primeiro deles inclui decisões como em quem votar; qual carreira escolher; se somos a favor ou contra a pena de morte; onde investiremos nosso dinheiro. Em casos tais, quanto mais opções existirem, maior a dificuldade da escolha.

O segundo grupo de decisões complexas inclui o raciocínio referente à resolução de um problema matemático, à construção de um prédio ou de uma aeronave. Verifica-se, portanto, que os dois subgrupos abarcam decisões intrinsecamente diferentes.

A diferença fundamental entre os dois subgrupos apresentados pode ser resumida no envolvimento das decisões com o ambiente social. Em outras palavras, quando decidimos em quem votar, onde investir ou qual carreira escolher, tomamos uma decisão eminentemente social. Essa decisão afeta tanto a nossa vida quanto a

dos demais. Por outro lado, quando resolvemos um complicado problema aritmético, não estamos diretamente envolvidos com outras pessoas.

A decisão social é aquela que envolve o ambiente social, ou seja, que afeta tanto a nossa vida quanto a dos demais. Embora esse tema seja objeto de análise mais detalhada na seção 4.2, a seguir será feita uma explicação do que caracteriza uma decisão social e a distingue das demais.

Decidir significa escolher agir em determinada direção quando há mais de uma ação possível. Mas nem todas as decisões têm a mesma complexidade. Algumas são simples, como as escolhas relacionadas aos apetites (DAMÁSIO, 1996, p. 198). Os apetites, que englobam a fome e a sede, pressupõem um mecanismo corporal que estimula um impulso para atender as necessidades. Atos reflexos, como a decisão de fechar os olhos ao perceber um objeto se aproximando, são instantâneos e também não são dotados de complexidade (DAMÁSIO, 1996, p. 199). Nas decisões relacionadas a atos reflexos e apetites o uso do raciocínio é de menor importância.

A faculdade de raciocinar está presente em dois grupos de decisões, mais complexas que as já apresentadas (DAMÁSIO, 1996, p. 199). O primeiro inclui situações como a resolução de um problema matemático, a elaboração de uma planta para a construção de um edifício.

O segundo envolve decisões como a escolha de uma carreira, de um candidato ou de um parceiro amoroso. Constata-se que a diferença fundamental entre os dois subgrupos apresentados pode ser resumida no envolvimento das decisões com o ambiente social. São essas as decisões objeto de análise nesta dissertação.

Feitas essas considerações preliminares, a seguir a ciência cognitiva será objeto de análise.

2.2 A CIÊNCIA COGNITIVA COMO SUPERAÇÃO DAS LIMITAÇÕES METODOLÓGICAS DO BEHAVIORISMO

Nesta seção serão apresentados o conceito, percurso histórico e fundamentos metodológicos da ciência cognitiva.

A ciência cognitiva, nascida em meados do século XX como contraposição ao behaviorismo de Skinner, representou a união de pesquisas provenientes de diversas áreas do saber, como a psicologia, a linguística, a antropologia, a sociologia e a medicina, com o intuito de tentar compreender os mecanismos inerentes à mente

humana. Antes da ciência cognitiva, as tentativas de compreender e explicar o funcionamento da mente humana estavam restritas a teólogos e filósofos.

O psicólogo cognitivo e educacional Howard Gardner, da Universidade de Harvard, define a 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 seu emprego (GARDNER, 2003, p. 19).

Gardner relaciona o termo ciência cognitiva sobretudo a esforços para a compreensão do conhecimento humano, embora, às vezes, o termo seja ampliado para incluir todas as formas de conhecimento: animado ou inanimado, humano ou não humano (GARDNER, 2003, p. 20). Seguindo a concepção de Gardner, esta dissertação trabalhará com as implicações da ciência cognitiva no conhecimento humano.

A ciência cognitiva tem como objetivo, segundo Candiotto, "constituir-se como ciência natural da mente, a saber, construir uma teoria dos fenômenos mentais que alcance explicações aceitáveis para uma abordagem naturalista das propriedades da mente" (CANDIOTTO, 2011, p. 75). Verifica-se o destaque dado à abordagem naturalista, o que significa, ainda segundo Candiotto, que os cientistas cognitivos, em geral, buscam "integrar suas teorias sobre os fenômenos mentais à luz das explicações do mundo natural, com o intuito de fundamentar as ciências naturais da cognição" (2011, p. 75).

Cumpre ressaltar que a ciência cognitiva objeto de análise nesta dissertação é a ciência cognitiva clássica, apresentada por Gardner na obra *A nova ciência da mente*.

2.2.1 A perspectiva histórica da ciência cognitiva

O surgimento da ciência cognitiva está relacionado ao cenário científico da época. No final do século XIX e início do século XX, influenciados pelos escritos do médico e filósofo Wilhelm Maximilian Wundt, investigadores, dentre eles Edward Bradford Titchener, começaram a defender que o estudo de questões sobre a consciência e o pensamento deveriam ser feitos com base em métodos experimentais

rigorosos. Contudo, o método que Wundt e seus discípulos propuseram foi a introspecção, ou seja, uma análise detalhada dos próprios padrões de pensamento.

A psicologia behaviorista surgiu como uma refutação aos métodos de introspecção, e teve como precursor John B. Watson. Ele negou a importância da introspecção no estudo da psicologia, e desprezou qualquer tentativa de estudo dos chamados estados mentais, que, por serem introspectivos, não poderiam ser analisados cientificamente. Para Watson, somente o comportamento observável era relevante. Por meio da realização de experimentos condicionantes, ele foi capaz de fazer animais e até seres humanos se comportarem como queria, e isso convenceu grande parte da comunidade científica americana a adotar o behaviorismo. A relação de estímulo e resposta parecia ser tudo o que era necessário para a psicologia. Evitavam-se termos do vocabulário mentalista, tais como comportamento intencional ou consciência. Assim, tudo que não fosse observável deveria ser posto de lado.

O behaviorismo foi motivado por três concepções que influenciavam os intelectuais da época. A primeira foi uma reação contra o dualismo de substâncias[1]. A segunda, a ideia Positivista de que o significado de qualquer sentença é, em última análise, uma questão de averiguar se as circunstâncias observáveis tenderiam a confirmar a sentença. A terceira foi a suposição de que a maioria, se não todos os problemas filosóficos eram o resultado de uma confusão conceitual ou linguística, e, dessa forma, para resolver os problemas filosóficos bastaria uma análise cuidadosa da linguagem na qual o problema é expresso (CHURCHLAND, 1988, p. 23).

O behaviorismo representou uma legítima expressão do desconforto que a comunidade científica sentia com o método da introspecção. Esse desconforto, somado aos resultados positivos obtidos por Watson nos experimentos

---

[1] A explicação dualista da mente engloba mais de uma teoria, mas todas possuem como núcleo essencial a concepção de que a natureza da consciência reside em algo não físico.
Uma importante teoria dualista é o dualismo de substâncias, teorizado sobretudo pelo filósofo René Descartes. Para essa teoria, a consciência inteligente é feita de uma substância que não se confunde com o corpo material (CHURCHLAND, 1988, p. 8). Maiores detalhes a respeito do dualismo cartesiano serão vistos na seção 4.2.1, nota de rodapé n.º 11.
Outra teoria é o dualismo de propriedades, que preconiza que, embora mente e corpo tenham a mesma substância, possuem propriedades diferentes. Tais propriedades incluem, por exemplo, a dor, a sensação do vermelho e o desejo, e não podem ser reduzidas ou explicadas apenas em termos materiais. (CHURCHLAND, 1988, p. 11).
O dualismo de propriedades foi proposto no final da década de 1970, com trabalhos de Karl Popper (1977) e Joseph Margolis (1978), depois do surgimento behaviorismo. Dessa maneira, a psicologia behaviorista era diretamente contrária ao dualismo de substâncias.

condicionantes que realizou, fez com que o behaviorismo se tornasse a doutrina dominante nas décadas de 1920, 1930 e 1940.

Se por um lado o behaviorismo foi uma resposta positiva aos critérios pouco avaliáveis do método da introspecção, por outro representou um obstáculo para a discussão de temas como linguagem e imaginação. A comunidade científica não estava aberta para estudos de aspectos da cognição que não fossem observáveis por meio do comportamento.

Em setembro de 1948, no Simpósio Hixon, o psicólogo Karl Lashley criticou a abordagem behaviorista durante o discurso denominado “O problema da ordem serial no comportamento”. Para Lashley, a cadeia de estímulo e resposta típica do behaviorismo não era capaz de explicar comportamentos serialmente ordenados, como falar, jogar tênis ou tocar piano. Isso porque, como comportamentos serialmente ordenados ocorrem muito rápido, não há tempo hábil para buscar o passo seguinte no comportamento anterior (GARDNER, 2003, p. 26).

O discurso de Lashley foi bem aceito pela comunidade científica, e representou o início do declínio da psicologia behaviorista. A aceitação dos postulados por ele apresentados evidenciou que muitos pensadores já percebiam as dificuldades causadas pela relutância em analisar estados mentais introspectivos.

Enquanto o behaviorismo entrava em declínio, o advento dos computadores e as analogias entre eles e o processo cognitivo humano foi o sinal da mentalidade que se seguiria. Filósofos, psicólogos e pensadores de diversas áreas começaram a perceber semelhanças entre o funcionamento dessas máquinas e da mente humana. A partir da década de 50, começaram a ser publicados trabalhos que realizavam analogias entre cognição humana e o funcionamento do computador, tais como o livro póstumo de John von Neumann, *The computer and the brain* (1958), e o artigo *Minds and machines* (1960) do filósofo Hilary Putnam. Graças à receptividade da comunidade cientifica às analogias entre o computador e os processos cognitivos, em meados do século XX a ciência cognitiva passou a ser a corrente de pensamento dominante, em substituição ao behaviorismo.

Também merece destaque no surgimento da ciência cognitiva a evolução matemática e lógica do início do século XX, precursora do desenvolvimento computacional. A lógica do raciocínio silogístico, desenvolvida na época de Aristóteles, havia dominado o cenário matemático por quase dois mil anos. No entanto, no final do século XIX, o lógico alemão Gottlob Frege desenvolveu uma nova

forma de lógica que envolvia a manipulação de símbolos abstratos. No início do século XX, os lógicos matemáticos Bertrand Russell e Alfred North Whitehead tiveram considerável sucesso em reduzir as leis básicas da aritmética a proposições de lógica elementar, e acabaram influenciando uma geração de pensadores com orientação matemática, incluindo John von Neumann e Norbert Wiener (GARDNER, 2003, p. 31).

O estudo lógico matemático de Alan Turing, realizado na década de 1930, teve grande importância para a ciência cognitiva. Ele desenvolveu uma máquina simples, posteriormente denominada “máquina de Turing”, que funcionava da seguinte maneira:

> Só eram necessários uma fita e um *scanner* (varredor) para ler o que estava na fita. A fita em si era dividida em quadrados idênticos, cada um dos quais contendo em sua superfície algum tipo de símbolo. Para fins de ilustração, Turing considerou uma máquina que usava o código binário [...], porém a única restrição geral era de que o número de símbolos diferentes não podia ser infinito. A cada passo, dependendo de seu estado interno, a máquina mantém o símbolo que é lido pelo *scanner,* ou o substitui por outro, e em seguida passa a ler o quadrado à direita, ou à esquerda, ou o mesmo quadrado. Apenas com estas operações simples, a máquina era capaz de executar qualquer tipo de programa ou plano que pudesse ser expresso por meio de número finito de símbolos. (GARDNER, 2003, p. 32).

Portanto, a máquina de Turing poderia ser programada para realizar qualquer tarefa que pudesse ser expressa em passos claros.

Entusiasmado com o desenvolvimento das máquinas computadoras, em 1950 Turing teorizou a respeito da possibilidade de programar uma máquina de modo que seria impossível distinguir se as respostas eram da máquina ou de um ser humano, o que ficou conhecido como “teste de Turing”. O objetivo do teste é demonstrar que as máquinas podem pensar, pois, caso o observador não seja capaz de identificar se a resposta foi fornecida pela máquina ou por uma pessoa, diz-se que a máquina passou no teste de Turing.

2.2.2 Elementos da ciência cognitiva

Visto o panorama histórico que levou ao surgimento da ciência cognitiva, é possível passar para o estudo dos elementos básicos que ela apresenta. Gardner elenca cinco aspectos que considera os mais importantes para a identificação de um estudo científico-cognitivo. Caso todos os itens elencados estejam presentes, ou ao

menos a maioria deles, estaremos diante de um trabalho da ciência cognitiva. Gardner subdivide os elementos identificadores da ciência cognitiva da seguinte maneira:

a) crenças centrais da ciência cognitiva:

- uso de representações: atividades cognitivas humanas devem ser compreendidas por meio das representações mentais, que representam um nível de análise separado do biológico ou neurológico;

- uso do computador: o computador é o modelo mais viável de como a mente funciona.

b) características metodológicas da ciência cognitiva:

- importância de estudos interdisciplinares;

- busca respostas para questões que remontam ao início da tradição filosófica ocidental;

- exclusão de elementos complicadores para o funcionamento cognitivo, como emoção, contexto ou cultura.

A seguir, cada um dos elementos que compõem a ciência cognitiva será objeto de análise, começando pelo *uso de representações*.

Dizer que o cientista cognitivo se utiliza de representações significa afirmar que ele postula num nível de realidade separado, denominado *nível da representação*.

A ciência cognitiva trabalha com símbolos, imagens, regras. Os cientistas cognitivos partem do pressuposto de que os seres humanos agem com base em modelos cognitivos interiorizados do nosso ambiente físico e social, e estas estruturas de conhecimento internas são as representações.

O cognitivista Jerry Fodor, que elaborou a própria teoria da mente a partir de uma perspectiva funcionalista – ou seja, a constituição psicológica de um sistema não depende do *hardware*, mas sim do *software*[2] –, desenvolveu amplamente o conceito

---

[2] A teoria funcionalista exerceu grande influência na ciência cognitiva clássica.
Hilary Putnam foi o idealizador da teoria funcionalista, que preconizava que a constituição psicológica de um sistema não depende do *hardware,* mas sim do *software* (GARDNER, 2003, p. 95). O *hardware* é a parte física do sistema, enquanto o *software* é o programa implementado.
Jerry Fodor acrescente às ideias básicas de Putnam a abordagem do processamento de informações das ciências cognitivas. Para ele, as atividades mentais são realizadas e constituídas na manipulação de símbolos, que são as representações mentais.
O funcionalismo não é uma teoria dualista, uma vez que entende que pensamentos, sentimentos, crenças e afins, não devem ser analisados como um tipo de substância material ou imaterial, mas sim a partir do papel causal que exercem na vida mental de um organismo (FODOR, 2004, p. 174).
Por outro lado, embora não se trate de uma teoria necessariamente materialista, pois não afirma que as mentes são feitas de coisas materiais, ao alegar que as mentes devem se materializar em algo, e esse algo muito provavelmente será físico, é uma teoria em maior consonância ao materialismo (MATTHEWS, 2007, p. 55).

de representação. Para ele, as atividades cognitivas são constituídas por meio das representações mentais, as quais nada mais são do que a manipulação de símbolos. Além disso, não se deve procurar nas representações mentais alguma semelhança com a realidade; pelo contrário, os símbolos mentais devem ser entendidos como entidades abstratas, sem qualquer relação física com as entidades que representam (FODOR, 1975, p. 31).

Como assevera Gardner, “Fodor foi muito mais longe que seus contemporâneos em sua disposição para pensar sobre como pode ser a representação mental” (2003, p. 95).

Fodor propõe a existência de uma linguagem do pensamento, como condição necessária para a formulação de qualquer teoria em psicologia cognitiva (1975, p. 33). De acordo com ele, uma vez que as operações mentais são realizadas por meio da manipulação das representações, que não passam de símbolos, deve haver um sistema interno capaz de manipulá-las de maneira eficaz. No entanto, mais do que um meio formal de manipulação de símbolos, a linguagem do pensamento envolve a análise da representação dos conteúdos do mundo, uma vez que nossos pensamentos são sobre coisas que existem no mundo.

Em suma, podemos concluir que os processos mentais, para a ciência cognitiva:

> consistem em manipulação de símbolos, que são as representações mentais. O pensamento é a manipulação lógica de representações mentais que tem uma forma correlata com a linguagem proposicional comum: a linguagem do pensamento ou mentalês (Candiotto & Bastos, 2011, p. 155).

Assim como as representações, cientistas cognitivos entendem que o *uso do computador digital* possui um papel fundamental no estudo da cognição humana. Além de indispensáveis para a manipulação de dados e desenvolvimento de estudos, ele possui um uso mais profícuo, que seria fornecer um modelo de comparação para o funcionamento da mente humana. As analogias entre a inteligência artificial e a inteligência humana influenciaram praticamente todos os cientistas cognitivos.

Diferente do behaviorismo, que buscava explicar um estado mental apenas a partir do estímulo (*input*) e resposta (*output*), o funcionalismo acrescenta um terceiro elemento necessariamente existente: a referência a uma variedade de outros estados mentais com os quais estará causalmente conectado (CHURCHLAND, 1988, p. 36).

Segundo o funcionalismo, não importa a base física que propicie a consciência, mas sim que essa base desempenhe funções equivalentes aos estados mentais. Isso possibilitaria a existência de estados mentais em um ente sem propriedades biológicas.

A ciência cognitiva parte de uma perspectiva central de que a mente atua por meio do processamento de informação, ou seja, da computação. Como os processadores de informação têm a capacidade de, ao mesmo tempo, representar e transformar a informação, sob o prisma de uma perspectiva cognitivista a mente teria alguma forma de representação e processamento para agir sobre e manipular as informações (FRIEDENBERG & SILVERMAN, 2006, p. 3).

Seguindo com a análise dos elementos fundamentais da ciência cognitiva, passamos ao estudo das características metodológicas. A primeira é a *convicção em estudos interdisciplinares*. Com o trabalho conjunto de pensadores provenientes de áreas distintas, os cientistas cognitivos acreditam ser possível chegar a melhores conclusões do que com estudos de disciplinas isoladas.

Outra característica é o *estudo de problemas filosóficos clássicos*. Isso significa que os debates filosóficos tradicionais, que remontam à filosofia grega, constituem o centro da agenda da ciência cognitiva.

O último elemento e característica metodológica trata da *exclusão da emoção, do contexto, da cultura e da história dos estudos da ciência cognitiva*. Os cientistas cognitivos argumentam que a inclusão desses fatores poderia obscurecer ou tumultuar os estudos propostos. Na tentativa de explicar tudo, é possível não explicar nada. Esse aspecto, nas palavras de Gardner, representa:

> a decisão deliberada de não enfatizar certos fatores que podem ser importantes para o funcionamento cognitivo mas cuja inclusão nesse momento complicaria desnecessariamente o empreendimento cognitivo-científico. Estes fatores incluem a influência de fatores afetivos ou emoções, a contribuição de fatores históricos ou culturais, e o papel do contexto de fundo no qual ocorrem atitudes ou pensamentos particulares (GARDNER, 2003, p. 20)

A exclusão da emoção pode ser constatada ao analisarmos as considerações tecidas por Jerry Fodor na obra *The Language of Thought*. Nessa obra, Fodor afirma que as emoções são estados mentais que devem ser considerados externos ao domínio da explicação cognitiva. Segundo o autor:

> eu acho provável que haja muitos tipos de exemplos de relações causais-mas-não-computacionais entre estados mentais. Muitos processos associativos talvez sejam assim, como também, provavelmente, muitos dos efeitos da emoção sobre a percepção e crença. Se esse palpite estiver certo, então esses são exemplos de boa-fé de relações causais entre estados mentais que, no entanto, estão fora do domínio da explicação psicológica

> (cognitiva). O que a psicologia cognitiva pode fazer, é claro, é especificar os estados que são relacionados e dizer que eles são relacionados. Mas, do ponto de vista psicológico, a existência de tais relações é simplesmente uma questão de fato bruto; a explicação deles é deixada a um nível mais baixo de investigação (provavelmente biológica). (FODOR, 1975, p. 203).

Em resumo, seguindo o pensamento da ciência cognitiva, Jerry Fodor considera que fatores emocionais estão fora de uma explicação psicológica cognitiva, e devem ser investigados por outras ciências, possivelmente a biologia. Não caberia, assim, tratar de emoções no estudo do processo cognitivo humano.

Vale ressaltar que os cientistas cognitivos tentaram explicar o processo de cognição como um todo. A cognição é a terminologia empregada na ciência cognitiva para abranger todos os processos mentais, incluindo o processo de decisão social, que é objeto desta dissertação. Como o processo de decisão social não passa de uma modalidade de cognição, também foi estudado a partir das crenças centrais e características metodológicas gerais da ciência cognitiva.

Verifica-se, em suma, que a ciência cognitiva clássica procurou compreender todos os processos cognitivos, incluindo a decisão social, utilizando-se do modelo computacional. Ao trabalhar com o conceito de representações, estabelecendo como possível a existência de estados mentais não neurofisiológicos, os cientistas cognitivos procuraram afastar tanto explicações socioculturais quanto biológicas ou neurológicas para a cognição (GARDNER, 2003, p. 20). O afastamento do estudo das bases biológicas do processo de cognição levou à opção metodológica de se ignorar o fator emocional no processo de decisão, inclusive na decisão social.

Na próxima seção será objeto de análise a crítica feita por John Searle ao modelo computacional da ciência cognitiva, bem como a teoria que Searle desenvolveu buscando explicar o processo de decisão social. Veremos que Searle refuta as analogias entre mente e computador digital e pretende fornecer uma explicação biológica para a decisão, embora não inclua fatores como a evolução e a emoção em sua teoria da racionalidade.

3 JOHN SEARLE: DA CRÍTICA À CIÊNCIA COGNITIVA AO DESENVOLVIMENTO DE UMA EXPLICAÇÃO PARA A DECISÃO SOCIAL

Nesta seção será feita uma análise da crítica de John Searle ao uso da inteligência artificial como modelo para a cognição humana. Searle refutou a presença de intencionalidade nas máquinas, o que inviabilizaria o sucesso das analogias entre cognição humana e computador, e defendeu a aplicação do naturalismo biológico – ou seja, uma explicação totalmente biológica – para a compreensão da cognição. Ele procurou embasar sua teoria nos mecanismos biológicos do cérebro e na intencionalidade, sem recorrer ao nível da representação simbólica. Não se convencendo dos argumentos utilizados pela ciência cognitiva, Searle desenvolveu uma teoria da racionalidade própria para tentar explicar o processo de decisão humano, inclusive a decisão social, teoria essa que também será objeto de estudo nesta seção.

John Searle, filósofo e escritor norte-americano, professor da Universidade de Berkeley, na Califórnia, Estados Unidos, argumentou contra um dos principais enunciados da ciência cognitiva: a possibilidade de criarmos máquinas inteligentes.

Para bem compreender a crítica formulada por Searle, antes serão expostos os conceitos de atos de fala e de intencionalidade por ele propostos. Searle se utilizou da teoria que desenvolveu sobre os atos de fala como base para elaborar a própria teoria da intencionalidade, e empregou a intencionalidade para formular a crítica ao modelo computacional de cognição proposto pela ciência cognitiva. Para Searle, o computador digital jamais teria a intencionalidade que um ser vivo possui.

Também será apresentado o conceito de naturalismo biológico formulado por Searle. Ele argumentou que buscava explicar o processo cognitivo a partir de uma perspectiva exclusivamente biológica. Para Searle, os estados mentais seriam produzidos da mesma maneira que qualquer outro processo biológico.

Ao final desta seção, serão expostas críticas ao conceito de intencionalidade de John Searle, dentre elas a formulada por Daniel Dennett, que afirma que o naturalismo biológico proposto por John Searle é inconsistente, uma vez que Searle não obtém êxito ao tentar fornecer uma explicação biológica aos estados mentais.

Vimos que o objetivo central desta dissertação é analisar a viabilidade do modelo proposto pelos cognitivistas para explicar as decisões sociais. Uma vez que as críticas de John Searle atingem pontos essenciais da ciência cognitiva, mostra-se

relevante analisá-lo neste estudo. Searle se dedica a elaborar argumentos para demonstrar as falhas do modelo computacional, e a explicar os estados mentais por meio da intencionalidade e dos mecanismos biológicos, sem recorrer às representações, como fazem os cientistas cognitivos.

Porém, Searle não apenas criticou a ciência cognitiva, mas também elaborou uma teoria própria para explicar a decisão humana, ou, como ele denominou, a racionalidade prática. A teoria da racionalidade desenvolvida por ele será objeto de análise, embora não seja adotada nesta dissertação. Isso porque, como será exposto, Searle ignorou o papel das emoções e da seleção natural na teoria que formulou, elementos que este estudo entende como essenciais para a compreensão do processo de decisão social.

Feitas essas considerações iniciais, na seção a seguir passaremos à crítica de John Searle em face da ciência cognitiva.

3.1 A CRÍTICA DE JOHN SEARLE AO USO DO MODELO COMPUTACIONAL

A crítica que Searle tece em face da ciência cognitiva é fundada no conceito de intencionalidade que desenvolve. Por esse motivo, antes de ingressar no estudo da crítica propriamente dita, será exposta a teoria da intencionalidade que ele formulou.

John Searle conceitua a intencionalidade como “aquela propriedade de muitos estados e eventos mentais pela qual estes são dirigidos para, ou acerca de, objetos e estados de coisas no mundo” (SEARLE, 2002, p. 1). A intencionalidade é a direção dos estados e eventos mentais. Por exemplo, se alguém tem uma crença, essa crença é a respeito de um fato; se tem um medo, é o medo de que algo ocorra; se tem uma intenção, é uma intenção de fazer alguma coisa.

O termo intencionalidade foi utilizado pela primeira vez no século XIX, pelo padre-filósofo Franz Brentano. Brentano desenvolveu o conceito de intencionalidade com o intuito de refutar a hipótese de que é possível analisar os elementos da consciência dividindo-os em compartimentos estanques. Conforme Brentano, não se pode conceber a mente sem um objeto, sem intenções a ela direcionadas:

> [...] a peculiaridade que, acima de tudo, é geralmente característica da consciência, é que ela mostra sempre e em toda parte, ou seja, em cada uma de suas partes separáveis, um certo tipo de relação, relacionando um sujeito a um objeto. Essa relação também é referida como "relação intencional". Para cada consciência pertence essencialmente uma relação (BRENTANO, 1995, p. 21).

Portanto, Brentano preconiza que não é possível um ato mental sem um objeto para o qual seja direcionado. O objeto para o qual o ato mental é direcionado deve ser entendido como um objeto interno, que pode não existir na realidade exterior. Para evitar confundir o objeto do ato mental com um objeto existente, Brentano denomina como “imanente” o objeto do ato mental (BRENTANO, 1995, p. 22).

Searle, contudo, afirma que desenvolve o conceito de intencionalidade de maneira peculiar, própria. (SEARLE, 2002, p. 1).

Em primeiro lugar, Searle esclarece que não são todos os estados mentais que possuem intencionalidade, apenas alguns. Crenças, medos, esperanças e desejos são intencionais. Por outro lado, uma ansiedade difusa ou depressão generalizada não podem ser consideradas intencionais, pois uma pessoa pode estar ansiosa ou deprimida sem que este estado esteja relacionado a alguma coisa. Se o estado mental não é sobre algo, se não tem um objeto específico, então não pode ser considerado intencional (SEARLE, 2002, p. 2).

Outra característica da intencionalidade é que ela não se confunde com consciência (SEARLE, 2002, p. 2). Por um lado, é possível que existam estados conscientes e não intencionais, como é o caso da já citada hipótese de uma ansiedade difusa. A pessoa está consciente de que está ansiosa, mas não é capaz de relacionar o estado de ansiedade a nenhum objeto específico.

Por outro lado, é possível que haja estados intencionais que não são conscientes, como no caso de determinadas crenças, que podem nunca ter sido analisadas conscientemente por um indivíduo, que nem por isso deixa de tê-las. Searle fornece o seguinte exemplo de um estado intencional não consciente: “acredito, por exemplo, que meu avô paterno tenha passado a vida inteira no território continental dos Estados Unidos, mas até este momento nunca havia formulado ou considerado conscientemente esta crença” (SEARLE, 2002, p. 3). Searle ressalta que as crenças não são inconscientes, ou seja, não estão ocultas da consciência. Uma crença não consciente nada mais é do que uma crença sobre a qual não pensamos normalmente. Assim, para Searle, nem toda consciência é consciência *de*, de modo que não há identidade entre consciência e intencionalidade.

O estado intencional não se identifica com o objeto para o qual está sendo direcionado, como, por exemplo, o medo de cobras não é o mesmo que cobras reais. No caso da ansiedade difusa, faltaria o objeto para o qual ela deveria ser direcionada,

de modo que não pode ser considerada um estado intencional, embora seja um estado consciente (SEARLE, 2002, p. 3).

Intencionalidade também não se confunde com intenção, apesar da similaridade dos termos. As intenções são apenas uma modalidade de intencionalidade, que deve sempre ser entendida no sentido de direção[3].

Após apresentar as principais características da intencionalidade, Searle tenta responder à pergunta: qual a relação entre os estados intencionais e os objetos e estados de coisas aos quais estão direcionados? (SEARLE, 2002, p. 6). Para tanto, o autor introduz o conceito de intencionalidade como representação, seguindo a teoria dos atos de fala.

Como Searle fundamenta a relação entre estados intencionais e objetos direcionados nos atos de fala, é importante conceituar os atos de fala antes de prosseguir no estudo dos estados intencionais. Além disso, como esclarece Barry Smith, Searle utiliza a teoria dos atos de fala como fundamento básico para as teorias posteriormente desenvolvidas, relacionadas à consciência, ao mental, à realidade institucional e social, e, mais recentemente, à racionalidade e ao livre arbítrio (SMITH, 2003, p. 2). Por tais razões, neste ponto será feita uma breve incursão no conceito de atos de fala como proposto por Searle.

O conceito de atos de fala foi desenvolvido por John Langshaw Austin. Para explicá-los, Austin dividiu os atos de fala em três modalidades: atos locutórios, ilocutórios ou perlocutórios. Os atos locutórios são aqueles que se realizam ao utilizarmos a fala (AUSTIN, 1990, p. 88), ou seja, nada mais são do que a pronúncia dos sons peculiares a cada língua. No entanto, podemos empregar a fala de diversas formas, e saber a maneira específica que está sendo utilizada naquele momento é fundamental para a compreensão dos sons emitidos. Nas palavras do autor:

> faz uma grande diferença saber se estávamos advertindo ou simplesmente sugerindo, ou, na realidade, ordenando; se estávamos estritamente prometendo ou apenas anunciando uma vaga intenção, e assim por diante (AUSTIN, 1990, p. 85).

Os atos ilocutórios são a forma específica de expressão que ocorre quando o ato de fala se reveste de um sentido direto. Nos atos ilocutórios, não se trata apenas

---

[3] Para evitar confusões entre os termos *intencionalidade* e *intenção*, John Searle utiliza as palavras intencionalidade e seus derivados com a primeira letra maiúscula. Apesar disso, a referida palavra será grafada com a inicial em minúsculo na presente dissertação, considerando a difusão do termo intencionalidade na literatura filosófica, salvo em citações literais.

do ato *de* dizer algo – o que seria o caso dos atos locutórios –, mas sim do ato *ao* dizer algo (AUSTIN, 1990, p. 89). Eles se caracterizam por possuírem uma força considerável, que pode ser de ameaça, ordem, aconselhamento etc.

Os atos perlocutórios não expressam uma ordem ou advertência direta, mas de alguma forma influenciam os ouvintes. As palavras ditas de forma frequente, até mesmo natural, podem fazer com que aqueles que as ouçam se sintam de alguma forma atingidos por elas. Os atos perlocutórios têm como objetivo justamente exercer essa influência (AUSTIN, 1990, p. 90).

Searle desenvolve a teoria de Austin sobre os atos de fala. No artigo *What is a Speech Act?*, Searle apresenta as regras que entende necessárias para a existência dos atos de fala, e considera que são praticados por meio da pronúncia de sons que possuem significado por estarem submetidos a determinadas regras constitutivas.

Searle não segue a classificação de Austin de atos locutórios, ilocutórios ou perlocutórios. No artigo *A Taxonomy of Illucutionary Acts*, Searle apresenta uma classificação dos atos de fala com base na direção de adequação existente entre a realidade e a linguagem, que pode ser da palavra para o mundo ou do mundo para a palavra. A direção de adequação é a condição para que o ato de fala seja verdadeiro ou não.

Para John Searle, os atos de fala possuem conteúdo proposicional e força ilocutória (1965, p. 5). Uma frase pode ser dita em diversos sentidos, como uma afirmação, uma ordem, uma pergunta ou um aviso. Por exemplo, a frase "*S* sairá da sala" pode ser pronunciada como uma pergunta: *S* sairá da sala?; como uma ordem: *S,* saia da sala!; como uma afirmação: *S* sairá da sala. Em todos os casos existe um *conteúdo proposicional* comum: existe um sujeito *S* que sairá da sala. Mas a *força ilocutória* em cada uma das sentenças é diversa: no exemplo, ela representa uma pergunta, uma ordem e uma afirmação. Essa variação de sentidos é a força ilocutória do ato de fala.

Feitas essas considerações a respeito dos atos de fala, é possível retornar à análise da comparação que Searle faz entre atos de fala e estados intencionais. Segundo ele, "os estados Intencionais representam objetos e estados de coisas no mesmo sentido de 'representar' em que os atos de fala representam objetos e estados de coisas" (2002, p. 6).

Não é a linguagem que dá origem à intencionalidade, mas a linguagem que deriva da intencionalidade (SEARLE, 2002, p. 8). Embora a linguagem seja um derivado da intencionalidade, ambas teriam vários pontos de semelhança.

Assim como os atos de fala, a intencionalidade também seria dotada de conteúdo e força ilocucionária. Por exemplo, eu posso querer que o sujeito *S* saia da sala, temer que *S* saia, acreditar que *S* sairá. Caso esta proposição seja pronunciada, será um ato de fala com conteúdo proposicional – que *S* saia da sala – distinto da força ilocucionária. Caso esta proposição seja apenas uma crença, um desejo ou um temor, ou seja, apenas um estado intencional, o conteúdo (neste caso denominado *representativo* e não proposicional: que *S* saia da sala) permanece o mesmo, enquanto o modo psicológico (que seria correspondente à força ilocucionária dos atos de fala) varia (SEARLE, 2002, p. 8).

Searle aplica as direções de adequação para os estados intencionais. Caso a crença de alguém se mostre equivocada, o problema está na crença, e não no mundo. Portanto, pode-se concluir que a direção de adequação de uma crença é mente-mundo. Já em relação a desejos, caso eles não possam ser realizados, a situação se modifica apenas por meio da modificação do desejo. Não há desejos falsos ou verdadeiro, mas apenas desejos que possam ou não ser cumpridos, de modo que a direção de adequação dos desejos é mundo-mente (SEARLE, 2002, p. 11).

Há ainda mais uma relação entre o ato de fala e o estado intencional: a condição de sinceridade do ato de fala. Na realização de um ato ilocutório com um conteúdo proposicional (ato de fala), expressamos um determinado estado intencional relacionado a esse conteúdo, e esse estado intencional é a condição de sinceridade desse ato (SEARLE, 2002, p. 12).

Por fim, atos de fala e estados intencionais também se assemelham quanto à aplicação da noção de condições de satisfação. No tocante aos atos de fala, verifica-se que um enunciado é satisfeito *se, e somente se,* for verdadeiro; uma ordem é satisfeita *se, e somente se*, for obedecida. As condições de satisfação também se aplicam aos estados intencionais. Por exemplo, uma crença será satisfeita se, e somente se, as coisas forem como acredito. (SEARLE, 2002, p. 14)

Com tais argumentos, Searle tem como objetivo confirmar a imagem de estado intencional fornecida inicialmente: “todo estado Intencional compõe-se de um conteúdo representativo em um certo modo psicológico” (SEARLE, 2002, p. 15).

É a partir do conceito de intencionalidade que Searle contesta os fundamentos da ciência cognitiva. Uma vez que os fundamentos básicos da intencionalidade já foram apresentados, a seguir passaremos à crítica formulada por Searle em face do modelo proposto pelos cognitivistas.

John Searle desenvolveu sua filosofia nas décadas de 70 e 80, quando preponderava o modelo computacional da ciência cognitiva, analisado na seção 2.2 desta dissertação. Dessa forma, a concepção predominante em filosofia e psicologia era de que o cérebro humano funcionava de modo semelhante aos computadores digitais, e as analogias entre ambos proliferavam. A versão mais extrema dessas analogias se refere ao modelo funcionalista, visto na seção 2.2, nota de rodapé 2. O funcionalismo preconiza que o cérebro seria como um computador digital, o *hardware*, enquanto a mente seria o programa de computador, o *software*. Searle dá a essa concepção o nome de “Inteligência Artificial forte” ou “IA forte” (1992, p. 36).

A conclusão inevitável dessa assertiva é de que não haveria nada de essencialmente biológico acerca da mente humana. Qualquer sistema físico com as características corretas – ou seja, com as mesmas entradas e saídas de uma mente – apresentaria uma mente nos moldes da humana. Com isso, naturalmente o desenvolvimento da ciência levaria à construção de uma máquina com as mesmas características cognitivas do ser humano. Como visto, havia, no âmbito da ciência cognitiva, um entusiasmo genuíno a respeito da expansão dos limites da tecnologia computacional. Esse entusiasmo levou os estudiosos a acreditarem que era apenas questão de tempo o desenvolvimento de uma tecnologia capaz de reproduzir integralmente a mente humana.

Searle refuta essa concepção. Para ele, um computador digital jamais deixará de ser apenas um computador digital, guiado por regras puramente formais, descritas em símbolos abstratos, pelos quais a máquina está programada para operar – uma sequência numérica de zeros e uns, por exemplo. Aparecendo certa sequência de símbolos na tela do computador, ele executa determinada operação. No entanto, os símbolos não possuem qualquer significado para a máquina, não têm conteúdo semântico (SEARLE, 1992, p. 38).

Essa característica dos computadores é justamente a que lhes define, e, ao mesmo tempo, o que diferencia os programas dos processos mentais. Não existe conteúdo numa sequência numérica – ao menos não para a máquina, que apenas

executa a ação relacionada à sequência. Por outro lado, os estados mentais têm, necessariamente, um conteúdo. Por exemplo:

> Se estou a pensar em Kansas City, ou se desejo beber uma cerveja fresca, ou se estou a imaginar que vai haver uma baixa nas taxas de juro, em cada caso, o meu estado mental tem um certo conteúdo mental, além de quaisquer estruturas formais que possa ter. Isto é, mesmo se os meus pensamentos ocorrem em séries de símbolos, deve haver algo mais no pensamento do que as séries abstratas, porque as séries por si mesmas não têm qualquer significado. Se meus pensamentos são acerca de alguma coisa, então as séries devem ter um significado, que faz que os pensamentos sejam a propósito dessas coisas. Numa palavra, a mente tem mais do que uma sintaxe, possui também uma semântica. (SEARLE, 1992, p. 39)

Portanto, programa de computador algum tem capacidade de ser como a mente, já que os computadores não têm a questão semântica própria da mente humana.

Para defender seu posicionamento, Searle desenvolveu o argumento do quarto chinês, que se tornou bastante conhecido entre os estudiosos de filosofia da mente. Nesse argumento, Searle sugere que imaginemos a existência de um programa para falar chinês que o faça tão bem como qualquer falante nativo de chinês – um programa capaz de simular a compreensão do chinês pelo computador. Em seguida, imaginemos alguém que não saiba falar chinês fechado num quarto, onde há caixas com símbolos chineses, um grande livro em português, onde está explicado o programa de computador para falar chinês, e uma abertura no quarto para os *inputs* e *outputs*. Esse alguém, fechado no quarto, executaria o programa de computador. De vez em quando são introduzidos símbolos no quarto, representando perguntas feitas pelos falantes de chinês fora do quarto. Então o sujeito consultaria o grande livro – o programa –, pegaria outros símbolos e levaria a resposta correta para fora. Para as pessoas fora do quarto o sujeito seria como um falante nativo, embora, na realidade, ele não entenda uma palavra em chinês (SEARLE, 1980, p. 418).

Se está correto dizer que o indivíduo no quarto não entende chinês, como poderia estar correto afirmar que o computador que implementa o mesmo programa compreende chinês? Já que não é possível afirmar que o indivíduo no quarto compreende chinês, é necessário estender esse raciocínio e concluir que tampouco o computador compreende. O computador é capaz de manipular símbolos, no entanto, essa manipulação apenas faz sentido a partir do ponto de vista de um telespectador. Para o próprio computador, trata-se apenas de manipular os símbolos de acordo com

a programação prévia da máquina, pois o computador não é capaz de absorver o sentido das ações que executa (SEARLE, 1980, p. 418). Essa, portanto, é uma diferença fundamental entre a inteligência artificial e a mente humana.

Searle afirma (1992, p. 49) que os computadores não possuem intencionalidade. Como não é possível atribuir conteúdo representativo e modo psicológico aos computadores digitais, não lhes é possível atribuir intencionalidade.

Contudo, da perspectiva de um observador externo, os computadores digitais atuam *como se* possuíssem intencionalidade. Em virtude dessa aparência de intencionalidade, Searle desenvolve os conceitos de intencionalidade intrínseca e derivada.

A intencionalidade intrínseca é irredutível, já que está ligada aos estados da mente e não às capacidades representacionais externamente impostas. Ela estabelece e possibilita o significado. Na obra *Intencionalidade*, Searle apresenta a percepção (SEARLE, 2002, p. 53) e a ação intencional (2002, p. 111) como formas primitivas de intencionalidade intrínseca. A explicação realizada no início desta seção se refere à intencionalidade intrínseca.

A intencionalidade derivada, por sua vez, depende da existência de um observador. É a intencionalidade atribuída ao computador digital: embora não seja capaz de compreender o conteúdo das ações que realiza, um terceiro que observe as ações implementadas pela máquina será capaz de entender. A intencionalidade é derivada na medida em que só existe a partir da perspectiva do observador, que dá sentido à ação.

A ciência cognitiva parte da crença central de que os processos mentais funcionam de maneira análoga ao computador. Mas, para Searle, o pensamento não se limita à simples tarefa de encontrar os corretos *inputs* e *outputs* do cérebro, pois envolve a compreensão do conteúdo representativo (SEARLE, 1992, p. 45). Portanto, não seria possível reproduzir o processo cognitivo humano num computador digital, mesmo com os avanços da tecnologia computacional.

Searle também argumenta que o uso de representações, da forma proposta pelos cientistas cognitivos, é desnecessário para a compreensão dos processos mentais (1992, p. 55). Para os cientistas cognitivos, pensar é processar informação, e processar informação é manipular símbolos. Uma vez que o computador manipula símbolos, estudar o pensamento é estudar os programas computacionais de manipulação de símbolos, existam eles em computadores ou em cérebros.

A ciência cognitiva entende que, para a compreensão do pensamento (ou da cognição, na terminologia utilizada pelos cognitivistas), existem três níveis de estudo: o dos estados mentais conscientes; o das células nervosas e um terceiro, o nível do processamento do sistema de informação. O nível do processamento do sistema de informação, que nada mais é do que o nível das representações, fica entre o nível dos estados mentais conscientes e o nível biológico, das células nervosas. O nível da representação é o principal objeto de estudo das ciências cognitivas.

Searle afirma que o terceiro nível de estudo da mente acrescentado pelos cognitivistas – o nível do processamento de informação – é irrelevante. Para Searle, aqueles que aceitam a existência do nível da representação pressupõem que, por trás de todo comportamento significativo, existe uma teoria interna que o coordena. No entanto, essa teoria interna seria algo dispensável, que poderia ser substituída por uma hipótese mais simples: a de que a estrutura fisiológica do cérebro instiga determinados comportamentos ou ações, sem a necessidade de um nível intermediário de regras ou teorias.

Searle usa o fato de que os seres humanos não podem ver infravermelhos para exemplificar a teoria de que não existe o nível da representação. É sabido que os seres humanos não são capazes de ver infravermelhos; contudo, não é necessário que exista uma regra universal predizendo que os seres humanos não possam ver o infravermelho, simplesmente não dispomos de um aparato visual que nos capacite a tanto (SEARLE, 1992, p. 64). Então, por que a capacidade de ver infravermelhos seria diferente das manifestações do nível da *folk psychology,* ou psicologia popular? Não há argumento suficiente para concluir que, num caso, exista o nível do processamento de informação, e, no outro, não.

Quando tentamos replicar uma característica humana num computador, necessitamos de um intrincado processamento de informações. Mas isso não significa que nos seres humanos a expressão dessa característica venha acompanhada de semelhante processamento. A conclusão mais simples é que o processo biológico seja capaz de expressar tal característica sem a necessidade de um processamento de informações intermediário. Searle, portanto, busca fornecer uma explicação biológica aos fenômenos mentais.

Searle denominou “naturalismo biológico” a teoria que criou para explicar a relação entre os estados mentais e o mundo físico (SEARLE, 2002, p. 366). Para o

naturalismo biológico os estados mentais são como quaisquer outros fenômenos biológicos, como a digestão, por exemplo.

A teoria do naturalismo biológico representa um desenvolvimento do argumento de que a intencionalidade é inerente ao cérebro, argumento este que Searle já expôs em 1980, no artigo *Minds, brains and programs*. Para Searle, a intencionalidade intrínseca não pode advir de objetos que não têm um cérebro biológico, pois "apenas algo que tenha os mesmos poderes causais do cérebro pode ter intencionalidade" (SEARLE, 1980, p. 423). Como o cérebro é o responsável pela intencionalidade, é responsável também pelos estados mentais, e não existe intencionalidade nem estados mentais em um sistema que não possua um aparato biológico equivalente ao cérebro.

O naturalismo biológico lida com o problema mente-corpo argumentando que estados mentais são ao mesmo tempo causados pelas operações no cérebro e realizados na estrutura cerebral. A consciência seria um produto causal da atividade do cérebro, mas o cérebro não apenas *causa* os estados mentais como também os *realiza*. Isso impediria a existência de qualquer lacuna temporal entre as alterações neurobiológicas e os estados mentais (SEARLE, 2002, p. 370).

Trabalhando com os conceitos de atos de fala e intencionalidade, John Searle refutou as analogias entre a mente humana e o computador. Mas, se o processo de decisão não pode ser explicado por meio da metáfora computacional, como o seria? Para responder a essa pergunta, Searle desenvolve a própria teoria da racionalidade, utilizando-se dos pressupostos da intencionalidade. Para ele, a decisão se processa por meio de uma lacuna, sem que exista qualquer causa anterior suficiente.

Searle defendeu, portanto, o naturalismo biológico em sua explicação da mente humana. A partir da hipótese de que a explicação da mente deve ser feita por meio de processos biológicos Searle desenvolve uma teoria da racionalidade.

3.2 A TEORIA DA RACIONALIDADE

John Searle dedicou a obra *Rationality in Action* para tentar formular a própria teoria da racionalidade. Os fundamentos dessa obra são os mesmos que ele já havia exposto em estudos anteriores: o naturalismo biológico, a intencionalidade e os atos de fala.

Analisando como ocorre a ação, Searle conclui que, quando nós explicamos uma ação, normalmente não citamos condições antecedentes suficientes. Em uma dada situação, embora o sujeito *A* possa decidir por *X* ou por *Y*, ele decide por *X*. Não existe nada que o obrigue a optar por *X*. Essa ação, portanto, não é causalmente determinada.

John Searle argumenta que não existe causalidade embutida na estrutura de ações voluntárias (2011, p. 67). Todas as pessoas, quando fazem algo, salvo em condições excepcionais, têm a nítida sensação de que poderiam estar fazendo outra coisa. A pessoa que está lendo um livro poderia estar caminhando, lendo outra coisa; enfim, praticando uma série de atividades alternativas. Todos estamos imersos nesse senso de possibilidades variadas, e isso nos dá a convicção de que temos liberdade de agir.

No entanto, se as condições anteriores são insuficientes, como explicar que a pessoa escolha uma ação e não outra?

Em resposta a tal pergunta, Searle sugere a existência de uma lacuna, um salto, cuja nomenclatura em inglês é *gap*[4] [5]. Tal lacuna poderia ser descrita de duas maneiras. Em primeiro lugar, em relação ao futuro. Nesse caso, seria a característica da decisão e da ação na qual temos a impressão que decisões e ações alternativas nos estão disponíveis.

Em segundo lugar, a lacuna pode ser relatada em relação ao passado. Aqui, a lacuna seria a característica da decisão e da ação na qual experienciamos que as razões precedentes não eram causalmente suficientes para as decisões e ações (SEARLE, 2011, p. 61). O sujeito pode explicar os motivos que o levaram a fazer algo, mas essa explicação não é uma causa suficiente. Em circunstâncias normais, existe uma explicação, mas não existe uma causa que tenha determinado o comportamento presente.

---

[4] O termo *gap* utilizado por John Searle não se confundo com a expressão *explanatory gap*, introduzida pelo filósofo Joseph Levine no artigo *Materialism and qualia: The explanatory gap*. Levine usou o termo para expressar a dificuldade teórica de teorias fisiológicas em explicar como eventos físicos poderiam ensejar a experiência como é sentida pelo sujeito. Ele usou como exemplo a frase "Dor é a ativação das fibras C", argumentando que, embora tal assertiva seja fisiologicamente válida, não auxilia a entender como a dor é sentida. Para ele, embora a *gap* não seja algo fechado, a consciência é ao mesmo tempo física, significando que pode existir uma lacuna epistemológica, mas não ontológica.

[5] Embora as palavras “lacuna” e “salto” possuam, em língua portuguesa, significados distintos, se adequam perfeitamente à palavra inglesa *gap*, sobretudo no contexto utilizado por John Searle. Por esse motivo, nesse estudo, ambos os termos serão empregados de maneira intercambiável como tradução para *gap*, embora com predominância da utilização do vocábulo lacuna.

Existe uma visível diferença entre perguntar “por que você fez isso” e “por que isso aconteceu”. A primeira pergunta (por que você fez isso) pressupõe que é causalmente possível que outra coisa tivesse acontecido. Significa que várias ações poderiam ter sido realizadas, mas só uma foi. Então, perguntar “por que você fez isso” significa perguntar: em quais razões você, como ser racional, se apoiou? A resposta é uma demonstração de como um ser racional age na lacuna, não a tentativa de dar uma causa suficiente (SEARLE, 2011, p. 88). Embora as razões fornecidas não sejam condições causais suficientes, normalmente são perfeitamente adequadas como explicação. Já quando alguém pergunta “por que isso aconteceu”, não parte do mesmo pressuposto de que várias ações poderiam ter ocorrido. Pelo contrário: há uma crença prévia de que existe uma condição causal capaz de me explicar o motivo daquela ocorrência.

A lacuna pode se manifestar em três ocasiões distintas. A primeira ocorre no momento da tomada de uma decisão racional. Nesse caso, há uma lacuna entre o processo deliberativo e a decisão em si. A segunda ocasião é quando alguém decidiu fazer algo (ou, usando a terminologia de John Searle, formou uma *prior intention*). Aqui, há uma lacuna entre a decisão e o início da ação. A terceira manifestação da lacuna ocorre quando alguém está no curso de uma ação que se prolonga no tempo: há uma lacuna entre o início da ação e o processo de executá-la em todas as suas fases (SEARLE, 2011, p. 62 e 63).

Para Searle, uma clara manifestação da lacuna na vida real estaria relacionada a situações nas quais há mais de uma razão para agir. Quando uma pessoa possui vários motivos para a ação, pode escolher um só deles como o motivo pelo qual age. Se o sujeito quer comprar um carro novo, por exemplo, pode estar preocupado com a própria segurança, pode querer ser capaz de chegar mais rápido a seus compromissos, ou pode estar em busca de prestígio social. Ainda que as três razões sejam relevantes, a razão determinante pode ser a busca de prestígio. Essa razão pode nunca ser revelada pelo sujeito.

Em situações assim, não são as crenças e desejos que levam a pessoa a agir de determinado modo, mas sim é a própria pessoa que decide o desejo em relação ao qual ela queria agir. É o sujeito quem decide, dentre todas as causas existentes, qual será a efetiva. Assim, todas as razões existentes são consideradas efetivas *pelo agente*, cabendo a ele escolher com base em qual delas irá agir (SEARLE, 2011, p. 65).

Searle não concorda com as premissas do que ele denomina Modelo Clássico de Racionalidade. Para o Modelo Clássico as ações são causadas por crenças e desejos, mas, para Searle, se essa "causa" for considerada "causa suficiente", essa premissa é simplesmente falsa, já que em situações onde há liberdade não há causa suficiente, mas sim explicação das razões para agir (SEARLE, 2011, p. 70).

Se crenças e desejos fossem suficientes para causar ações, simplesmente assistiríamos a ação acontecer em nós. Tudo se passaria como se estivéssemos vendo um filme. Não haveria esforço para decidir nada, as decisões simplesmente surgiriam por si mesmas. Mas ocorre o contrário: as decisões são fruto de um contínuo e inevitável esforço pessoal.

Searle argumenta que as crenças não podem ser um antecedente causal necessário e eficiente da ação em virtude de que a direção de satisfação da crença não existe na ação (2011, p. 60). Foi visto na seção 3.1 que a crença é um estado intencional que possui direção de adequação mente-mundo. A condição de satisfação da crença é que corresponda ao que acontece no mundo, o que a torna verdadeira. Mas a ação não se processa da mesma forma.

A ação é composta por dois elementos: uma intenção-em-ação e um movimento corporal. A intenção-em-ação é a intenção do sujeito no momento em que está executando a ação. Caso a ação seja premeditada, haverá uma intenção prévia, ou *prior intention* (SEARLE, 2011, p. 44). Tanto a intenção-em-ação quanto a intenção prévia são causadas por um estado intencional antecedente.

O estado intencional seria o elemento antecedente à ação. Ele dá origem à intenção prévia (facultativa, presente apenas quando a ação é premeditada). No momento da ação, entram em cena a intenção-em-ação (também decorrente do estado intencional) e o movimento corporal. Essa seria a estrutura da ação (SEARLE, 2011, p. 49).

Ao contrário do que acontece com as crenças, as ações não possuem condição de satisfação. A intenção-em-ação e intenção prévia sim: elas serão satisfeitas se o movimento corporal ocorrer. A ação é a condição de satisfação dessas duas modalidades de intenção. Portanto, as crenças não podem ser consideradas causas suficientes para uma ação (SEARLE, 2011, p. 60).

Searle exclui o computador da linha da racionalidade (2011, p. 66). Para tanto, usa de fundamentos semelhantes aos que apresentou para refutar a existência de processos mentais em computadores, conforme analisado na seção 3.1. O

computador não seria racional ou irracional, pois seu comportamento está totalmente determinado por seu programa e a estrutura de seu *hardware*. Assim como o computador só possui intencionalidade a partir do ponto de vista de um observador, também só na perspectiva do observador o computador pode ser considerado racional.

Cumpre destacar que, embora Searle não use o termo “decisão social” na obra *Rationality in action*, muitos dos exemplos que fornece se enquadram a esta modalidade de decisão. É o caso da eleição de um candidato, exemplo que ele repete em diversas ocasiões, como nas páginas 13, 50, 65 e 72, e é retomado na seção 5.3 deste estudo. Dessa forma, pode-se concluir que a teoria também abarca a decisão social, e, portanto, está relacionada aos estudos desta dissertação.

Utilizando-se dos conceitos de atos de fala e intencionalidade, Searle desenvolve uma teoria da racionalidade na qual não é possível comparar o processo de decisão humano ao de um computador digital. O computador atua de acordo com a programação recebida, sem intencionalidade. Já os seres humanos agem com intencionalidade e sem vinculação necessária a uma causalidade anterior.

Contudo, o fato de Searle se apoiar na intencionalidade para excluir o computador da linha de racionalidade foi objeto de muitas críticas, e a próxima seção irá abordar algumas delas.

3.3 AS CRÍTICAS AO CONCEITO DE INTENCIONALIDADE DE JOHN SEARLE

Searle foi alvo de crítica por diversos autores ao relacionar a intencionalidade ao cérebro humano. Dentre eles, Zenon Pylyshyn, filósofo e cientista cognitivo, buscou expor a fragilidade das explicações de Searle quanto à relação entre intencionalidade e cérebro. No artigo *The “causal power” of machines*, Pylyshyn argumenta que é obscura a relação entre cérebro e intencionalidade, da maneira como Searle a define, e critica a teoria de Searle de que somente sistemas equivalentes ao cérebro humano podem ter intencionalidade, como se o cérebro *causasse* a intencionalidade (PYLYSHYN, 1980, p. 442).

Pylyshyn se utiliza do exemplo da substituição gradual de partes do cérebro por chips de computador como argumento contra a teoria de que o cérebro causa a intencionalidade. Supondo que fosse possível substituir partes do cérebro por chips de computador, e que tais chips funcionassem exatamente como neurônios,

permitindo ao seu portador, por exemplo, falar por intermédio de tais chips, seria adequado afirmar que o indivíduo que fala não quer dizer algo – ou seja, perde a intencionalidade? Se, como quer Searle, a intencionalidade fosse uma substância própria do cérebro, essa pergunta deveria ser respondida afirmativamente. No entanto, para Pylyshyn, nada diferenciaria um cérebro normal de um cérebro repleto de chips de silício no tocante à intencionalidade do sujeito (PYLYSHYN, 1980, p. 442).

O cientista cognitivo Douglas Hofstadter, por sua vez, critica a experiência do quarto chinês apresentando a *resposta dos sistemas*, ou seja, que o entendimento deveria ser atribuído ao sistema todo, não às partes dele (HOFSTADTER, 2000, p. 374). A falha essencial no raciocínio de Searle é facilmente identificada se compararmos o sistema do quarto chinês ao funcionamento mental dos seres humanos: nossa consciência também ocorre a partir de neurônios que, individualmente, não tem compreensão do que estão fazendo.

Considerando que Searle busca embasar sua teoria nos mecanismos biológicos do cérebro e na intencionalidade, sem recorrer ao nível da representação simbólica – como fazem quase todos os cognitivistas –, para Gardner, as afirmações de Searle a respeito da intencionalidade não ganham contornos científicos verdadeiros (GARDNER, 2003, p. 191). Para que os argumentos alcançassem uma verdadeira cientificidade, seria necessário que Searle explicasse com exatidão o que entende por intencionalidade, e fosse capaz de demonstrar que ela se restringe a cérebros orgânicos.

Daniel Dennett também formulou críticas em face do conceito de intencionalidade de John Searle. Na obra *A perigosa ideia de Darwin*, Dennett observa como John Searle ignorou os postulados da teoria da evolução pela seleção natural[6]. Isso é ainda mais digno de nota se observarmos que Searle, afastando-se das teorias computacionais da mente, procurou fornecer aos processos mentais explicações naturalistas.

Daniel Dennett defende que a seleção natural pode ser vista como um algoritmo[7] (1998, p. 55). Mas não se trata de algoritmos que comprovadamente

---

[6] A teoria da evolução pela seleção natural será objeto de análise na seção 5. Por isso, nesse momento não serão tecidos maiores esclarecimentos a respeito da teoria, além dos considerados essenciais para a compreensão da crítica de Daniel Dennett ao conceito de intencionalidade formulado por John Searle.

[7] Um algoritmo é um processo ordenado por regras pré-definidas, que determina como proceder para a realização de uma tarefa (resolução de problemas, cálculos etc.). A palavra deriva do nome do matemático persa Al Khowarizmi, autor de um manual de álgebra no século IX (TEIXEIRA, 1998, p. 20).

computem determinadas funções matemáticas, e sim algoritmos que classifiquem e construam coisas.

Nesse sentido, Dennett expõe dois exemplos que ilustram o argumento: o de um torneio de tênis e de um torneio de cara ou coroa. Suponhamos que exista um torneio de tênis nacional, no qual todos podem participar. Como o tênis é um esporte que pressupõe habilidade, embora uma parcela de sorte seja parte inevitável do jogo, é de se concluir que o jogador mais habilidoso irá vencer o torneio. Assim, se houver outro torneio no ano seguinte, é bastante provável que pelo menos alguns dos resultados se repitam; ou seja, os primeiros jogadores tendem a permanecer nos primeiros lugares, salvo se algo excepcional acontecer (DENNETT, 1998, p. 57).

Por outro lado, num torneio nacional de cara ou coroa, é apenas o elemento sorte que se destaca. Não há habilidade em se vencer na cara ou coroa, de modo que, a cada torneio, os resultados finais apresentarão vencedores totalmente diferentes.

Apesar disso, tanto numa quanto noutra situação, o que se verifica é um processo eliminatório, onde o vencedor derrota o oponente sucessivas vezes. Se o torneio tiver trinta etapas, seja qual for o elemento exigido do jogador – sorte ou habilidade – há uma única certeza quanto ao resultado final: o vencedor será aquele que vencer trinta vezes seguidas.

O algoritmo, no contexto acima, “considera como dados de entrada um conjunto de competidores e garante terminar identificando um único vencedor” (DENNETT, 1998, p. 55). A questão central, aqui, é entender que o vencedor nem sempre será o portador de maior habilidade ou sorte. O algoritmo não precisa ter objetivo nem propósito. Portanto, é perfeitamente possível que haja um torneio no qual aqueles que seguem para a próxima fase sejam os perdedores. Embora não seja uma regra inteligente, não contraria o conceito de algoritmo, que pressupõe unicamente que exista mais de uma entrada e uma única saída, considerando algum tipo de regra, por mais sem sentido que a regra aparente ser.

Dennett apresenta a seleção natural como um processo algoritmo. Sem finalidade, sem a presença de uma Mente primordial que guie o estado das coisas. É justamente esse o ponto de maior controvérsia após a disseminação dos estudos de

---

O algoritmo estabelece passos que devem ser necessariamente seguidos para a solução do problema. Portanto, ele tem como essência a apresentação um conjunto de regras que deve ser cegamente obedecido (TEIXEIRA, 1998, p. 22).

Charles Darwin. Enquanto aceitar a concepção da seleção natural seria possível, partindo de um corte histórico – o momento atual, ou o surgimento do homem –, aceitar a mesma concepção desde o surgimento da vida na Terra levaria a conclusões que escapariam ao campo da biologia. Pensar no surgimento da vida em si como fruto de um elemento aleatório, e no aparecimento do ser humano como decorrência do processo de seleção natural, proporcionaria um necessário alargamento da teoria da evolução pela seleção natural para outros campos, tais como a cosmologia, a política, a filosofia, a psicologia e a religião (DENNETT, 1998, p. 66). Por esses motivos, muito da resistência ao darwinismo se exteriorizou por meio dessa contínua busca pela inteligência primordial, se não agora, no aparecimento da vida. Contudo, para Dennett, com base nos pontos apresentados, existem elementos suficientes para extirpar a ideia de uma mente inteligente, primordial, que tenha criado a complexidade da vida como a conhecemos hoje, ou seja, de uma Mente original.

Para Dennett, Searle é um dos estudiosos que busca a Mente original, a causa primeira: inclusive teria denominado essa Mente primordial de intencionalidade original (DENNETT, 1998, p. 415). Dennett argumenta que a divisão formulada por Searle entre intencionalidade original e derivada é incompatível com a teoria darwinista, uma vez que, pelos pressupostos da seleção natural, descendemos de autômatos sem intencionalidade (1998, p. 415). A explicação de Searle seria, portanto, um contrassenso, pois deveria explicar quando e como deixamos para trás nosso passado de autômatos e nos transformamos em seres com intencionalidade.

O ponto de vista de Searle a respeito da ausência de função do que existe na natureza é fundamental para a crítica de Dennett. Para Searle, não existem funções intrínsecas na natureza: elas existem apenas do ponto de vista do observador (SEARLE, 1996, p. 14). Desenvolvendo esse raciocínio, Searle expõe o seguinte:

> Estamos cegos para este fato por causa da prática, especialmente na biologia, de se falar de funções como se fossem intrínsecas à natureza. Mas exceto para aquelas partes da natureza que são conscientes, a natureza não sabe nada da função. É, por exemplo, intrínseco à natureza que o coração bombeia o sangue, e com isso faz com que ele atravesse o corpo. É também um fato intrínseco da natureza que o movimento do sangue está relacionado com uma série de outros processos causais que têm a ver com a sobrevivência do organismo. Mas quando, ao invés de dizermos “O coração bombeia o sangue”, dizemos, “A função do coração é bombear o sangue”, estamos fazendo algo mais do que registrar esses fatos intrínsecos. Estamos situando esses fatos em relação a um sistema de valores que possuímos (1996, p. 14-5).

Em resumo, de acordo com Searle, o coração não possui função alguma: nós lhe atribuímos uma função. Searle argumenta que apenas artefatos feitos por artífices humanos conscientes possuem funções reais: as asas do avião foram feitas para voar; as de um pássaro, não. Segundo Dennett, a partir desse pensamento de Searle, pode-se concluir que "o discurso de função na biologia, como mero discurso de intencionalidade do tipo *como se fosse*, não é para ser levado a sério" (1998, p. 417). À conclusão de Searle, de que na natureza não existe uma variedade de mecanismos exibindo projetos, acrescenta Dennett que para Searle "só os artefatos humanos têm essa honra, e apenas porque [...] é preciso que haja uma Mente para fazer algo que tenha uma função!" (DENNETT, 1998, p. 417).

Dennett afirma que a teoria de John Searle a respeito da intencionalidade original, entendida como uma "propriedade inatingível em princípio por qualquer processo de pesquisa e desenvolvido para construção de algoritmos cada vez melhores" (DENNETT, 1998, p. 418), não passa de uma versão da crença na ideia de Mente primordial. As mentes, conforme Searle, seriam fontes originais, que explicariam e dariam sentido aos projetos – não seriam o resultado do projeto, conclusão inafastável da teoria da evolução pela seleção natural.

Dennett apresenta um exemplo com o intuito de criticar as objeções apontadas por Searle contra a IA forte (DENNETT, 1998, p. 442-447). Suponhamos uma situação onde um ser humano, utilizando-se de uma avançada tecnologia de criogenia, resolve congelar a si mesmo, com o objetivo de acordar após quatrocentos anos. Como garantir que o corpo permanecerá num local seguro por quatro séculos? Certamente não é possível contar com a boa vontade dos descendentes para tanto, tampouco existe uma empresa que com certeza permanecerá todo esse tempo em atividade. A melhor solução seria que o corpo ficasse em uma cápsula, e que essa cápsula fosse capaz de proteger o corpo contra os perigos ambientais.

Existem duas possibilidades para a movimentação dessa cápsula: ela pode permanecer fixa ou ser capaz de se mover, afastando-se, assim, do perigo iminente. Essa é basicamente a diferença existente entre plantas e animais. Suponhamos que exista tecnologia avançada o suficiente para fazer com que a cápsula se mova. Essa cápsula seria, então, equipada com sensores de calor, frio e umidade, de modo que se afastaria de um local sempre que notasse que ali não estaria mais em segurança.

Essa cápsula, portanto, assumiu ares de um robô extremamente complexo. Esse robô, após ter sido colocado em atividade, passa por diversas experiências de perigo, e tem sucesso em escapar de cada uma delas. Com o passar do tempo, o robô é capaz de relacionar situações de perigo no passado com situações de potencial perigo no presente. Ele é capaz de prever situações perigosas de forma mais acurada. O robô, portador do corpo do humano que o criou, se torna um especialista em sobrevivência.

Mas, quem garante que esse indivíduo será o único decidido a preservar o próprio corpo pelos próximos séculos? É possível que muitos outros compartilhem de tal desejo, e, para isso, desenvolvam robôs semelhantes, que começarão a disputar os recursos existentes. Dessa forma, cada robô terá que aprender a interagir com os demais, calcular se lhe é vantajoso cooperar com outros robôs, ou mesmo se deve tentar destruí-los. É possível que cada robô desenvolva elevado nível de autocontrole, tudo para ser capaz de permanecer existindo.

Para um observador externo, o robô parece estar apenas preocupado com a própria sobrevivência. É até mesmo possível que o próprio robô, como agente autônomo, esqueça-se da meta principal – manter o corpo que transporta vivo por quatro séculos – e se empenhe na realização das próprias metas secundárias, destinadas à manutenção de si mesmo. Nada impede que o robô se torne imprudente, até mesmo suicida, agindo como se a manutenção da própria existência fosse tudo o que importasse.

Poderíamos atribuir inteligência a tal robô? Caso utilizemos o conceito de Steven Pinker, que define inteligência como o empenho para atingir objetivos em face de obstáculos (PINKER, 1998, p. 393), a resposta seria afirmativa. Os objetivos seriam imprescindíveis à inteligência; sem eles, a inteligência não tem razão de ser. Então, podemos nos perguntar: quem cria tais objetivos? No caso do robô, é o sujeito que queria permanecer vivo por quatrocentos anos. Quando se trata de inteligência artificial, o programador é o criador dos objetivos. No caso dos animais em geral, e dos seres humanos em particular, encontrar o “programador” pode ser mais complexo.

O objetivo supremo dos seres vivos vem da seleção natural (PINKER, 1998, p. 393). O cérebro de um ser vivo vai se esforçar para colocar o dono em circunstâncias como aquelas que levaram seus ancestrais a se reproduzir. Não é que o cérebro *queira* se reproduzir: ele apenas repetirá comportamentos que levaram os ancestrais

a obter o êxito reprodutivo – embora tais comportamentos possam, a longo prazo, ser prejudiciais ao indivíduo ou à própria espécie.

A criação de um robô com as características descritas, embora improvável diante da tecnologia atual, não é impossível. Caso ele exista, se analisarmos a intencionalidade dele com base na teoria de John Searle, teremos que afirmar que nosso robô não possui intencionalidade original, mas apenas derivada. Por mais que o robô seja hábil em administrar os próprios interesses, por mais envolvido que esteja com as próprias metas secundárias, a intencionalidade dele ainda seria derivada. Dennett, então, estende o raciocínio para os seres humanos: nossos genes seriam a fonte original de nossa intencionalidade (DENNETT, 1998, p. 446), assim como o indivíduo congelado foi a fonte da intencionalidade do robô. Para Dennett, nós, seres humanos, nos vemos como senhores absolutos de nossa vontade, no entanto somos seres criados para proteger nossos genes, de modo que a distinção entre intencionalidade original e derivada, nos exatos termos propostos por Searle, levaria a concluir que a intencionalidade dos humanos também não passa de uma intencionalidade derivada.

Portanto, partindo de pressupostos relacionados à teoria da evolução pela seleção natural, que será analisada na seção 5 desta dissertação, Dennett apresenta argumentos no sentido de que a separação entre a intencionalidade original e derivada não é tão acentuada como Searle quer levar a crer. Searle procurou explicar os estados mentais por meio do conceito de intencionalidade, e atribuiu a verdadeira intencionalidade exclusivamente aos seres humanos, afirmando que máquinas possuem apenas intencionalidade derivada, que existe apenas da perspectiva do observador. Mas, ao levar em consideração a seleção natural, Dennett enfraquece a justificativa de Searle para a existência de uma intencionalidade original.

Searle teve o mérito de apresentar uma perspectiva crítica sobre um elemento da ciência cognitiva, a analogia entre o funcionamento do computador e o processo cognitivo humano. Ele argumentou que o fator biológico não pode ser desconsiderado, como quiseram os cientistas cognitivos clássicos. Contudo, Searle rejeito a metáfora computacional com base na teoria da intencionalidade, sem levar em conta a seleção natural ou o elemento emocional.

Contudo, as críticas tecidas por Zenon Pylyshyn, Douglas Hofstadter e Daniel Dennett ao conceito de intencionalidade de John Searle apontam no sentido de que Searle não teria conseguido alcançar o intento de fornecer uma explicação totalmente

biológica aos estados mentais, e que há dúvidas sobre a diferença entre a intencionalidade originária e derivada ser tão marcante como Searle propõe. Adotando as opiniões dos críticos de Searle, o presente estudo entende que a argumentação dele é insuficiente para refutar a crença da ciência cognitiva de que o computador é o modelo mais adequado para a compreensão do processo cognitivo. É preciso, portanto, buscar outras teorias para prosseguir com a análise a respeito da capacidade a ciência cognitiva de explicar o processo de decisão social.

Na próxima seção, serão apresentados estudos neurocientíficos que indicam que a emoção é inerente ao processo de decisão social. Será visto como, ao introduzir o elemento emocional, as teorias formuladas pelos neurocientistas tornam o modelo proposto pela ciência cognitiva insuficiente para explicar as decisões sociais.

4 A INFLUÊNCIA DAS EMOÇÕES NO PROCESSO DE DECISÃO SOCIAL NA PERSPECTIVA DA A NEUROCIÊNCIA

Na primeira seção viu-se como os cognitivistas se utilizaram das analogias entre o computador digital e a mente humana para tentar explicar a cognição, aqui incluída a decisão social. Na segunda seção foi apresentada a crítica de John Searle ao modelo cognitivista, e a teoria da racionalidade por ele formulada. Notou-se que, embora Searle tenha destacado o papel do naturalismo biológico em suas teorias, se fundou verdadeiramente no conceito de intencionalidade para negar a racionalidade aos computadores. No entanto, os críticos de Searle argumentam que alegar que há intencionalidade original para seres humanos e derivada para máquinas não é suficiente para desacreditar as analogias entre mente e computador.

Por isso esta seção analisará como a neurociência introduz o elemento emocional no processo de decisão social, contrariando a característica metodológica da ciência cognitiva que preconiza a exclusão das emoções dos estudos.

A explicação biológica dos fenômenos mentais atingiu um novo patamar na década de 1990, com o desenvolvimento dos estudos neurocientíficos. O surgimento de novas ferramentas permitiu estudar do cérebro humano de maneira inédita, e os neurocientistas começaram a observar a influência do fator emocional no processo de decisão social. A neurociência, portanto, não corroborou uma das características metodológicas da ciência cognitiva.

A importância da apresentação de estudos neurocientíficos nesta dissertação reside em que a neurociência foi capaz de fornecer dados empíricos indicando que a emoção é essencial para a tomada de decisão social. Ademais, o desenvolvimento dos estudos neurocientíficos mostrou que as analogias entre o computador digital e os processos cognitivos não se justificam, já que os estados e processos mentais não ocorrem no cérebro de maneira análoga aos processos computacionais.

Na presente seção, será objeto de análise a teoria formulada pelo neurocientista António Damásio na obra *O Erro de Descartes*, onde ele apresenta o fator emocional como fundamental ao processo decisório humano. Também serão expostos alguns experimentos, tais como o relacionado aos dilemas morais, realizado por Joshua Greene[8] e sua equipe, publicado no artigo *An fMRI investigation of*

[8] Joshua Greene é professor de psicologia na Universidade de Harvard, e dirige o Departamento de Cognição Moral. Seus estudos focam na junção entre psicologia, neurociência e psicologia.

emotional engagement in moral judgment (GREENE et al., 2001, pp. 2105-2107), e a variação do experimento de Greene, realizada por Michael Koenigs [9] no artigo *Damage to the prefrontal córtex increases utilitarian moral judgements* (KOENIGS et al., 2007, pp. 908-911), que também apontam para a relação entre decisão social e emoção.

Ao avançarmos no estudo da presente seção, também veremos os problemas das explicações neurocientíficas, e os motivos pelos quais devem ser analisadas com cautela.

4.1 A FILOSOFIA E A NEUROCIÊNCIA

Os estudos neurocientíficos influenciaram a discussão filosófica sobre a cognição humana. A neurociência, embora seja uma disciplina relativamente jovem, possui grande importância atualmente, sobretudo se consideramos a constante exposição de seus resultados na mídia.

Mas os avanços da neurociência não seriam possíveis sem o desenvolvimento das técnicas de neuroimagem. Como explica João de Fernandes Teixeira na obra *Filosofia do Cérebro,* tais técnicas surgiram na metade da década de 1990 e provocaram uma verdadeira revolução na capacidade de compreensão do cérebro humano. Antes, o cérebro só era observável por meio de autópsias, que geralmente se realizavam em pacientes com alguma disfunção cognitiva (TEIXEIRA, 2012, p. 11).

A partir da neuroimagem, cujas técnicas principais são o PET (*Positron Emission Tomography*) e o fMRI (*Functional Magnetic Ressonance Imaging*), esse panorama mudou. Foi possível visualizar o cérebro em atividade. Tais técnicas funcionam, basicamente, da seguinte maneira: o paciente (humano ou não) realiza alguma atividade cognitiva enquanto tem o cérebro escaneado. Eventos neurais requerem oxigênio, portanto fluxo sanguíneo aporta para os locais em atividade. O exame permite verificar a área do cérebro onde houve maior aporte de fluxo sanguíneo, localizando a atividade no tecido cerebral. Em tese, tais técnicas possibilitariam o mapeamento integral do cérebro por meio da identificação de todas as áreas envolvidas em quaisquer atividades humanas.

[9] Michael Koenigs é professor de psiquiatria na Universidade de Wisconsin-Madison. Tem estudos publicados na área de neurociência social e afetiva.

O desenvolvimento da neurociência mudou o cenário do estudo da mente. Entre os neurocientistas, há um entusiasmo generalizado – por muitos considerado ingênuo (TEIXEIRA, 2012, p. 16) – de que todas as respostas para o problema mente-cérebro serão dadas pela neurociência. Na filosofia, por exemplo, Patricia e Paul Churchland, que serão estudados a seguir, são de opinião que a neurociência resolverá todos os problemas filosóficos. Os neurocientistas passaram a tentar responder questões filosóficas clássicas a partir de estudos neurocientíficos. Isso deu origem a uma nova disciplina, a *neurofilosofia*, que será objeto de análise a seguir.

O termo neurofilosofia é atribuído a Patricia Smith Churchland, que o desenvolveu na obra *Neurophilosophy.* A neurofilosofia é uma tentativa de criar uma teoria unificada entre a filosofia e a neurociência para explicar o problema mente-cérebro, utilizando-se de estudos neurocientíficos para responder problemas clássicos de filosofia, tais como o funcionamento da mente e o processo de tomada de decisão.

Os neurofilósofos tendem a apresentar conclusões condizentes com a *doutrina do neurônio.* Tal termo, cunhado pelos filósofos Ian Gold e Daniel Stoljar[10] (1999, p. 803), expressa uma doutrina que pressupõe que a mente será integralmente explicada pela neurociência. Os que se filiam à doutrina do neurônio acreditam que surgirá uma teoria neurocientífica inteiramente biológica que explicará a mente em sua completude.

Dentre as teorias que atribuem a existência dos eventos mentais exclusivamente ao cérebro estão o materialismo reducionista e o materialismo eliminativista.[11]

---

[10] Ian Gold é um pesquisador canadense, das áreas de filosofia e psiquiatria, na Universidade McGill em Montreal.
Daniel Stoljar é professor de filosofia na Universidade Nacional da Austrália.
Juntos, eles escreveram o artigo *A neuron doctrine in the philosophy of neuroscience*. Graças ao artigo, o termo "doutrina do neurônio" se popularizou, designando doutrinas que pretendem explicar a mente apenas por meio da neurociência.

[11] A neurofilosofia é uma explicação monista para o problema mente-corpo. O problema mente-corpo é uma questão filosófica que busca compreender como os estados mentais se relacionam ao corpo físico. Ontologicamente, resolver o problema mente-corpo seria explicar a real natureza dos estados e processos mentais (CHURCHLAND, 1988, p. 7).
Existem duas correntes filosóficas que buscam responder ao problema mente-corpo: o monismo e o dualismo.
O dualismo explica que os estados mentais residem em algo não físico. Embora não seja uma teoria atualmente aceita pela comunidade científica em geral, é uma teoria intuitiva da mente. A população em geral tende a acreditar no dualismo, que está presente na maioria das religiões (CHURCHLAND, 1988, p. 7).

O materialismo reducionista, ou teoria da identidade, preconiza que cada estado mental é idêntico a um estado físico. Em 1956, Ullin T. Place publicou o artigo *"Is consciousness a brain-process?"* (referenciado na versão de 2002). No artigo, Place defende uma explicação inteiramente científica aos processos mentais (2002, p. 55), o que é alcançado ao identificá-los com os processos cerebrais. Uma vez que as pessoas só possuem experiências conscientes quando certos processos ocorrem em seus cérebros, é possível concluir que os processos cerebrais explicam inteiramente a experiência consciente.

Como expressa Richard Boyd no artigo *Materialism without reductionism*, se considerarmos verdadeira a proposição "Água = H_2O", também a proposição "Dor = ativação da fibra-C" será necessariamente verdadeira (BOYD, 1980, p. 83). A ativação da fibra-C *é o mesmo que* a dor, de modo que os eventos mentais são o mesmo que o estado físico.

Seguindo os pressupostos da teoria da identidade, Paul e Patricia Churchland desenvolveram a teoria que denominaram *materialismo eliminativo*. Paul Churchland conceitua o materialismo eliminativo já no primeiro parágrafo de um de seus mais conhecidos artigos, *Eliminative Materialism and the Propositional Attitudes*:

> é a tese de que a nossa concepção de senso comum dos fenômenos psicológicos constitui uma teoria radicalmente falsa, uma teoria fundamentalmente tão defeituosa que tanto os princípios quanto a ontologia dela serão eventualmente substituídos, ao invés de homogeneamente reduzidos, por uma neurociência completa (Churchland, 1981, p. 67).

Para o casal Churchland, o materialismo reducionista é falso, pois é impossível estabelecer uma correspondência exata do que ocorre no cérebro com os estados

---

O monismo, ao contrário, afirma que a explicação dos fenômenos mentais está na matéria. Existem várias teorias monistas, dentre as quais podemos destacar o behaviorismo filosófico, o materialismo reducionista, o funcionalismo e o materialismo eliminativo.

O behaviorismo filosófico foi analisado na seção 2.2. Para Paul Churchland (1988, p. 7), o behaviorismo não foi tanto uma teoria que tentou explicar o que são os estados mentais, mas sim uma teoria sobre como analisar ou entender o vocabulário que utilizamos para falar sobre eles. Mas a exclusão da introspecção foi vista como uma falha do behaviorismo, o que levou à criação de novas teorias para tentar explicar o problema mente-corpo.

O funcionalismo também foi visto na seção 2.2. Segundo a concepção funcionalista, os estados mentais não dependeriam do *hardware*, ou seja, da concepção física, mas sim do *software*, que seria o programa.

O materialismo reducionista e o materialismo eliminativo são teorias monistas que relacionam os estados mentais a eventos neurológicos. Portanto, tanto o materialismo reducionista quanto o eliminativo são concepções relacionadas à neurofilosofia, motivo pelo qual serão analisados no corpo do texto desta seção.

mentais da maneira como são percebidos pelo senso comum. Isso porque o senso comum interpretaria os estados físicos do cérebro de maneira equivocada. Para evitar essa interpretação errônea eles propõem a eliminação da terminologia relacionada aos fenômenos psicológicos.

Cumpre ressaltar que, ao conceituar o materialismo eliminativo, Paul Churchland afirmou que as concepções atuais dos fenômenos psicológicos não serão *reduzidas,* mas sim *substituídas* pela neurociência completa. O materialismo eliminativo não é, portanto, uma teoria reducionista. O reducionismo pressupõe que uma teoria seja abarcada por outra. Ernest Nagel, na obra *The Structure of Science*, explica que a redução pode ser homogênea ou heterogênea (1995, p. 338). A redução homogênea ocorre quando uma teoria antiga é absorvida por uma nova, mais abrangente. Desse modo, os eventos explicados pela teoria antiga passam a ser explicados pela nova, sem alterar substancialmente o sentido da velha teoria, que é incorporada.

Por outro lado, as reduções heterogêneas ocorrem quando a teoria nova, que abarca a antiga, havia sido criada para lidar com fenômenos qualitativamente distintos da antiga teoria. Os conceitos da teoria antiga não são os mesmos da nova. Dessa forma, é preciso introduzir regras de correspondência para que a redução possa ser efetuada.

O que o materialismo eliminativo pretende não é reduzir os termos da antiga teoria, a psicologia popular, à nova, neurocientífica. O objetivo é *eliminar* a teoria anterior. Uma vez que os conceitos da psicologia popular são falsos e causam inúmeras confusões conceituais, não devem ser preservados: devem ser eliminados e integralmente substituídos por conceitos da neurociência.

De acordo com o materialismo eliminativo, em breve poderemos abrir mão de termos psicológicos para designar estados da consciência. Estados intencionais, como crenças e desejos, serão integralmente explicados pela neurociência. Não será mais necessário, portanto, usar dos termos imprecisos da psicologia popular – em inglês, *folk psychology.* Todo e qualquer estado psicológico poderá ser explicado em termos materiais. Com isso, a psicologia desapareceria: não seria preciso fornecer explicações psicológicas para eventos que podem ser integralmente explicados pela neurociência.

Para exemplificar o que irá acontecer com termos da psicologia popular, Paul Churchland cita alguns exemplos de conceitos que caíram em descrédito com o

passar do tempo. Dentre eles, está o conceito de bruxa e possessão demoníaca. Segundo ele:

> A psicose é um distúrbio razoavelmente comum entre os seres humanos, e, séculos atrás, suas vítimas eram regularmente vistas como casos de possessão demoníaca, como instâncias do próprio espírito de Satã, observando-nos malevolamente através dos olhos de suas vítimas. A existência das bruxas não era questão de controvérsia. Elas eram vistas ocasionalmente, em cidades ou aldeias, envolvidas em comportamentos incoerentes, paranoicos, ou mesmo homicidas. Mas, observáveis ou não, por fim chegamos à conclusão de que bruxas não existem. [...] As teorias modernas dos distúrbios mentais resultaram na eliminação das bruxas de nossa ontologia séria. (CHURCHLAND, 2004, p. 81)

Portanto, de acordo com as premissas do materialismo eliminativo, no futuro os conceitos da psicologia popular, tais como desejo, crença e sensação, serão também eliminados do vocabulário científico. Com o desenvolvimento da neurociência, esses termos serão substituídos por vocábulos neurocientíficos. Será o fim da confusão conceitual que o uso de termos psicológicos proporciona.

Saulo de Freitas Araújo, na obra *Psicologia e Neurociências*, efetua diversas críticas ao materialismo eliminativo. Dentre elas, podemos ressaltar o próprio conceito de psicologia popular no qual os Churchlands se apoiam (ARAÚJO, 2011, p. 55): a psicologia popular seria, para ele, mais ampla do que a versão empobrecida apresentada pelos Churchlands. A eliminação pretendida se torna mais complicada se pensarmos na extensão dos termos que compõem a psicologia popular.

Ademais, os Churchlands desconsideram a diferença entre redução nomológica e ontológica (ARAÚJO, 2011, p. 57). Uma redução nomológica fornece a explicação necessária para o fenômeno empírico, de modo que é possível dizer que uma sensação decorre de determinado evento neural. Isso é diferente de dizer que a mesma sensação *é* o evento neural, o que seria uma redução ontológica.

Araújo questiona a impossibilidade de convivência entre a neurociência e a psicologia preconizada pelos Churchlands, e afirma que não há motivo justificável para que ambas não coexistam. Pelo contrário, podem perfeitamente evoluir em conjunto e proporcionarem o desenvolvimento mútuo.

Mas são os obstáculos metodológicos os maiores empecilhos ao materialismo eliminativo. Em primeiro lugar, a capacidade de mapeamento integral do cérebro está muito distante da realidade. Como argumenta João de Fernandes Teixeira (2012, p. 48), o cérebro tem complexidade neural tão alta, um número tão elevado de neurônios

e sinapses[12], que seria muito difícil, diante do cenário científico atual, reproduzir o cérebro em sua integralidade. Uma vez que a reprodução do cérebro se mostra um projeto tão longínquo, o mapeamento cerebral integral, pressuposto do materialismo eliminativo, é uma suposição remota.

Finalmente, quando o neurocientista se utiliza da neuroimagem, precisa que o paciente se utilize dos termos da psicologia popular para explicar o que está acontecendo. O paciente deve dizer ao experimentador se está feliz, assustado, com raiva etc. Não fazer uso dos termos da psicologia popular representaria um retorno ao behaviorismo, o qual, como visto na seção 2.2, não admite a utilização de descrições subjetivas.

Por outro lado, as pessoas que têm seus cérebros examinados não são capazes de identificar os sinais luminosos dos exames com o que sentiram ou pensaram naquele momento. O medo que eu sinto não é o mesmo que o sinal luminoso que se acende quando eu o sinto, ainda que ambos estejam correlacionados. Isso mostra a interdependência existente entre neurociência e termos da psicologia popular. Ainda que se possa estabelecer uma correlação entre o estado subjetivo e o estado cerebral, isso não é o mesmo que eliminar os estados subjetivos (ARAÚJO, 2011, p. 59).

Os autores Max R. Bennett e Peter M. S. Hacker apresentam outra perspectiva para a relação entre neurociência e filosofia. Bennett é neurocientista e professor de fisiologia na Universidade de Sidney, enquanto Hacker é filósofo e vinculado ao St. John's College, Oxford. Juntos eles escreveram a obra *Fundamentos filosóficos da neurociência*, onde analisam problemas conceituais no âmbito da neurociência cognitiva, entendida como a parte da neurociência que explica "as condições neurais que tornam possíveis as funções perceptivas, cognitivas, cogitativas, afetivas e volitivas" (BENNETT & HACKER, 2005, p. 15).

Bennett & Hacker entendem que é tarefa da neurociência estabelecer empiricamente o funcionamento das estruturas e operações neurais (2005, p. 15). A filosofia tem atribuição diversa, de esclarecer:

> [...] questões conceptuais (respeitantes, por exemplo, aos conceitos de mente ou memória, pensamento ou imaginação), a descrição das relações lógicas entre conceitos (tais como entre os conceitos de percepção e sensação, ou os conceitos de consciência e autoconsciência) e o exame das relações

[12] O número total de neurônios é algo próximo a cem bilhões, e cada neurônio tem, em média, sete mil conexões sinápticas como os outros (TEIXEIRA, 2012, p. 48).

> estruturais entre diferentes campos conceptuais (tais como entre o psicológico e o neural, ou o mental e o comportamental) (BENNETT & HACKER, 2005, p. 16).

É necessário distinguir um erro de observação de um erro teórico. Enquanto o erro de observação leva à falsidade da conclusão, o erro teórico leva à falta de sentido da teoria proposta. A verdade e a falsidade científicas não são o mesmo que a falta de sentido para a filosofia. Dessa forma, caberia à filosofia utilizar-se de conceitos adequados para fornecer sentido aos estudos neurocientíficos, sem apreciar a veracidade empírica dos resultados (BENNETT & HACKER, 2003, p. 20).

Assim como Bennett & Hacker, entende-se neste estudo que os estudos neurocientíficos têm importância para a filosofia, mas cabe à filosofia estabelecer questões conceituais, lógicas e relacionais. A filosofia pode se servir dos estudos neurocientíficos, ao mesmo tempo em que cabe a ela refletir sobre os limites da investigação neurocientífica.

A neurociência, portanto, expandiu a temática de investigação filosófica. Neurocientistas passaram a tentar explicar estados mentais utilizando-se dos estudos do cérebro. Eventuais confusões conceituais ou a vinculação a teorias eliminativistas não esvaziam os resultados obtidos pelos estudos neurocientíficos, pois é inegável haver descobertas da neurociência que podem ter importantes implicações para a filosofia. É o caso do estudo que será apresentado a seguir, do neurocientista António Damásio, que desenvolve uma teoria para tentar comprovar que a emoção é parte integrante do processo de decisão social.

Feitas essas considerações preliminares a respeito da relação entre filosofia e neurociência, a seguir será apresentada a *hipótese do marcador-somático*, desenvolvida pelo neurocientista António Damásio, que considera a emoção um elemento indissociável do processo de decisão social.

4.2 A HIPÓTESE DO MARCADOR-SOMÁTICO

Estudando pacientes sob seus cuidados, o neurocientista António Damásio desenvolveu uma hipótese segundo a qual as emoções ocupam papel fundamental no processo decisório humano.

A hipótese formulada por Damásio foi exposta na obra *O erro de Descartes,* e tem por base as consequências observadas em pacientes que sofreram danos

cerebrais severos, a maioria deles causados por uma lesão no córtex pré-frontal. Os efeitos nefastos dos danos ao córtex pré-frontal foram descritos pela medicina pela primeira vez no emblemático caso do inglês Phineas Gage (DAMÁSIO, 1996, p. 23).

Gage era um trabalhador da construção civil, envolvido na construção de uma estrada de ferro. Considerado pelos superiores um empregado eficiente e dedicado, possuía uma atribuição de grande importância e dificuldade: era encarregado de preparar as detonações na rocha necessárias para a posterior passagem dos trilhos da estrada de ferro.

Durante o trabalho de detonação, Gage introduzia a pólvora em uma barra de ferro e, em seguida, a cobria com areia. Com a pólvora coberta, Gage iria pressioná-la dentro da barra de ferro, compactando-a. Após, colocaria a barra num orifício feito na rocha e acenderia o rastilho para a detonação.

Contudo, no verão de 1848, durante uma detonação, Phineas Gage sofreu um grave acidente. Por equívoco, a areia não foi colocada sobre a pólvora, de modo que ele pressionou a pólvora diretamente dentro da barra de ferro. Como resultado, uma pequena faísca foi provocada, causando a explosão da pólvora. A explosão lançou a barra de ferro diretamente para o rosto de Gage.

A barra adentrou pela face esquerda de Gage e saiu pelo topo da cabeça, levando consigo parte do cérebro dele. Apesar disso, algo incrível aconteceu: mesmo com o cérebro parcialmente arrancado, foi levado ao hospital consciente, falando e andando normalmente. Não houve qualquer dano motor ou mental aparentes. Apesar da gravidade da lesão, Phineas Gage sobreviveu, e a sobrevivência dele revelou que aquela parte do cérebro destruída não era, como inicialmente se chegou a pensar, inútil.

A lesão sofrida por Phineas Gage teve uma consequência reveladora: Gage teve uma abrupta mudança de personalidade (DAMÁSIO, 1996, p. 27). Antes educado e responsável, agora usava palavras obscenas, era rude com os colegas e sem respeito pelos superiores. Nunca mais conseguiu levar adiante um plano para ações futuras. Estava claro que ele não mais era capaz de seguir as normas sociais, era como se houvesse se transformado em outra pessoa. Como nunca mais conseguiu se fixar em outro emprego, Gage terminou seus dias sendo sustentado por familiares, vindo a falecer em 1861, após um ataque epilético.

A história de Phineas Gage é marcante na compreensão do funcionamento do cérebro humano. Antes do acidente, muito já se havia dito a respeito da importância

do cérebro em funções motoras ou sensoriais. Contudo, o acidente produziu uma lesão cerebral que em nada alterou a fala, a audição ou qualquer função corporal da vítima. Houve, sim, uma profunda mudança na personalidade de Gage.

Damásio tinha sob seus cuidados pacientes com lesões semelhantes às de Phineas Gage, e observou que havia uma recorrente mudança de personalidade após a lesão, e que a mudança estava relacionada com a capacidade desses pacientes de tomar decisões adequadas.

Os vários pacientes com danos no córtex pré-frontal que estavam aos cuidados de Damásio eram constantemente submetidos a testes psicológicos. Um desses testes tinha como objetivo a solução de problemas sociais hipotéticos, para descobrir se o paciente era capaz de fornecer as opções de ação adequadas para a situação. Era comum que os pacientes fossem capazes de responder adequadamente a tais testes, o que significa que eles entendem o que se espera em uma determinada situação social. Contudo, numa situação real, o paciente se mostra perplexo e anormalmente indeciso diante de qualquer mínima escolha que tenha que fazer (DAMÁSIO, 1996, p. 75).

Damásio cita um caso específico ocorrido com um de seus pacientes, para ilustrar o nível de indecisão a que está se referindo. Segundo o autor, quando solicitado a um de seus pacientes que escolhesse entre duas datas possíveis para a próxima consulta, a atitude do sujeito deixou os investigadores perplexos:

> Durante quase meia hora, o doente enumerou razões a favor e contra cada uma das datas: compromissos anteriormente assumidos, proximidade de outros compromissos, possíveis condições meteorológicas, praticamente tudo o que se pudesse imaginar a respeito de uma simples data. [...] acabamos por lhe dizer calmamente que deveria vir na segunda das duas datas alternativas. Sua resposta foi da mesma forma calma e pronta. Limitou-se a dizer: “Está bem”. (DAMÁSIO, 1996, p. 226).

Esse exemplo mostra que o paciente, embora pareça agir normalmente, não é capaz de se decidir em questões que para a maioria das pessoas são de menor importância.

Damásio elaborou uma hipótese para explicar por que os pacientes com danos no córtex pré-frontal apresentam um processo decisório diferente das pessoas com funcionamento cerebral normal, e a denominou hipótese do marcador-somático.

Para desenvolver a hipótese, Damásio apresenta as modalidades de decisões que entende inerentes aos seres humanos. Este assunto já foi objeto de análise na seção 2.1, e por esse motivo será retomado neste momento de forma breve.

A decisão social é aquela que envolve o ambiente social, ou seja, que afeta tanto a nossa vida quanto a dos demais.

Decidir significa escolher agir em determinada direção quando há mais de uma ação possível. Mas nem todas as decisões têm a mesma complexidade. Algumas são simples, como as escolhas relacionadas aos apetites (DAMÁSIO, 1996, p. 198). Os apetites, que englobam a fome e a sede, pressupõem um mecanismo corporal que estimula um impulso para atender as necessidades. Atos reflexos, como a decisão de fechar os olhos ao perceber um objeto se aproximando, são instantâneos e também não são dotados de complexidade (DAMÁSIO, 1996, p. 199). Nas decisões relacionadas a atos reflexos e apetites o uso do raciocínio é de menor importância.

A faculdade de raciocinar está presente em dois grupos de decisões, mais complexas que as já apresentadas (DAMÁSIO, 1996, p. 199). O primeiro inclui situações como a resolução de um problema matemático, a elaboração de uma planta para a construção de um edifício.

O segundo envolve decisões como a escolha de uma carreira, de um candidato ou de um parceiro amoroso. Constata-se que a diferença fundamental entre os dois subgrupos apresentados pode ser resumida no envolvimento das decisões com o ambiente social. São essas as decisões objeto de análise nesta dissertação.

Assim como esta dissertação tem como objetivo estudar o subgrupo da decisão social, Damásio desenvolve a hipótese do marcador-somático para tentar explicar esse subgrupo de decisões.

Suponhamos que um indivíduo precise fazer uma escolha que envolva o ambiente social, como mudar ou não de emprego. Segundo Damásio (1996, p. 202), diversas imagens mentais se apresentariam na consciência dele. Em cada uma dessas imagens ele vislumbraria uma consequência para a decisão a ser tomada. Por exemplo, uma das cenas poderia apresentar uma situação favorável, onde ele encontra um emprego melhor e se desenvolve. Outra poderia ser uma cena pessimista, na qual se vê desempregado e sem dinheiro. Inúmeras imagens mentais podem lhe aparecer, e essa sucessão de imagens fará com que o indivíduo tome a decisão que lhe pareça mais adequada.

Como se pode observar, a decisão é tomada com base num conhecimento prévio do indivíduo, não surge a partir de um vazio mental. Entretanto, isso não responde à questão: como a decisão é tomada? Segundo Damásio, há duas possibilidades. A primeira é por meio da “razão nobre”. A segunda é a “hipótese do marcador-somático” (1996, p. 202).

Ao se utilizar da expressão “razão nobre”, Damásio quer aludir à filosofia clássica. Ou seja, uma vez que o indivíduo possui diante de si inúmeras opções, ele as submete a um processo racional de eliminação. Com base em ponderações objetivas, o sujeito elimina as decisões que se apresentam como piores, até chegar àquela decisão que possui menos aspectos negativos e mais aspectos positivos.

Entretanto, efetuando uma análise mais profunda, essa forma de decidir não é tão simples quanto parece. Voltemos ao exemplo inicial, da mudança de emprego. O indivíduo que precisa decidir pode visualizar diversos cenários futuros. Há um incontável número de opções que se colocam diante dele no momento da decisão, e muitos dos cenários imaginados são aleatórios, pois não é possível afirmar com segurança o que irá ocorrer no futuro. As opções, portanto, dependem da constante criação de cenários imaginários, cada vez mais distantes da realidade atual.

Uma decisão exclusivamente embasada na racionalidade, por conseguinte, não vai funcionar em situações como essa. Se o uso da “razão pura” significa decidir com base nos prós e contras de cada opção, a razão não nos dá as ferramentas necessárias para a decisão. É muito comum que as situações apresentem vantagens e desvantagens relativamente equilibradas, cujas consequências futuras não são aferíveis de imediato. É, portanto, imprescindível que a pessoa se sirva de outros mecanismos que a permitam encerrar a especulação e decidir.

Apesar da insuficiência de uma decisão puramente racional, o fato é que as decisões sociais fazem parte da vida de todos os seres humanos. Embora haja indivíduos indecisos, e o grau de indecisão varie de pessoa para pessoa, todos decidem inúmeras vezes durante a vida. Qual mecanismo estaria por trás dessas escolhas?

Para responder a essa pergunta, António Damásio desenvolve a hipótese do marcador-somático. A hipótese parte do pressuposto de que o estado somático, que nada mais é que o estado do corpo (soma, em grego, significa corpo), influencia no momento da tomada de uma decisão social. Ela une, portanto, a necessidade de decidir às sensações corporais que ocorrem no instante da decisão.

A hipótese de Damásio é que, antes da análise objetiva dos prós e contras de uma decisão, ocorre uma alteração no estado somático do sujeito. Ao visualizar um determinado cenário mental, é possível que a pessoa sinta um forte mal-estar, ou, nas palavras do autor, uma “sensação visceral desagradável” (DAMÁSIO, 1996, p. 205). Esse mal-estar pode ser tão intenso a ponto de paralisar a ação do indivíduo. O marcador somático funciona como um sinal automático, que dessa maneira:

> protege-o de prejuízos futuros, sem mais hesitações, e permite-lhe depois *escolher entre um número menor de alternativas*. A análise custos/benefícios e a capacidade dedutiva adequada ainda têm o seu lugar, mas só *depois* de esse processo automático reduzir drasticamente o número de opções (DAMÁSIO, 1996, p. 205).

Assim, a pessoa “marca” essa imagem ao estado somático desagradável. Por isso a denominação “marcador-somático”: a sensação corporal, seja positiva ou negativa, imprime no sujeito uma forte apreciação ou rejeição da opção a ela associada (DAMÁSIO, 1996, p. 205).

O marcador-somático tornaria a decisão mais fácil para o indivíduo. Caso se mostre desagradável, pode funcionar como um sinal de “pare”, indicando o perigo de uma determinada escolha. Já um marcador-somático positivo pode servir como elemento motivador, capaz de guiar a pessoa numa sequência de decisões futuras. O papel da razão seria exercido apenas depois do sujeito ter eliminado uma gama de opções por meio do marcador-somático, de modo que as opções racionais se desenvolveriam dentro do que é “sentido” como melhor pelo indivíduo.

Os marcadores-somáticos, contudo não tomam decisões por nós (DAMÁSIO, 1996, p. 206). Como a razão ainda desempenha um papel no momento da decisão final, e como é sempre possível decidir ainda que de maneira contrária às emoções despertadas pelas imagens mentais, as decisões ainda são tomadas racionalmente pelos indivíduos. O que os marcadores propiciam é uma predisposição, uma tendência a decidir de uma ou outra forma. Nada impede que o indivíduo decida da maneira que ele considere a correta, ainda que tal decisão desencadeie um estado somático desconfortável.

Diante disso, a hipótese não priva os indivíduos da responsabilidade pelas decisões pessoais. Os marcadores-somáticos apenas limitam a ampla gama de alternativas que uma pessoa possui antes de tomar determinada decisão social. O marcador-somático possibilita ao sujeito escolher dentro de um número razoável de

opções, sem ter que eliminar conscientemente uma infinidade de ações possíveis.

Um marcador-somático é positivo ou negativo de acordo com a emoção que ele desperta. Dessa forma, o elemento emocional teria um papel essencial na decisão social.

Damásio, portanto, conclui que o fator emocional é indissociável do processo de decisão social. Mas o estudo de Damásio não é o único experimento neurocientífico que conclui pela relevância do fator emocional. A importância das emoções também foi constatada quando neurocientistas analisaram o cérebro de pacientes submetidos a dilemas morais, como veremos adiante.

Mas antes de apresentar os estudos neurocientíficos dos dilemas morais, serão objeto de análise as críticas à teoria de António Damásio. Embora tenha tido o mérito de incluir as emoções no processo de decisão social, a hipótese do marcador-somático também foi alvo de críticas, dentre as quais as formuladas por Bennett & Hacker e Pedro M. S. Alves[13], que serão expostas a seguir.

4.2.1 As críticas a António Damásio

Bennet & Hacker (2005) e Pedro M. S. Alves (1996), por meio de argumentos diferentes, criticaram os fundamentos sob os quais António Damásio construiu a hipótese do marcador-somático.

Bennet & Hacker, na já mencionada obra *Fundamentos Filosóficos da Neurociência*, afirmam que, quando os neurocientistas formulam suas teorias, muitas vezes ignoram o que já foi discutido em filosofia. Isso leva à confusão conceitual na formulação da teoria, que pode comprometer os próprios resultados apresentados (BENNETT & HACKER, 2005, p. 20). Esse é o elemento central da crítica de Bennet & Hacker à teoria de Damásio.

A emoção é um elemento fundamental na hipótese do marcador-somático. Caso Damásio tenha se equivocado ao definir o que é emoção, pode ter se equivocado também nas conclusões que apresenta. Bennet & Hacker alegam que Damásio apresentou um conceito errôneo de emoção, o que prejudicou a construção da hipótese do marcador-somático (BENNETT & HACKER, 2005, p. 233).

---

[13] Pedro M. S. Alves é Doutor em Filosofia e Professor da Faculdade de Letras da Universidade de Lisboa.

Bennet & Hacker dedicam parte da obra *Fundamentos Filosóficos da Neurociência* para definir a emoção. Para eles, o sentimento é gênero do qual emoção é espécie. O sentimento, que representa tudo aquilo que podemos sentir, engloba:

a) as sensações: têm localização corporal definida, e podem nos informar sobre o estado do nosso corpo;
b) a percepção tátil: proporciona sensações capazes de detectar características do ambiente;
c) os apetites: representam uma união entre sensação e desejo. Exemplos são a fome, a sede e o desejo sexual. Apresentam sensações com localização específica: a fome se localiza no estômago, é impossível que uma sensação em outro lugar do corpo seja considerada fome. São formas de desconforto que nos predispõem a agir, no caso, a satisfazer o apetite. Apresentam um tipo amplo de desejo: ao faminto, qualquer comida lhe saciará a fome. Possuem padrão de ocorrência, saciedade e recorrência;
d) os vícios: são apetites induzidos, não naturais (BENNETT & HACKER, 2005, p. 220);
e) afetos: subdividem-se em:

- emoções: apresentam sensações que não têm localização específica. São diretamente relacionadas a objetos particulares: por exemplo, alguém pode ter medo de cobras, mas não de todos os animais. Por não apresentarem o mesmo componente fisiológico dos apetites, as emoções não têm o mesmo padrão de ocorrência, saciedade e recorrência. Finalmente, as emoções têm uma dimensão cognitiva diferente dos apetites, já que são influenciadas por crenças. (BENNETT & HACKER, 2005, p. 221);

- agitações: são perturbações afetivas momentâneas, geralmente causadas por um evento inesperado. Ocorrem quando nos sentimos temporariamente assustados, revoltados ou espantados, por exemplo. Não induzem à ação como as emoções, mas inibem temporariamente a ação motivada. Podemos nos comportar de determinada maneira quando estamos excitados ou chocados, mas não agimos devido à excitação ou ao choque no sentido em que agimos devido ao amor ou à gratidão (BENNETT & HACKER, 2005, p. 222). Não ocasiona uma ação, mas sim uma reação: um grito de horror, uma paralisia devido a um choque;

- disposições: consistem nos estados de espírito em que nos encontramos em grande parte do dia, durante as horas de vigília, quando nos sentimos felizes, melancólicos, deprimidos ou irritáveis. As disposições "não dão motivos para a ação, mas são exibidas na maneira como fazemos tudo o que fazemos" (Bennet & Hacker, 2005, p. 222). Não são uma emoção, mas podem induzir a uma tendência a apresentar certa emoção.

Todas as espécies de sentimentos se influenciam mutuamente. Como exemplo, podemos citar a influência das emoções sobre os apetites, que ocorre quando algumas questões, como de ordem religiosa, impedem que as pessoas se alimentem com determinadas comidas. A partir daí, a comida passa a ter uma dimensão cognitiva que antes não possuía. Isso não significa que a dimensão cognitiva assumida retire do alimento o componente fisiológico capaz de saciar a fome, mas sim que a ingestão de um alimento considerado proibido pode fazer emergir emoções como a culpa e a tristeza.

Bennet & Hacker não negam que "as fronteiras entre a emoção, a agitação e a disposição não estão bem definidas*"* (2005, p. 223). A interconexão entre as três é inegável: uma agitação pode dar origem a uma emoção, uma emoção pode dar origem a uma disposição, várias combinações podem ser feitas. A diversidade conceitual dos afetos decorre, em parte, da complexidade prática de tais distinções.

Comparando os conceitos apresentados por Bennet & Hacker com os de António Damásio, observa-se que as disposições são bastante semelhantes ao que Damásio caracterizou como *sentimentos de fundo*. Para Damásio, os sentimentos de fundo não se confundem com as emoções, e têm as seguintes características:

> [...] não são nem demasiado positivos nem demasiado negativos, ainda que se possam revelar agradáveis ou desagradáveis. Muito provavelmente, são esses sentimentos, e não os emocionais, que ocorrem com mais frequência ao longo da vida. Apenas nos damos conta sutilmente de um sentimento de fundo, mas estamos conscientes dele o suficiente para sermos capazes de dizer de imediato qual é sua qualidade. Um sentimento de fundo não é o que sentimos ao extravasarmos de alegria ou desanimarmos com um amor perdido; os dois exemplos correspondem a estados do corpo emocionais. Ao contrário, ele corresponde aos estados do corpo que ocorrem *entre* emoções. Quando sentimos felicidade, cólera ou outra emoção, o sentimento de fundo é suplantado por um sentimento emocional. O sentimento de fundo é a imagem da paisagem do corpo quando essa não se encontra agitada pela emoção. (DAMÁSIO, 1996, p. 181)

As disposições, portanto, estão conosco de maneira mais permanente do que as emoções. O surgimento da emoção ou da agitação sobrepõe o estado de

disposição do momento. Na ausência desses estados, voltamos a sentir alguma disposição.

Bennet & Hacker afirmam que o problema no conceito de emoção elaborado por Damásio está na ênfase dada ao mecanismo neural subjacente, como se vê na citação abaixo:

> Vejo a *essência* da emoção como a coleção de mudanças no estado do corpo que são induzidas numa infinidade de órgãos por meio das terminações das células nervosas sob o controle de um sistema cerebral dedicado, o qual responde ao conteúdo dos pensamentos relativos a uma determinada entidade ou acontecimento (DAMÁSIO, 1996, p. 168).

Damásio entende as emoções como as mudanças somáticas induzidas pelas células nervosas, e denomina sentimento a capacidade de sentir a resposta emocional (DAMÁSIO, 1996, p. 169).

Bennet & Hacker criticam a dicotomia proposta por Damásio entre emoção e sentimento. Para eles, conceituar emoção como as mudanças somáticas causadas pelo pensamento significaria dizer que, para aprender o significado de palavras emocionais, e, por consequência, aprender seu modo de uso, teríamos que aprender os nomes das complexas mudanças corporais que ocorrem (BENNET & HACKER, 2005, p. 234). Para eles não é isso que ocorre: quando se aprende sobre o medo, não se aprende sobre nomes de sensações corporais, mas sim sobre objetos perigosos ou ameaçadores.

Não haveria, portanto, fundamento em dividir a emoção da capacidade de senti-la: ter uma emoção é sentir a emoção. Além disso, é possível sentir uma emoção e não ter reações corporais correspondentes durante todo o tempo. Alguém pode, por exemplo, amar o próprio filho, mas não sentirá alteração no ritmo cardíaco toda vez que olhar para ele. Tal característica é ainda mais acentuada em emoções como orgulho e admiração. (BENNET & HACKER, 2005, p. 236).

Isso não significa que não exista uma relação entre certas emoções e mudanças somáticas, mas sim que não se deve identificar as mudanças somáticas com a emoção em si (BENNET & HACKER, 2005, p. 235).

Bennet & Hacker também criticam o conceito de imagens mentais proposto por Damásio. Ao tentar explicar como nos lembramos de algo, Damásio propõe que evocamos imagens mentais aproximadas do que experimentamos anteriormente. As imagens mentais seriam tentativas imprecisas de replicar padrões que já foram experienciados, e normalmente são retidas na consciência de forma passageira. O

surgimento das imagens mentais faria disparar os marcadores somáticos do corpo, influenciando, assim, a decisão a ser tomada.

Bennet & Hacker discordam da existência de imagens mentais, e argumentam que é incorreto “pressupor que percepcionar um objeto ou percepcionar que as coisas são assim-e-assim envolve imagens de qualquer coisa” (2005, p. 235). Nós não pensamos em imagens ao pensarmos em alguma coisa, da mesma maneira que, quando dizemos algo não falamos antes, para nós mesmos e mentalmente, o que vamos dizer. Dessa forma, nem a percepção nem o pensamento envolvem, necessariamente, imagens mentais, de modo que Damásio erra ao afirmar que as agitações emocionais são sempre causadas por imagens mentais.

O que nos faz sentir a emoção não é um pensamento ou uma imagem mental, mas a circunstância ou o objeto da emoção. Portanto, também seria um equívoco de Damásio pressupor que as mudanças somáticas referentes à emoção são causadas pelo pensamento (BENNET & HACKER, 2005, p. 234).

As emoções não são a respeito de alterações somáticas, mas sim a respeito do objeto que as causa. Em outras palavras, se alguém está amedrontado, o objeto do medo é o objeto do mundo exterior, real ou imaginário, que o apavora. As mudanças somáticas apenas ocorrem porque aquele objeto é entendido pela pessoa como assustador o suficiente para aterrorizá-la.

Diante de tais argumentos, Bennet & Hacker concluem que a hipótese do marcador-somático é incorreta (2005, p. 237), pois Damásio se equivoca ao dizer que emoções são imagens somáticas que nos dizem o que é bom ou ruim.

Isso não significa que Bennet & Hacker entendam que a as emoções não interferem nas decisões sociais. No entanto, eles atribuem a associação entre emoções e racionalidade ao interesse em perseguir determinados objetivos, como se pode ver no trecho a seguir:

> Embora Damásio possa estar perfeitamente certo ao associar a capacidade de racionalidade no raciocínio prático e na prossecução de objetivos com a aptidão de sentir emoções, a ligação reside numa característica comum subjacente aos dois. [...] é implausível pressupor que o que está errado com os pacientes que sofreram lesões no sector ventromediano do córtex pré-frontal é que as suas reações somáticas sejam inadequadas ou desinformadoras para eles [...]. Mas o que deve ser investigado é se a lesão cerebral no tipo de pacientes que Damásio estudou afeta a capacidade de se interessarem ou de continuarem a interessar-se por finalidades e objetivos. Porque essa deficiência tanto afetaria as emoções dos pacientes como suas aptidões para prosseguir objetivos ao longo do tempo (BENNET & HACKER, 2005, p. 237).

Portanto, Bennet & Hacker relacionam as emoções à tomada de decisão social, mas não nos termos propostos por Damásio. Para eles, não é correto afirmar que os marcadores somáticos dos pacientes com lesão no córtex pré-frontal não lhes fornecem informações adequadas, mas sim que os pacientes com lesão cerebral perdem a capacidade de perseguir objetivos de longo prazo.

Pedro M. S. Alves, no artigo *Que Verdade no Erro de Descartes*, elaborou outra crítica à obra de António Damásio. A crítica de Pedro Alves tem como fundamento principal o pensamento de René Descartes, e como Damásio interpretou o dualismo cartesiano[14].

Alves inicia seu argumento expondo como há pouco espaço na atualidade para qualquer autor defender algo semelhante à dualidade de substâncias na forma conceituada por Descartes (ALVES, 1996, p. 171). Apesar disso, a ciência cognitiva discutiu o dualismo cartesiano, como assevera Howard Gardner:

---

[14] O dualismo cartesiano é o nome dado à teoria, criada por René Descartes, que defende que mente e corpo possuem substâncias distintas.
Descartes conclui que a mente e o corpo têm substâncias diferentes ao buscar definir fundamentos seguros para o saber. Examinando os fundamentos do saber tradicional, inicia pela experiência sensível, e conclui que ela não pode ser fundamento do saber, pois nossos sentidos às vezes nos enganam (DESCARTES, 2009, p. 58). Outro possível fundamento seria o poder discursivo da razão, ou o pensamento. Contudo, Descartes conclui que mesmo o pensamento pode nos enganar, pois não há garantias de que os pensamentos que nos ocorram sejam mais verdadeiros do que as ilusões dos sonhos (2009, p. 58).
Depois de colocar tudo em dúvida, Descartes finalmente encontra o que entende ser o novo fundamento do saber. Se tudo o que o homem pensa pode ser falso, algo há que existe necessariamente: o sujeito pensante. É a partir de tal argumento que Descartes concluirá que o corpo e a alma possuem substancias distintas, com propriedades incompatíveis.
Descartes toma a proposição *Penso, logo existo*, como basilar e indiscutível. Isso porque, ainda que duvidemos de tudo, não há como duvidar que estamos duvidando. Como dúvidas são pensamentos, não podemos duvidar que pensamos enquanto estamos duvidando. Por conseguinte, não há como *pensar que não pensamos* (TEIXEIRA, 2011, p. 29). *Penso, logo existo* é uma proposição sobre a qual não cabe nenhuma refutação.
Ao concluir que era possível que ele mesmo não existisse e que a única coisa que lhe garantia a existência era o pensar, Descartes identificou o ser humano a uma substância cuja natureza é o pensamento, e que não depende de nada material, inclusive do corpo, para existir.
Descartes então conclui que o homem é um ser dual. O primeiro argumento que usa para defender essa conclusão é que a mente, sendo mais fácil de ser conhecida, deve ser diferente do corpo (DESCARTES, 2010, p. 150).
Ainda na obra *Meditações*, Descartes apresenta outro argumento como comprovação da dualidade do homem: as substâncias materiais, incluindo nossos corpos, são divisíveis, enquanto o mental é indivisível, de modo que o espírito não pode ser concebido a não ser como coisa única e inteira (DESCARTES, 2010, p. 200).
Esse dualismo de substâncias levou à necessidade de se responder à seguinte pergunta: sendo corpo e alma tão diferentes, como é possível que a alma interaja causalmente com o corpo? Na obra *Paixões da alma,* Descartes argumenta que a alma movimenta o corpo por intermédio da única ligação que com ele possui: a glândula pineal (DESCARTES, 2010, p. 313).

> Em suas discussões das ideias e da mente, da experiência sensorial e do corpo, do poder da linguagem e da importância de um ser que organiza e duvida, Descartes formulou uma agenda que dominaria as discussões filosóficas e influenciaria a ciência experimental nas décadas e séculos seguintes. Além disso, ele propôs a imagem vívida e controversa da mente como um instrumento racional que, contudo, não pode ser simulado por qualquer máquina imaginável – uma imagem que ainda hoje é debatida na ciência cognitiva (GARDNER, 2003, p. 66)

Alves argumenta que apenas após uma leitura das obras de ciência natural de Descartes é possível compreender os reais motivos que o levaram a formular o conceito segundo o qual corpo e mente possuiriam substâncias distintas e inconfundíveis. Na obra *Tratado do Homem*, Descartes apresenta explicações físicas detalhadas a respeito do funcionamento de todo o corpo humano. Para Descartes, os processos orgânicos podem ser integralmente explicados apenas por mecanismos físico-naturais (ALVES, 1996, p. 172).

Por outro lado, embora Descartes explique tudo o que é observável no homem baseado em nada mais do que uma causalidade natural, ele mesmo não se convence de que este homem orgânico seja o homem verdadeiro. Um homem que não possuísse nada além da causalidade natural não passaria de um autômato, e Descartes entende que o homens e autômatos não se confundem[15].

Os animais respondem de maneira automática às situações da vida, enquanto os seres humanos respondem com criatividade. Descartes não entendia possível que o conjunto dos órgãos físicos fosse capaz de proporcionar aos humanos uma gama tão ampla de respostas às diferentes situações da vida. A solução que encontrou foi estabelecer a existência de uma alma imaterial, que forneceria a racionalidade necessária ao comportamento humano, nos tornando, assim, essencialmente

---

[15] Descartes dedicou parte de sua filosofia à compreensão dos autômatos, que são seres semelhantes aos humanos, mas desprovidos de alma, que é uma prerrogativa do ser humano. Descartes entende que, como os animais não possuem alma, morrem juntamente com o corpo. Já os seres humanos possuem uma alma imortal, que não é destruída com a destruição do corpo (DESCARTES, 2009, p. 99).

A ausência da alma imortal faz com que os autômatos não tenham pensamento. Assim sendo, os autômatos são incapazes de desenvolver processos de raciocínio e não possuem estados mentais subjetivos.

Descartes acreditava que, embora humanos e animais possuam estrutura fisiológica semelhante, existem duas diferenças essenciais que impedem a confusão entre o homem e o autômato. A primeira é a capacidade de usar palavras, ou outros sinais, para expressar os próprios pensamentos (DESCARTES, 2009, p. 96). Não é impossível que uma máquina ou um animal, como um papagaio, profira palavras. Contudo, tanto a máquina quanto o animal não são capazes de compreender o sentido das palavras que proferem.

A segunda diferença é o conhecimento, a capacidade de agir com base na razão. Ainda que os autômatos sejam capazes de fazer algumas coisas tão bem quanto, ou até melhor que nós, o fato de não agirem com acerto em tantas outras prova justamente que não têm espírito, e sim “é a natureza que neles opera de acordo com a disposição de seus órgãos” (DESCARTES, 2009, p. 98).

distintos dos autômatos. Isso possibilitou a ele explicar como o ser humano é capaz de agir sem se fixar a padrões estereotipados de comportamento (ALVES, 1996, p. 176).

Assim, o autômato só se faria igual ao homem se tivesse *consciência* daquilo que faz e das circunstâncias em que age; em última análise, se tivesse consciência de si (ALVES, 1996, p. 177). Para o autômato, não existe consciência, não existe um centro de comando que dirige a vida em determinada direção, segundo os ditames da racionalidade.

O cerne do problema cartesiano, que fez Descartes desenvolver a teoria do dualismo, seria, portanto, a compreensão da consciência. Contudo, apesar de Damásio ter dado ao seu livro o título de *O erro de Descartes,* expressa que a obra "não é sobre a consciência" (DAMÁSIO, 1996, p. 267).

Embora os estudos neurocientíficos tenham avançado na compreensão do cérebro humano, a explicação do funcionamento da consciência ainda é um desafio. Alves conclui sua argumentação expondo o seguinte a respeito do tema:

> Perante esse limite óbvio da abordagem neurológica, há apenas duas saídas possíveis. Há a saída confiante, optimista, a que considera que, no fundo, o que não sabemos responder hoje sabê-lo-emos amanhã, e *sem mudar* o próprio modelo de abordagem utilizado. E há a saída cartesiana, que nos diz apenas: procurai noutra direcção, e a partir de outros princípios (ALVES, 1996, p. 178).

Descartes procurou a resposta para a consciência em outra direção, e formulou a teoria dualista da mente.

Uma vez que o pensamento de Damásio se alinha à neurofilosofia, analisada na seção 4.1, ele acredita que as respostas serão integralmente fornecidas com o avanço da ciência. Contudo, essa conclusão não é tão evidente: é possível que a explicação detalhada de todos os mecanismos neurológicos ainda não seja suficiente para estabelecer uma ponte real entre a matéria e a experiência consciente.

Ao intitular a própria obra como *O erro de Descartes,* António Damásio talvez tenha atribuído a Descartes um erro que, se bem analisado, é parte de um enigma que nem a neurociência nem a filosofia foram ainda capazes de solucionar.

Percebe-se que as críticas apresentadas refutam aspectos pontuais da hipótese do marcador-somático, sem excluir a importância das emoções na decisão social. Portanto, há elementos indicando que existe relação entre decisão social e

emoção, elementos esses que são reforçados por meio das conclusões dos estudos neurocientíficos apresentados na próxima seção.

4.3 A INFLUÊNCIA DAS EMOÇÕES NAS DECISÕES DE DILEMAS MORAIS

O dilema moral é uma situação na qual não existe uma dimensão que se imponha com legitimidade moral (LA TAILLE, 2007, p. 80). Em um dilema moral, dois princípios morais entram em conflito, e não há uma ação possível capaz de preservar ambos, de modo que um dos princípios deverá ser sacrificado em prol do outro.

Os neurocientistas têm se utilizado de dilemas morais com o escopo de compreender o processo decisório humano, examinando o cérebro do indivíduo no momento em que é exposto a um dilema moral. O estudo com dilemas morais tem apresentado resultados que corroboram a importância das emoções no processo decisório.

Dentre os dilemas morais existentes, há um bastante conhecido na atualidade, sobretudo após ter sido discutido pelo filósofo Michael Sandel na obra *Justiça – O Que é Fazer a Coisa Certa* (2012, p. 30). Imagine a seguinte situação: um trem desgovernado está indo em direção a um local onde há cinco trabalhadores. Caso nenhuma providência seja tomada, os cinco morrerão. Contudo, diante de você está uma alavanca que permite mudar a direção do trem e fazê-lo correr por outros trilhos. Ocorre que, nos outros trilhos, há um trabalhador que certamente morrerá. A pergunta é: nessas circunstâncias, você moveria a alavanca, alterando o trajeto do trem e poupando a vida dos cinco trabalhadores em detrimento da vida de um?[16] Ao fazer essa pergunta a inúmeros entrevistados, a maioria respondeu que sim: quase todos moveriam a alavanca para salvar os cinco, ainda que um fosse sacrificado.

Mas, suponhamos uma segunda situação, similar à primeira[17]. Você está diante dos trilhos do trem e observa que há cinco trabalhadores mais adiante, fazendo alguns ajustes. Repentinamente, surge um vagão do trem sem controle. Se você não fizer nada, os cinco trabalhadores serão mortos. Mas, bem diante de você, há um sujeito bastante corpulento. Você percebe que, se o jogasse nos trilhos do trem, o vagão pararia, poupando a vida dos cinco trabalhadores, embora matasse o sujeito à sua frente. Nesse caso, você empurraria o sujeito nos trilhos do trem, salvando os

[16] Esse dilema é conhecido, em inglês, como *footbridge dilemma.*
[17] A variação do dilema recebe a denominação de *trolley dilemma* em inglês.

trabalhadores, ou se omitiria? A resposta da grande maioria foi não: poucos entrevistados seriam capazes de empurrar alguém para a frente do vagão, ainda que essa decisão poupasse a vida de mais pessoas.

Embora as situações apresentadas sejam semelhantes, o fato é que as respostas fornecidas pelos entrevistados são distintas. É possível argumentar que uma decisão exclusivamente racional não faria distinção nos casos apresentados, uma vez que, no tocante ao número de mortos, as duas situações são idênticas. [18] Sendo assim, o que cada um dos casos possui para ensejar reações diferentes?

Michael Sandel apresenta várias hipóteses para tentar explicar a diferença de comportamento verificada. Por exemplo, a diferença poderia ocorrer em virtude da *intenção* (SANDEL, 2012, p. 32). O sujeito, ao mover uma alavanca, não tem a intenção de matar um terceiro, enquanto tal intenção estaria presente ao empurrar alguém nos trilhos. No entanto, o próprio Sandel considera o problema em tal argumento: no empurrão, a intenção também não era a morte de um, mas sim salvar a vida dos outros cinco. Ele argumenta, ainda, que a diferença poderia residir na *sensação de obrigarmos alguém a agir contra a própria vontade*, o que ocorreria ao empurrarmos o indivíduo (2012, p. 31). Mais uma vez, ele próprio refuta tal raciocínio, já que o sujeito que estava apenas trabalhando e teve o trem para ele redirecionado provavelmente também não estaria disposto a dar a vida pelos outros funcionários.

A respeito da perplexidade causada em decorrência da verificação de resultados tão diferentes para situações similares, Sandel diz o seguinte:

> Não é fácil explicar a diferença moral entre esses casos – por que desviar o bonde parece certo mas empurrar o homem da ponte parece errado.

---

[18] A concepção de que é preferível salvar o maior número possível de pessoas se adequa aos postulados da *ética utilitarista*.
O utilitarismo é uma corrente filosófica desenvolvida por pensadores ingleses, e apresenta como foco principal o estudo dos problemas éticos, embora também aborde questões lógicas. A filosofia utilitarista foi fundada por Jeremy Bentham e desenvolvida por John Stuart Mill, cuja obra *Utilitarismo* é emblemática na tradição de tal corrente, sendo hoje considerada um texto clássico e de extrema importância para a filosofia moral (GALVÃO, 2005, p. 10).
A ética utilitarista preconiza que a finalidade das ações humanas é o prazer. Consequentemente, é considerado bom aquilo que é útil e capaz de proporcionar prazer aos homens. Que disto não se conclua tratar-se de uma ética egoísta. Ao contrário: a ética utilitária tem por fundamento a utilidade, ou o Princípio da Maior Felicidade, como fundamento da moralidade. Segundo John Stuart Mill, "as ações são corretas na medida em que proporcionem a felicidade de um maior número de pessoas, e erradas ao produzirem o reverso da felicidade" (MILL, 2005, p. 48).
Logo, o princípio básico da ética utilitarista é que o agir humano deve ter como diretriz o bem-estar do maior número de pessoas. Caso as circunstâncias sejam tais que impossibilitem um aumento da felicidade, ao menos se deve tentar minimizar os inconvenientes gerados pela ação necessária. Trata-se de uma ética bem-intencionada, que preconiza a maximização do prazer e a minimização da dor e do sofrimento.

> Entretanto, note a pressão que sentimos para chegar a uma distinção convincente entre eles – e se não pudermos, para reconsiderar nosso julgamento sobre a coisa a fazer em cada caso (2012, pp. 32-33).

Como observado por Sandel, a diferença no padrão observado gera uma pressão para compreendermos o resultado de tal disparidade. Com o intuito de alcançar tal compreensão, Joshua Greene realizou uma pesquisa que forneceu uma nova perspectiva a respeito do dilema do trem desgovernado.

O filósofo e psicólogo evolutivo Joshua Greene e sua equipe analisaram os cérebros dos indivíduos no momento em que são submetidos aos dilemas morais apresentados, bem como outros semelhantes (GREENE et al., 2001, p. 2105), e constataram uma diferença na atividade cerebral quando o sujeito é submetido a um ou outro tipo de situação.

Greene apresentou aos entrevistados 60 dilemas para serem resolvidos. Os dilemas foram divididos em “morais” e “não-morais”. Os morais, por sua vez, subdividiam-se em “morais impessoais” e “morais pessoais”. Os dilemas morais impessoais são equivalentes ao primeiro caso do trem desgovernado – o caso da alavanca. Já os morais pessoais são equivalentes à segunda situação, quando o sujeito deveria ser empurrado para os trilhos (GREENE et al., 2001, p. 2106).

O estudo concluiu que na maioria dos casos em que um sujeito se vê diante de um dilema moral pessoal, há fortes estímulos no córtex pré-frontal medial, no cingular posterior e na amígdala, áreas do cérebro conhecidas por processarem emoções. Por outro lado, quando o dilema proposto é moral impessoal, os estímulos ocorrem em partes do cérebro relacionadas às funções cognitivas, como atenção e memória de curto prazo (GREENE et al., 2001, p. 2107).

Assim, a conclusão de Joshua Greene é no sentido de que a diferença entre os resultados dos dilemas morais apresentados reside no *engajamento emocional* do indivíduo. Segundo ele:

> Em cada uma das áreas do cérebro identificadas em ambos os experimentos 1 e 2, a condição moral-pessoal teve um efeito significativamente diferente tanto das condições morais-impessoais quanto das não-morais. [...] Os dados comportamentais fornecem evidências adicionais para o aumento da participação emocional na condição moral-pessoal, revelando um padrão de tempo de reação que é exclusivo para essa condição e que foi previsto por nossa hipótese sobre a interferência emocional (GREENE et al, 2001, p. 2107).

Em outras palavras, na primeira situação, quando a atitude solicitada é apenas mover uma alavanca, a decisão não possui forte conteúdo emocional, pois o entrevistado não se sente diretamente responsável pela morte daquele que vem a ser sacrificado. Na segunda hipótese, o fato de empurrar diretamente o sujeito para os trilhos do trem acarreta uma sensação de mal-estar que funciona como um freio para a ação, e é incapaz de superar o desejo de salvar os cinco trabalhadores.

E quanto a pessoas com lesões no córtex pré-frontal, como Phineas Gage? Se tais pessoas realmente possuem, conforme afirmou Damásio, um processamento emocional diferente dos indivíduos com funcionamento cerebral normal, que interfere na tomada de decisão, reagiriam eles de forma diferente quando submetidas a dilemas morais?

De acordo com outro experimento, realizado por Michael Koenigs e sua equipe, a resposta é sim: pessoas com danos no córtex pré-frontal apresentam uma tendência a decidir de forma utilitária, sempre sacrificando um em detrimento da maioria, independente de se tratar de dilemas morais pessoais ou impessoais (KOENIGS et al., 2007, p. 909).

Esse resultado corrobora a hipótese de que as emoções estão intrinsecamente ligadas às decisões que envolvem o ambiente social do indivíduo. Desprovidos da barreira emocional verificada em pessoas com funcionamento cerebral normal, pacientes com danos no córtex pré-frontal decidem os dilemas morais pessoais movidos apenas por fatores matemáticos.

Mas que disso não se conclua que os pacientes com danos no córtex pré-frontal apresentam um completo embotamento emocional. No mesmo estudo de Koenigs et al. (2007, p. 910), concluiu-se ser possível que esses pacientes apenas possuam uma diminuição no processamento das emoções sociais, sendo estatisticamente mais vingativos e sensíveis à injustiça. Essa conclusão decorreu da comparação dos resultados obtidos por pessoas com e sem danos no córtex pré-frontal no “jogo do ultimato”.

O jogo do ultimato consiste, primeiramente, em dar uma quantidade de dinheiro a um jogador *A. A* deve dividir o dinheiro com o jogador *B*. Mas *A* não precisa dividir o dinheiro igualmente: suponhamos que *A* receba 100, ele pode entregar a *B* apenas 1, ficando com 99 para si. Contudo, *B* tem o poder de rejeitar a oferta. Se *B* rejeita a oferta, tanto *A* quanto *B* ficam sem nada. Diante da possibilidade de rejeição de *B, A* deve entregar uma quantia razoável para o outro jogador, sob pena de ambos ficarem

sem dinheiro algum. Caso *B* se sinta ultrajado com a proposta de *A*, certamente rejeitará a oferta.

Foram submetidas ao jogo do ultimado pessoas com funcionamento cerebral normal e outras com danos no córtex pré-frontal. O resultado foi que as pessoas sem qualquer lesão cerebral se mostraram propensas a aceitar uma divisão desigual do dinheiro, desde que não muito distante da metade. Já os pacientes com lesões no córtex pré-frontal tendem a rejeitar com mais frequência propostas minimamente desiguais (KOENIGS et al., 2007, p. 910).

O resultado levou os pesquisadores a concluir que as pessoas com danos no córtex pré-frontal não apresentam um embotamento geral das emoções, pois, se isso fosse verdade, elas não se sentiriam tão insultadas pela divisão desigual do dinheiro. Na verdade, ocorreu justamente o contrário: as pessoas com as referidas lesões cerebrais foram mais sensíveis a se sentirem ultrajadas, indicando que são menos capazes de controlar a frustração e tendentes a manifestar raiva exagerada (KOENIGS et al., 2007, p. 910).

A partir desse resultado, Jorge Moll e Ricardo de Oliveira-Souza (2007, p. 320) formularam a hipótese de que indivíduos com lesões no córtex pré-frontal não possuem uma diminuição generalizada das emoções, mas apenas das emoções sociais. Essas emoções, que incluem a culpa e a empatia, estariam fortemente relacionadas à parte interna do lobo frontal. Já o córtex pré-frontal dorsolateral e o córtex orbitofrontal seriam importantes para emoções autocentradas e aversivas, por exemplo raiva e frustração.

Essa hipótese explicaria porque os indivíduos que apresentam as lesões são "'mais racionais' quando julgam um dilema moral 'pessoal' mas mais emocionais no jogo do ultimato" (MOLL & OLIVEIRA-SOUZA, 2007, p. 320). Portanto, não seria o caso de um embotamento emocional de tais indivíduos, mas sim de uma redução da capacidade de possuir emoções sociais.

Percebe-se que os resultados das pesquisas neurocientíficas apresentadas apontam no sentido de que o fator emocional é indissociável do processo decisório. Sendo assim, a ciência cognitiva, ao eleger como característica metodológica a exclusão das emoções dos, limitou a própria capacidade de explicar a decisão social.

Embora a neurociência tenha apresentado estudos empíricos inovadores, capazes de fornecer explicações para processos mentais por meio do estudo do cérebro, por outro lado, como foi exposto no início desta seção, os neurocientistas, ao

adentrarem na seara filosófica, o fizeram a partir de uma perspectiva materialista, acarretando todos os problemas já mencionados. Além disso, as experiências neurocientíficas não podem deixar de ser analisadas com parcimônia, uma vez que são realizadas numa pequena parcela da população, e, diversas vezes, os estudos têm por base o funcionamento cerebral de indivíduos com problemas neurológicos, como foi o caso da hipótese do marcador-somático.

A explicação da neurociência apresenta, portanto, algumas limitações. A próxima seção tem como objetivo proporcionar mais argumentos que apontem para a relevância das emoções no processo decisório, e, com isso, refutar a capacidade da ciência cognitiva em explicar a decisão social, em razão da característica metodológica de excluir as emoções de seus estudos. Na seção seguinte será apresentada uma teoria segundo a qual as emoções ocupam um relevante papel na decisão social: a teoria do altruísmo recíproco.

5 A TEORIA DO ALTRUÍSMO RECÍPROCO E O PROCESSO DE DECISÃO SOCIAL

Na seção anterior, com o fim de analisar se o pressuposto metodológico da ciência cognitiva de excluir a emoção prejudica a explicação da decisão social, viu-se como os estudos empíricos do cérebro realizados pela neurociência indicam que a decisão social está atrelada à emoção. Nesta seção será apresentada a teoria do altruísmo recíproco, desenvolvida por Robert Trivers, que se funda no processo de seleção natural para justificar a presença das emoções no processo de decisão social.

A seguir, para localizar historicamente a teoria do altruísmo recíproco, serão tecidas algumas considerações sobre o caminho que teoria da seleção natural percorreu para sair do campo da biologia e ingressar no âmbito das ciências humanas.

5.1 A TEORIA DA SELEÇÃO NATURAL E AS CIÊNCIAS HUMANAS

A teoria da seleção natural ficou conhecida a partir dos estudos do naturalista britânico Charles Darwin. Mas ele não foi o primeiro a propor uma teoria que abarcasse a evolução biológica. Antes de Darwin publicar a obra *A origem das espécies* em 1859, pensadores do século XVIII já haviam defendido a existência da evolução na biologia, dentre eles o Conde de Buffon e Erasmus Darwin, avô de Charles Darwin (MAYR, 1998, p. 371 e 381). Contudo, faltava sistematicidade às ideias, de modo que nenhum deles conseguiu desenvolver uma teoria da evolução.

O primeiro autor a sistematizar uma teoria da evolução foi Jean-Baptiste de Lamarck, na obra *Pholosophie zoologique*, de 1809. Lamarck afirmou ter desenvolvido a própria teoria para tentar explicar dois fenômenos biológicos: a perfeição crescente dos animais e a grande diversidade dos organismos. Lamarck entendia por perfeição crescente “o aumento gradual em ‘animalidade’, dos animais mais simples aos que possuíam a mais complexa organização, culminando com o homem” (MAYR, 1998, p. 387).

Lamarck defendeu a transformação lenta e gradual das espécies, ocorrida ao longo das gerações. A capacidade de transformação era decorrente do uso e desuso das estruturas, de maneira que quando uma estrutura se fazia necessária era gerada espontaneamente, e, caso não fosse mais relevante para o organismo, desaparecia com o passar das gerações (MAYR, 1998, p. 398).

A teoria da seleção natural desenvolvida por Charles Darwin teve o mérito de apreciar cientificamente os mecanismos evolutivos. Na obra *A origem das espécies,* Darwin afirma que ocorrem mutações em animais e plantas com o passar das gerações, e que as mutações podem ser positivas ou negativas para a adaptação no ambiente em que vivem. Quando a mutação é adaptativa ao ambiente, o indivíduo tem maior probabilidade de sobreviver e transmitir seus genes. Por outro lado, quando a variação é prejudicial, o indivíduo tende a ser destruído. Segundo o autor, "a esta conservação das diferenças e variações individualmente favoráveis e a destruição das que são prejudiciais a chamei eu *seleção natural*" (DARWIN, 2009, p. 78, grifos no original).

Darwin não defendeu o progresso no processo evolutivo, como fez Lamarck, e também abordou com cautela os seres humanos em seus escritos, evitando falar de evolução social. Apesar disso, a ideia de progresso esteve inicialmente presente em autores que tentaram transpor a teoria biológica darwinista às ciências humanas.

Herbert Spencer foi um influente divulgador da noção de progresso social em termos de seleção natural. O pensamento de Spencer obteve notoriedade e aprovação na segunda metade do século XIX, mas foi duramente criticado no início do século XX. Richard Hofstadter, na obra *Social Darwinism in American Thought*, originalmente publicada em 1944, popularizou o termo "darwinismo social", cujas ideias centrais atribuiu sobretudo a Spencer. O darwinismo social seria a tentativa de aplicar o darwinismo nas sociedades humanas, de modo que a luta pela existência resultaria na sobrevivência dos mais aptos. Graças à divulgação do pensamento de Hofstadter, o darwinismo social ficou associado à defesa de atos de eugenia, racismo e imperialismo. Dessa forma, até a primeira metade do século XX, principalmente após a Segunda Guerra Mundial, houve grande resistência à aplicação da teoria da seleção natural em ciências humanas.

O autor Yves Christen afirma que o darwinismo social ainda é bastante repudiado por causa das supostas consequências morais que acarretaria. Apesar disso, Christen entende que "nenhum argumento científico importante veio contradizer a teoria do darwinismo social" (1981, p. 144), o que não impede que continue rejeitado.

A resistência em utilizar conceitos evolutivos em ciências humanas começou a ser superada na década de 1960, com a publicação de artigos dos biólogos evolucionistas William D. Hamilton (1964) e George C. Willians (1966). Os artigos de Hamilton e Willian representaram o início das tentativas de se entender o

comportamento social humano por meio dos conceitos de seleção em relação ao indivíduo.

Na esteira da abordagem evolutiva, Robert Trivers desenvolveu, em 1971, a teoria do altruísmo recíproco, que será analisada adiante e constitui um importante embasamento teórico ao presente estudo. A teoria do altruísmo recíproco explica como é possível que seres não aparentados atuem em benefício uns dos outros. Auxiliar outro ser da mesma espécie é uma decisão social, e Trivers explica o processo de decisão social por meio da evolução e da presença de emoções.

Um reflexo da influência da teoria da seleção natural nas ciências humanas foi o surgimento da psicologia evolucionista. Enquanto a psicologia tradicional dava ênfase nas relações familiares e culturais, a psicologia evolucionista introduz o fator evolutivo na tentativa de compreensão do funcionamento mental humano.

A união da revolução cognitiva das décadas de 1950 e 1960 com a da biologia evolucionista de 1970 e 1980 foi denominada pelo antropólogo John Tobby e pela psicóloga Leda Cosmides como “psicologia evolucionista”. A ciência cognitiva auxiliaria a compreender *como* é possível termos esse determinado tipo de mente, enquanto a biologia evolucionista auxiliaria a entender *por que* possuímos esse tipo específico de mente (PINKER, 1998, p. 34).

Na obra *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, os autores Cosmides, Tooby & Barkow dizem que a psicologia evolucionista é a psicologia informada pelo fato de que a arquitetura inerente à mente humana é produto do processo evolucionário (1992, p. 7).

Para formular e argumentar pela validade de uma hipótese evolucionista, é importante uma abordagem em níveis. O primeiro nível é o da teoria evolucionista geral, ou seja, da evolução pela seleção natural, visto a partir de uma perspectiva genética por autores mais recentes, como no caso de Hamilton (1964), Willians (1966) e Trivers (1971).

Em seguida, existe um nível intermediário, que compreende teorias que procuram explicar todo um domínio de funcionamento. É o caso da teoria do altruísmo recíproco, de Robert Trivers, que busca explicar todo o comportamento altruísta para não aparentados.

O terceiro nível abrange hipóteses evolucionárias específicas, ou seja, se a teoria se provar verdadeira, quais os efeitos que ela teria na evolução. É o nível das predições de comportamentos derivados das hipóteses. Se encontrarmos

comportamentos não excepcionais que contrariam integralmente a hipótese, ela se mostrará falsa (BUSS, 2008, p. 43).

A introdução da teoria do altruísmo recíproco nesta dissertação se justifica em virtude de que, embora a perspectiva neurocientífica apresente fortes argumentos no sentido da importância das emoções no processo decisório, a neurociência pode apresentar os problemas do eliminativismo, como exposto na seção anterior. Além disso, a neurociência apresenta uma explicação empírica do funcionamento das emoções no processo decisório, muitas vezes apresentando resultados que têm por base a análise de casos isolados, o que poderia levar ao questionamento da validade dos estudos na população em geral. Por fim, enquanto os estudos neurocientíficos dizem que *a emoção compõe* a decisão social, o altruísmo recíproco apresenta argumentos que explicam *a razão pela qual* as emoções estão presentes no processo de decisão social.

Entender a relação entre emoções e decisão é adequado a partir de uma perspectiva evolucionista, pois as emoções são adaptações (PINKER, 1998, p. 390). O enfoque evolucionário é capaz de explicar muitas expressões emocionais consideradas problemáticas.

A adaptabilidade do sistema do altruísmo recíproco depende de uma elevada capacidade de detectar e punir um traidor (entendido como aquele que, tendo sido beneficiado uma vez, não retribui quando tem a oportunidade), bem como de estimar a relação de custo-benefício de um ato altruísta, para evitar retribuições desproporcionais. Por isso as emoções que auxiliam nessas escolhas foram selecionadas evolutivamente e participam da decisão social. Para o sucesso de um sistema cooperativo como o altruísmo recíproco, é fundamental a presença de uma grande variedade de emoções que exerçam influência nas decisões sociais.

Expostas essas considerações iniciais, a seguir será apresentada a teoria do altruísmo recíproco, de Robert Trivers, que apresenta importantes conclusões a respeito dos motivos pelos quais a emoção pode ter influência no processo decisório.

5.2 O ALTRUÍSMO RECÍPROCO

Robert Trivers desenvolveu a teoria do altruísmo recíproco tendo por base as premissas da teoria da seleção natural, ou seja, de que os comportamentos que hoje existem foram selecionados ao longo do tempo porque se mostraram adaptativos.

Com a teoria, ele procurou explicar o motivo pelo qual os animais, incluindo nós, seres humanos, auxiliamos indivíduos não aparentados.

O elemento emocional desempenha importante papel nessa teoria, sobretudo as chamadas “emoções sociais” – aquelas que, conforme visto na seção 4.2, indivíduos com lesões no córtex pré-frontal não desenvolvem adequadamente. Por isso, com o intuito de reforçar o argumento desenvolvido até esse momento, a respeito da insuficiência da ciência cognitiva para explicar a decisão social em virtude da relevância das emoções nesse processo, na presente seção serão tecidas algumas considerações sobre a teoria do altruísmo recíproco, e como ela preconiza o entrelaçamento entre emoções e decisão social.

O comportamento altruísta, segundo Trivers, pode ser definido como o comportamento que beneficia outro organismo, não aparentado, e que parece prejudicial ao organismo que atua beneficiando o outro (TRIVERS, 1971, p. 35). A questão que se coloca é: por qual motivo um organismo age em prejuízo próprio, com o intuito de favorecer outro, não aparentado e até mesmo de espécie diferente? Partindo do pressuposto de que tal comportamento deve ter alguma utilidade, já que foi evolutivamente selecionado, qual seria ela?

Trivers não entende como altruísta, para os fins da teoria por ele desenvolvida, o comportamento que beneficia parentes próximos, sobretudo filhos. Agir em prol de parentes próximos é o que se denomina seleção por parentesco – em inglês, *kin selection*[19]. A seleção por parentesco é facilmente explicada pela intenção de perpetuar os próprios genes. É possível concluir que a seleção por parentesco teve importante papel na evolução de certos comportamentos altruístas, sobretudo se considerarmos as tribos caçadoras-coletoras, nas quais havia grandes chances de que todos se considerassem parentes (1971, p. 35).

Somente um comportamento dirigido em benefício de um ente não aparentado, com algum custo para o benfeitor, pode ser considerado uma ação altruísta. Nesse

---

[19] Hamilton (1964) propôs a teoria da seleção familiar ou da adaptação inclusiva (*inclusive fitness*). Embora Hamilton não tenha usado o termo *kin selection*, a terminologia se consolidou e hoje conceitua a seleção decorrente do grau de parentesco.
Hamilton ponderou que o sucesso na transmissão genética não estaria restrito à prole. Graças à divisão da carga genética com parentes próximos, alguém agindo em benefício de um familiar estaria aumentando o sucesso da transmissão dos próprios genes.
O comportamento altruísta, embora pudesse diminuir a adaptação de um indivíduo em particular, aumentaria a adaptação de seus parentes, melhorando a difusão genética. Isso levaria a uma proliferação do gene altruísta, elevando a probabilidade de comportamentos altruístas em prol de parentes nas gerações futuras.

ponto, encontramos o primeiro elemento importante para a eficácia de uma ação altruísta: a ponderação dos custos. Os benefícios da reciprocidade dependem dos custos, pois é preciso que haja menos custos para o altruísta do que ganhos para o beneficiário. Entenda-se custo como o grau de atraso na reprodução genética do altruísta e benefício o grau de aumento da possibilidade de reprodução genética do favorecido (TRIVERS, 1971, p. 36).

Mas o verdadeiro elemento de eficácia na ação altruísta é a reciprocidade. Um comportamento altruísta depende da expectativa de reciprocidade no futuro. Se o beneficiário não exprimir reciprocidade – ou seja, trair – o altruísta fica no prejuízo.

O papel de destaque que Trivers dá à reciprocidade é criticado por Moore (1984), que não concorda com a inserção da reciprocidade na definição de altruísmo. De acordo com Moore, o termo altruísta não corresponde a uma ação que tem por objetivo a aquisição de ganhos recíprocos, de modo que a teoria de Trivers não representaria propriamente o altruísmo.

Trivers (1971, p. 36) aponta três maneiras distintas pelas quais os altruístas podem distribuir seus benefícios: aleatoriamente, para aparentados, ou para não aparentados.

a) aleatoriamente: a distribuição aleatória de benefícios altruístas não traz benefícios, já que dela não se espera reciprocidade nem favorece o parentesco. O altruísta fica com todo o custo e toda a vantagem vai para o beneficiário;
b) para aparentados: agir altruisticamente em prol de aparentados é a *seleção por parentesco*, sobre a qual já foram tecidas algumas considerações. A *kin selection* beneficia a transmissão genética do altruísta, pois aquele que recebe o auxílio divide material genético com aqueles que o prestam;
c) para não aparentados: o comportamento altruísta não é aleatório e é exercido em benefício de não aparentados. É para esse tipo de comportamento que Trivers elaborou a teoria do altruísmo recíproco;

Para que casos de altruísmo para não aparentados tenham sucesso, ou seja, apresentem vantagem tanto para os altruístas quanto para os beneficiários, é importante que se verifiquem algumas circunstâncias (TRIVERS, 1971, p. 37). Dentre elas, podemos ressaltar:

a) uma vida longa, na qual haja chances de que dois indivíduos se encontrem em várias situações altruístas;

b) baixa taxa de dispersão, de modo que as mesmas pessoas convivam muito tempo juntas;
c) dependência mútua, para que haja variação de situações beneficiário e altruísta;
d) cuidado parental, com dependência entre pais e filhos acentuada e prolongada no tempo;
e) hierarquia de dominância, onde haja relações assimétricas, com um dos membros dominando o outro. O subordinado pode agir em benefício do dominante, esperando que ele lhe beneficie no futuro. Ou o dominante pode ser benevolente como o subordinado, para manter a fidelidade em caso de contestação da própria autoridade;
f) combate, significando que, por mais dominante que seja um indivíduo, ele pode ser atingido em combate, derrubado por união dos mais fracos ou por outros motivos.

Não faltam exemplos de situações de altruísmo entre humanos: a ajuda em épocas de perigo, em face de acidentes, predadores; a partilha de comida; o auxílio aos doentes, feridos e idosos; o compartilhamento de objetos e de conhecimentos.

Contudo, para que o altruísta não seja prejudicado, é preciso que, caso haja oportunidade numa situação futura, aquele que tenha recebido auxílio no passado ajude o benfeitor. Quando as situações se inverterem, se o beneficiário não retribuir o favor, ele será considerado, nos termos utilizados por Trivers, um traidor (1971, p. 46). Entenda-se por traidor, aqui, simplesmente o sujeito que, tendo obtido alguma espécie de auxílio, não retribui quando tem a oportunidade.

Existem duas formas pelas quais a traição pode ser exercida (1971, p. 46). A primeira é a traição grosseira, quando o traidor não retribui em nada – o altruísta não tem qualquer benefício. Por ser facilmente verificável, haverá forte seleção para eliminar o traidor grosseiro.

A segunda é a traição sutil. Aqui, há reciprocidade, mas sempre uma parte doa menos do que a outra – se as posições se invertessem, o recebedor não doaria tanto. O altruísta se beneficia um pouco, mas o traidor sutil é quem mais se beneficia da interação.

Tendo em vista que uma relação altruísta pode durar muito tempo, inclusive por toda uma vida, com centenas de trocas se realizando, não há como estabelecer uma tabela que defina exatamente os custos e benefícios de cada um dos envolvidos. Mas

é possível concluir que aquele que mais doa fica numa situação desfavorável. Caso ele interrompa a relação, vai perder todo o investimento altruísta que fez no passado. Além disso, perderá o benefício que recebe atualmente, que, embora seja pouco, é melhor do que nada. No entanto, se continuar na relação, a tendência é que ela lhe permaneça desfavorável, com uma perda progressiva das doações por ele efetuadas.

É, assim, necessária uma análise minuciosa das ações realizadas para identificar a traição sutil. Justamente a dificuldade em identificá-la e, uma vez identificada, em pôr fim a essa relação, diante da situação desfavorável do altruísta, que faz com que a traição sutil seja adaptável (TRIVERS, 1971, p. 47).

Pode parecer então que trair seja a estratégia mais adaptativa, mas não foi essa a conclusão de Axelrod na obra *The Evolution of Cooperation*. Nessa obra, foi proposto o estudo do Dilema do Prisioneiro[20] para estabelecer qual a estratégia mais eficaz em relações que perduram no tempo: trair ou cooperar. A conclusão (1981, p. 175) foi que a estratégia vencedora possui duas regras básicas:

a) o indivíduo coopera no primeiro encontro;
b) nos encontros posteriores faz exatamente o que o outro oponente fez no primeiro.

Axelrod denominou essa estratégia de *tit for tat* (expressão que teria o sentido da que conhecemos como “olho por olho”). O fato da estratégia *tit for tat* ter se mostrado vencedora no dilema do prisioneiro indica que nem o altruísmo aleatório nem a traição constante são as melhores opções. Com isso vemos a importância de que o indivíduo possua mecanismos que o ajudem a decidir se vai cooperar ou trair e, uma vez traído, que o incentivem a punir o traidor[21].

---

[20] “No Dilema do Prisioneiro existem dois jogadores, e cada um possui duas escolhas, denominas cooperar ou trair. Cada jogador deve fazer a escolha sem saber a que o opositor fará. Não importa o que o outro faça, a traição proporciona um maior retorno do que a cooperação. O dilema é que se os dois traírem, se saem pior do que se ambos cooperarem. ” (HAMILTON E AXELROD, 1981, p. 7-8).

[21] Tanto a teoria do altruísmo recíproco quanto o estudo do dilema do prisioneiro têm servido de base para contestar a teoria econômica neoclássica, que preconiza, a partir dos ensinamentos de Adam Smith na obra *A Riqueza das Nações*, que o autointeresse é a base do comportamento humano voltado para as relações de mercado.
Muitos estudos têm buscado relacionar as premissas econômicas a teorias evolutivas, sobretudo teorias sociobiológicas que incluem o altruísmo como elemento central (caso da teoria do altruísmo recíproco). Nesse sentido existem os estudos de Simon (1973, 1983), Becker (1976), Nelson & Winter (1982) e Lucas (1987) (CAMPELL, 1987, p.171).
A lógica do comportamento altruísta, defendido por teorias biológicas como o altruísmo recíproco e a adaptação inclusiva, refutariam a ideia do *homem econômico*, que decide sempre por razões egoístas.

O resultado reforça a teoria do altruísmo recíproco. De acordo com Axerold (1981, p. 173), o comportamento altruísta pode ser seletivamente vantajoso para um organismo, caso exista uma interação que perdure no tempo e haja expectativa de retorno futuro. Os indivíduos que se relacionam continuamente são capazes de ajustar o comportamento de acordo com que os outros fizeram no passado, sendo essa a base do altruísmo recíproco.

Portanto, se houver reciprocidade, a cooperação é a estratégia mais adaptativa. Essa situação faz surgir uma necessária capacidade de identificar e punir o traidor, de modo que a seleção natural irá rapidamente proporcionar um complexo sistema psicológico em cada um que regule tanto as disposições altruístas quanto as de traição, e também que seja eficiente em detectar a traição do outro. Com a evolução, serão selecionados tanto os melhores traidores quanto os melhores detectores, já que eles terão mais sucesso em transmitir os próprios genes por meio da descendência. Há uma tendência para que os indivíduos cada vez mais se beneficiem de trocas altruísticas e se previnam da traição. Assim, as pessoas não se dividem em mais ou menos altruístas ou traidores, mas em melhores ou piores detectores de uma situação – se devem doar ou trair.

Dessa forma, um elemento fundamental para o funcionamento do sistema do altruísmo recíproco é uma aguçada capacidade de estimar a relação de custo-benefício de um ato altruísta. Sem essa capacidade, atos desproporcionais prejudicariam um dos envolvidos na relação – que, na prática, estaria doando muito mais do que recebendo.

A adaptabilidade do altruísmo recíproco depende de que os traidores não se sobressaiam, de modo que os altruístas devem saber quando agir e quando parar de atuar em benefício de terceiros. As emoções são mecanismos que auxiliam a pessoa a tomar decisões relacionadas ao sistema do altruísmo recíproco, e o papel que elas representam nesse sistema serão analisadas a seguir.

5.2.1 A função das emoções no sistema do altruísmo recíproco

As emoções seriam um mecanismo selecionado para auxiliar cada indivíduo a calcular o custo-benefício dos atos altruístas que pratica. Um ato altruísta não engloba apenas um grande feito, como salvar alguém em perigo, mas também as pequenas decisões sociais que uma pessoa toma diariamente. As emoções serviriam como um

alerta para que a pessoa aja altruisticamente ou pare com a ação altruísta, caso não ocorra reciprocidade.

O êxito do altruísmo recíproco depende da presença tanto de emoções positivas quanto negativas. Observam-se emoções positivam em situações de amizade e empatia, que incentivam o agir altruísta. Sem a tendência para gostar de outros que não sejam parentes, para formar amizades, o altruísmo recíproco não funciona bem.

A seleção natural pode ter favorecido tanto o auxílio de estranhos em situações de desespero, quanto a tendência de agir em benefício e de gostar daqueles que são altruístas. Trivers inclusive cita pesquisas (1971, p. 48) nas quais se conclui que indivíduos tendem a ser mais altruístas diante de pessoas que gostam, e também que o altruísmo costuma ser predominantemente exercido em prol de amigos. Emoções positivas fazem com que o altruísta se sinta bem em favorecer a outrem.

O ato altruísta desperta emoções positivas naquele que o pratica, e tais emoções servem de estímulo para que o benfeitor continue a praticar cada vez mais ações altruístas. Por outro lado, aquele que recebe o benefício sentiria *gratidão*. Como os humanos devem ter sido selecionados para serem sensíveis ao custo e benefício de um ato altruísta, tanto para decidir quando fazê-lo e quando retribuir, a gratidão seria uma das emoções que regula a medida da retribuição. Assim, quanto maior a necessidade do favorecido, maior a tendência dele à retribuição. Também, quanto mais escassos os recursos do altruísta, maior o sentimento de gratidão, e, consequentemente, maior a possibilidade de que o beneficiário retribua (TRIVERS, 1971, p. 49).

Mas é possível que surjam traidores adaptados justamente para tirar proveito das emoções positivas que a ação altruísta gera no benfeitor. Isso coloca o altruísta numa posição frágil, pois pode ser facilmente explorado. Dessa forma, é preciso outro mecanismo, oposto ao das emoções positivas, para que a reciprocidade possa ser cobrada e o altruísta não fique no prejuízo. Trivers dá a esse mecanismo o nome de *agressão moralista*, e afirma que “muito da agressão humana tem tons morais. Injustiça, deslealdade e falta de reciprocidade frequentemente motivam a agressão ou indignação humana” (1971, p. 49). Pode-se verificar que as emoções negativas também funcionam como um regulador da reciprocidade no altruísmo.

A agressão moralista seria a origem da indignação humana, do forte senso de justiça que muitos carregam dentro de si. Ela serve a diversos propósitos: em primeiro

lugar, põe um freio na tendência do altruísta de continuar agindo, movido pelos já mencionados bons sentimentos, ainda que na ausência de reciprocidade. Também há um fator educativo, uma vez que o traidor aprende que seu comportamento não será tolerado, que o altruísta não mais o auxiliará em uma situação futura. Finalmente, em casos extremos, elimina o traidor, e a possibilidade de que transmita seus genes, matando-o ou exilando-o, por exemplo. Isso explica, em parte, porque os seres humanos têm uma raiva tão proeminente daqueles que traem. Por exemplo, em antigas tribos africanas estudadas, pequenas faltas como comportamento preguiçoso e injustiça na divisão da comida podiam causar agressões aparentemente desproporcionais à ofensa cometida (TRIVERS, 1971, p. 49).

Outra emoção que teria sido desenvolvida para integrar o sistema do altruísmo recíproco teria sido a *culpa* (TRIVERS, 1971, p. 50). Caso alguém, numa relação recíproca, venha a trair, e tal fato seja descoberto pelo parceiro, ou haja grande chance de que venha a ser, e supondo que o parceiro responda rompendo o relacionamento com o traidor, o traidor pagará um alto preço pela sua conduta. É, portanto, vantajoso ao indivíduo não realizar o ato de traição, ou, fazendo-o, não o repetir. A seleção teria agido nos traidores de modo a despertar emoções de culpa, para que eles evitem trair no futuro, e que atuem de modo a reparar a traição um dia perpetrada. Assim, caso o traidor repare o dano causado e não venha a trair no futuro, volta a ser benéfico para o altruísta prosseguir na relação com aquele que traiu.

Na obra *Como a mente funciona,* Steven Pinker concorda com os fundamentos do sistema do altruísmo recíproco, e entende que as emoções são adaptações úteis ao processo de decisão social. Pinker vai ainda mais longe na defesa do papel das emoções negativas no sistema do altruísmo recíproco, e afirma que não há mal funcionamento de uma emoção quando ela domina alguém, fazendo-o agir de maneira inadequada ao convívio social. Pelo contrário: esse agir que, aos olhos dos demais, pode ser considerado irracional, seria um comportamento adaptativo, ajustado aos objetivos primordiais dos seres humanos (PINKER, 1998, p. 394).

Pinker, a partir dos fundamentos da teoria do altruísmo recíproco, argumenta que emoções como a ira e o desejo de vingança são excelentes defesas contra humanos egoístas. Somos, por essência, seres que cooperam. O surgimento de um egoísta prejudica o grupo, e é preciso uma maneira de impedir que proliferem. Por outro lado, a vingança nem sempre é positiva, pois pode implicar em custos muito superiores aos de aceitar a ofensa. É preciso que uma forte emoção obrigue o sujeito

a se vingar, mesmo que isso seja, aparentemente, desvantajoso. A longo prazo, sociedades com indivíduos vingativos obtiveram vantagem, pois venceram os egoístas e, em última análise, promoveram a cooperação.

O mesmo se diga de emoções como a culpa e a vergonha. Tais emoções são adaptações que obrigam o indivíduo a agir conforme o esperado. As emoções, portanto, não são vestígios do passado animal, mas sim uma ferramenta que assume o controle para nos obrigar a cumprir e cobrar promessas (PINKER, 1998, p. 425).

Como as emoções ajustam os objetivos, e o pensamento estabelece ferramentas adequadas para atingi-los, não parece haver uma linha divisória nítida entre o pensar e o sentir (PINKER, 1998, p. 394). Tampouco o pensar precede o sentir ou vice-versa. A emoção mobiliza o pensamento e o corpo para agir de acordo com a necessidade, tendo como rota de orientação o objetivo supremo.

O problema das emoções, que leva muitos a acreditar que trabalham contra nós, é terem sido projetadas para propagar nossos genes, “e não promover a felicidade, sabedoria ou valores morais” (PINKER, 1998, p. 390). Entender o meio e assegurar a cooperação dos demais foram os comportamentos que fizeram com que nossos ancestrais transmitissem os próprios genes da melhor maneira possível (1998, p. 393), além, é claro, de comportamentos comuns a todos os animais, tais como luta, fuga, sexo, alimentação. As emoções entram em cena como importante elemento para que os comportamentos de cooperação, que envolvem decisões sociais, fossem exercidos da melhor forma possível.

Com base nos argumentos de Pinker e Trivers, pode-se afirmar que não apenas as emoções consideradas positivas, tais como amizade e generosidade, representam um importante papel para os humanos. Emoções como raiva e desejo de vingança desempenham um papel tão importante quanto as demais. Sem a raiva e a vingança os traidores triunfariam, e a cooperação seria impossível.

Verifica-se, portanto, que existe uma ampla gama de emoções que têm função relevante no processo de decisão social, e que são necessárias para que um sistema cooperativo como o do altruísmo recíproco possa prosperar.

É evidente que nem todas as emoções podem ser consideradas adaptativas em nossa sociedade atual. Não vivemos mais em uma sociedade tribal, e mecanismos que uma vez foram adaptativos hoje podem ser considerados inadequados. A evolução tende a conservar os comportamentos, ainda que eles não se mostrem mais adequados ao ambiente. É possível, portanto, que isso tenha ocorrido com algumas

emoções. Como argumenta Jonathan Cohen, professor de psicologia e codiretor do Instituto de Neurociência da Universidade de Princeton, muitas das emoções que fazem parte de nossa composição biológica podem ter sido eficientes no passado, auxiliando a sobrevivência da nossa espécie, mas uma vez que as circunstâncias de vida dos seres humanos mudaram tão profundamente, é possível que existam emoções que não sejam mais mecanismos capazes de nos conduzir com eficiência em direção aos nossos maiores interesses (COHEN, 2005, p. 6).

Apesar disso, é vantajoso para o grupo que os traidores sejam malsucedidos. Seja nos relacionamentos individuais, seja na sociedade como um todo, a traição gera prejuízos, pois diminui cada vez mais o desejo de realização de atos altruístas.

Novak & Sigmund (2005) propõem a reputação como importante elemento contra os traidores. Em um cenário no qual o Dilema do Prisioneiro se repita sucessiva vezes, mas com parceiros diferentes, a retaliação pode ocorrer por meio de um indivíduo que não tenha mantido contato com o traidor, mas que apenas tenha sido informado que aquele jogador traiu anteriormente. Por isso, a construção da reputação é um fator de fortalecimento da cooperação (2005, p. 1292).

Trivers aponta o surgimento de um elemento para aperfeiçoar a traição sutil: a *mímica*. Uma vez que a amizade, a agressão moralista, a culpa e a gratidão são elementos do altruísmo recíproco, a seleção irá favorecer indivíduos que consigam mimetizar esses comportamentos; capazes, assim, de influenciar a ação alheia em proveito próprio. Uma pessoa pode, por exemplo, exprimir um comportamento indignado de agressão moralista, sem que qualquer traição tenha ocorrido. Com isso, ela irá obter mais do que proporcionou numa relação. O mesmo pode ocorrer com falsas demonstrações de generosidade ou amizade, desacompanhadas de qualquer emoção, apenas pensando no futuro recebimento de algum benefício. O traidor pode fingir estar arrasado pela culpa, só para, no futuro, poder trair novamente. Ou alguém pode aparentar estar em péssima situação, para, com isso, despertar a piedade dos outros, fazendo-os agir altruisticamente em seu benefício. Como esclarece Trivers, esse comportamento pode até nem ser consciente: é possível que a pessoa, em dado momento, se convença de que não voltará a trair, e manifeste a culpa; ou que o falso moralmente agressivo acredite sinceramente que a traição tenha ocorrido (1971, p. 50).

Por isso, é importante que os indivíduos sejam dotados de técnicas de detecção do traidor sutil. A seleção precisa ter favorecido a habilidade de detectar e eliminar traidores sutis.

É fácil explicar o motivo pelo qual as pessoas se sentem ofendidas quando são vítimas de uma agressão moralista injustificada, uma vez que o agressor externa um comportamento de hostilidade diante de alguém que acredita não ter feito nada de errado. No entanto, é mais complexo explicar por que as pessoas rejeitam demonstrações de culpa, amizade ou generosidade, quando desprovidos do elemento emocional correspondente. Atos que, quando genuínos, são acompanhados de uma emoção específica, restam vazios de significado quando destituídos desse componente. Sem a emoção correspondente, manifestações de amizade ou de culpa não passam de ações cínicas e premeditadas, com o objetivo de benefícios futuros, e as pessoas tendem a desprezar manifestações de amizade e culpa desacompanhadas da emoção que as caracteriza.

Na prática, embora seja possível detectar o traidor por meio da psicologia do altruísmo recíproco, é mais fácil e mais comum que essa detecção aconteça em virtude da identificação de um comportamento inconstante da parte do traidor. Por exemplo, caso o indivíduo, após aparente manifestação de intensa culpa, volte a trair, é bem possível que seja um traidor sutil e esteja se utilizando de um comportamento mimético de culpa apenas para continuar a trair naquela relação.

O tipo de decisão envolvido no sistema do altruísmo recíproco é bastante semelhante ao que António Damásio denominou de decisões que envolvem o ambiente social, ou decisões sociais (DAMÁSIO, 1996, p. 200). Conforme visto na seção 4.2, ao formular a hipótese do marcador-somático, Damásio afirmou que as emoções eram um componente fundamental nas decisões relacionadas a outras pessoas, como as que envolvem em quem votar, qual carreira seguir ou com quem se casar.

Como nesta dissertação é adotado o conceito de decisão social formulado por Damásio, entende-se que a teoria do altruísmo recíproco é aplicável ao processo de decisão social.

É possível estabelecer um paralelo entre a hipótese do marcador-somático e a teoria do altruísmo recíproco, no sentido de que o envolvimento emocional numa decisão social pode decorrer de uma tentativa de estimar os ganhos e custos futuros de determinada relação. Portanto, decisões sociais seriam, de fato, mais emocionais

do que outras, já que no envolvimento com outras pessoas está sempre em jogo o quanto dar e o quanto receber naquela relação.

Como visto, as emoções desempenham um papel central no sistema do altruísmo recíproco, influenciando tanto nas ações altruístas quanto nas retaliações. O elemento emocional funciona como a engrenagem reguladora da eficácia do altruísmo recíproco, garantindo que uma das partes não seja explorada pela outra.

5.3 A CIÊNCIA COGNITIVA E O PROCESSO DE DECISÃO SOCIAL

Na presente seção será realizada uma retomada da ciência cognitiva, diante do que foi exposto nas seções 3, 4 e 5 desta dissertação. O objetivo desse cotejo é analisar, com base no que foi apresentado nas referidas seções, há elementos capazes de colocar em dúvida a capacidade da ciência cognitiva de explicar o processo de decisão social.

A ciência cognitiva, analisada na seção 2.2, apresenta como crenças centrais o uso de representações e o computador como o modelo mais viável para o funcionamento da mente. Além disso, uma de suas características metodológicas é a exclusão da emoção no estudo do processamento cognitivo.

John Searle criticou o uso de representações e a analogia entre mente e computador, como visto na seção 3. Searle propôs a intencionalidade, entendida como a “propriedade de muitos estados e eventos mentais pela qual estes são dirigidos para, ou acerca de, objetos e estados de coisas no mundo” (SEARLE, 2002, p. 1), como elemento essencial na diferenciação entre os estados mentais e o funcionamento das máquinas. Para ele, ao contrário das mentes humanas, as máquinas não têm intencionalidade. Searle buscou explicar os estados mentais e o processo decisório por meio de uma abordagem naturalista, calcada na intencionalidade. Contudo, os críticos a Searle apresentaram argumentos que enfraqueceram o intuito de Searle de fornecer uma explicação verdadeiramente biológica, e colocaram em dúvida a existência de uma diferenciação tão acentuada na intencionalidade humana.

As críticas à teoria da intencionalidade de John Searle, analisadas na seção 3.3, colocam em dúvida a existência de uma dicotomia marcante entre intencionalidade original e derivada. Por esse motivo, este estudo buscou outro caminho para a análise da validade dos postulados da ciência cognitiva para explicar

a decisão social, e se voltou para a característica metodológica de exclusão das emoções.

Para entender o papel das emoções na decisão social, a seção 4 apresentou estudos neurocientíficos, em especial a hipótese do marcador-somático. Tanto essa hipótese quanto os estudos relacionados aos dilemas morais indicaram que a emoção faz parte da decisão social.

É importante ressaltar que, de acordo com as teorias neurocientíficas apresentadas, as emoções *fazem parte* do processo de decisão social, de modo que há dúvidas a respeito da capacidade de tal elemento poder ser controlado ou voluntariamente retirado do processo decisório. Os experimentos cerebrais de pessoas realizando escolhas dilemas morais mostram que, independentemente do desejo da pessoa de se envolver ou não emocionalmente com um tema, determinadas situações desencadeiam reações emocionais no cérebro, sendo questionável a capacidade de separar a emoção da decisão.

As conclusões da teoria do altruísmo recíproco, apresentadas na seção 5, também se coadunam com a presença das emoções no processo de decisão social. As emoções são elementos reguladores nas relações interpessoais, equilibrando a ação altruísta e a reciprocidade, bem como auxiliando na retaliação dos traidores.

É possível analisar uma decisão social sob o prisma das duas teorias mencionadas acima (hipótese do marcador-somático e altruísmo recíproco) para ponderar a respeito da validade dos argumentos por elas propostos. Veja-se um exemplo desenvolvido na obra *Rationality in action*, de John Searle: a eleição de um candidato. Esse é um típico exemplo de decisão envolvendo o ambiente social. Escolher em quem votar é uma decisão difícil, já que entra em cena um elemento futuro imprevisível, o comportamento do candidato depois de eleito.

É possível que, com o intuito de escolher o melhor candidato, o eleitor analise as propostas de cada um e faça um esforço sincero no sentido de escolher o candidato mais adequado, que beneficie tanto ele quanto a comunidade. Mas a ponderação dos fatores objetivos pode não ser suficiente para uma decisão, já que, como acontece na maioria dos casos, cada um possui aspectos positivos e negativos.

Portanto, uma eleição apresenta os seguintes elementos: a) a existência de mais de um candidato, o que significa mais de uma opção; b) dificuldade em tomar uma decisão apenas a partir da análise dos prós e contras de cada concorrente, ou seja, uma decisão exclusivamente racional.

Enquanto argumenta que não existem causas antecedentes suficientes para uma ação, e que uma razão só é efetiva porque nós a fazemos efetiva (SEARLE, 2001, p. 16), Searle coloca o seguinte:

> É por isso que, eventualmente, a explicação de suas ações e a justificação podem não coincidir. Suponhamos que lhe pedissem para justificar ter votado em Clinton; você pode fazer isso apelando para a administração superior dele na economia. Mas pode ser o caso que a verdadeira razão pela qual você tenha agido seja que ele foi para sua velha Universidade em Oxford, e você pensa, "Lealdade aos colegas da Universidade vem primeiro" (SEARLE, 2001, p. 16).

Constata-se, portanto, que a escolha não foi imparcial, e sim com base no fato de que o candidato frequentou a mesma faculdade que o eleitor. Caso alguém pergunte os motivos da escolha, dirá que lhe parecia o melhor candidato por causa do conhecimento em economia, mas sabe que o verdadeiro motivo foi que ambos estiveram na mesma faculdade.

É possível tecer uma analogia entre a situação apresentada por Searle com a argumentação desenvolvida por António Damásio, conforme analisado na seção 4.2 dessa dissertação. Em suma, Damásio afirma que, diante de uma situação em que se exija uma decisão, e que tal decisão envolva o ambiente social – como é o caso da decisão em quem votar – diversas "imagens" se apresentariam na consciência do indivíduo. Essas imagens representariam cenários para uma decisão futura. A razão, para ele, não forneceria os meios necessários para decidir, já que, pela razão, não é possível saber quais dos cenários imaginados se concretizaria. Assim, entraria em cena o mecanismo emocional, que faria com que o sujeito decidisse com base em sensações viscerais (DAMÁSIO, 1996, p. 205). Essa é a essência da hipótese do marcador-somático.

No exemplo proposto por Searle, a decisão foi tomada com base na lealdade perante os colegas de faculdade. Não se trata, portanto, de uma decisão calcada naquilo que Damásio denomina de "razão nobre" (DAMÁSIO, 1996, p. 202), que seria uma decisão fundamentada nos pontos objetivamente positivos e negativos de uma situação. Pensando no que se passou com o eleitor em tal situação, não há como descartar que tenha sentido uma sensação visceral agradável, utilizando, mais uma vez, os termos de Damásio, ao pensar no candidato que frequentou a mesma escola

que ele. Escolher um candidato pelo fato de ter estudado na mesma faculdade que o eleitor parece mais um exemplo de decisão emocional.

Há um outro paralelo possível de traçar na situação apresentada: a decisão do eleitor e a teoria do altruísmo recíproco. Trivers insiste no fato de que atos altruístas tendem a ser feitos mais em benefício de amigos do que de desconhecidos. Relações altruístas se protraem no tempo e são vantajosas a todos os envolvidos, desde que um deles não seja um traidor.

Como descartar que o eleitor tenha se perguntado, ainda que não explicitamente: com quem será mais fácil estabelecer um laço de amizade, com alguém que estudou na mesma faculdade que eu ou com um sujeito com quem não tenho nada em comum?

Uma vez eleito um candidato com quem o eleitor tem algo em comum, maiores são as possibilidades de que seja beneficiário de um ato altruísta dele. O candidato, agora eleito, pode ser grato ao apoio dado ao eleitor, e pode simpatizar com ele pelo fato de ambos terem frequentado a mesma faculdade. Constata-se, portanto, que entram em cena algumas emoções importantes do altruísmo recíproco: simpatia, amizade, gratidão.

Nesse sentido, é possível ponderar que a evolução selecionou indivíduos que decidiam de forma emocional, já que essas decisões se mostraram benéficas com o passar do tempo. Decidir em prol de um candidato pelo fato de ele ter estudado na mesma faculdade pode vir a beneficiar o eleitor. O fato de os seres humanos terem decidido com base em elementos emocionais por gerações mostrou-se uma estratégia bem-sucedida, que hoje explica a razão pela qual as emoções estão tão interligadas à decisão social.

Portanto, na concepção do presente estudo, as emoções fazem parte da decisão social. Dessa maneira, escolher como característica metodológica a exclusão das emoções dos estudos teria representado uma limitação à capacidade da ciência cognitiva em explicar o processo de decisão social.

6 CONSIDERAÇÕES FINAIS

A proposta desta dissertação foi analisar se ciência cognitiva estaria apta a fornecer uma explicação científica para a decisão social. O objetivo desta seção é estabelecer conclusões a respeito do estudo desenvolvido e sugerir recomendações para ponderações futuras relacionadas ao tema.

Este estudo partiu do pressuposto de que há dois elementos da ciência cognitiva que poderiam representar um obstáculo para explicar as decisões que envolvem o ambiente social. Um dos elementos seria a crença central de que o computador é o modelo mais viável para a compreensão da mente humana, e outro a opção metodológica de excluir as emoções dos estudos.

A conclusão central desta dissertação é que a característica metodológica de excluir as emoções dos estudos representa o principal problema da ciência cognitiva para explicar a decisão social. Portanto, a crítica que este trabalho faz à ciência cognitiva é sobretudo metodológica.

Para desenvolver a questão principal, iniciou-se com a apresentação da crítica de John Searle ao modelo computacional da ciência cognitiva. Searle refutou os argumentos desse modelo sobretudo com base na própria teoria da intencionalidade. Contudo, os críticos a Searle foram capazes de questionar as bases da intencionalidade nos moldes por ele propostos, o que fragilizou os argumentos de Searle contra a ciência cognitiva.

Adotou-se nesta dissertação a argumentação dos críticos de Searle. As diferenças entre a intencionalidade original e derivada não são tão acentuadas como Searle alega, de modo que ele não foi capaz de refutar o modelo computacional proposto pela ciência cognitiva.

Com o intuito de encontrar outros argumentos para a análise da possibilidade de a ciência cognitiva explicar a decisão social, tomou-se um aspecto específico dos pressupostos metodológicos que ela abarca: a exclusão da emoção dos estudos. Em seguida, analisou-se a questão da importância das emoções no processo de decisão social. Para isso, num primeiro momento foi utilizado o embasamento teórico fornecido por estudos neurocientíficos, sobretudo a hipótese do marcador somático de António Damásio. Em seguida, apresentou-se a psicologia evolucionista, especificamente a teoria do altruísmo recíproco de Robert Trivers, para a análise da relação entre emoção e decisão social.

As teorias apresentadas nas seções 4 e 5 indicam que a emoção pode ser considerada parte integrante do processo de decisão social. A neurociência, por meio de estudos empíricos, indicaria que as emoções estão presentes em decisões dessa natureza, enquanto a teoria do altruísmo recíproco explicaria o porquê.

Dessa forma, pode-se dizer que a escolha metodológica de excluir o estudo da emoção teria representando uma limitação à ciência cognitiva, ao menos no tocante à capacidade de explicar o processo de decisão social.

A exclusão das emoções dos estudos da ciência cognitiva é uma questão metodológica, e pode ser difícil para a ciência cognitiva alterar essa característica metodológica, em virtude das características das representações. Como argumentam Friedenberg e Silverman (2006, p. 441), seria um desafio à ciência cognitiva incorporar os módulos emocionais nos módulos mentais, e especificar a maneira pela qual ambos interagem. O problema dessa formulação seriam as diferenças qualitativas entre cognição e emoção. A neutralidade dos pensamentos os enquadra bem à representação simbólica, enquanto as sensações presentes nas emoções poderiam necessitar de uma modalidade diferente de representação.

Por outro lado, poderíamos nos perguntar como considerar a validade de uma teoria que busca explicar a decisão social sem apreciar o aspecto emocional. Seria possível fornecer uma explicação científica para a decisão social sem incluir as emoções? Para uma análise dessa natureza, seria necessário teorizar a respeito do quanto as emoções influenciam na decisão, e se é possível alguém, de maneira voluntária, se distanciar totalmente das emoções ao decidir.

Também há que se considerar uma dificuldade em estudar as decisões humanas de forma indistinta, pois as emoções são mais relevantes quando o fator social está envolvido. Estudos que buscam compreender o processo de decisão humano seriam mais frutíferos se procurassem, desde o início, diferenciar as modalidades de decisão que pretendem estudar. Como visto na seção 2.1, há diferentes modalidades de decisão, e que o que é verdadeiro para os apetites não o será para atos reflexos, nem para decisões complexas não relacionadas ao ambiente social. Cada uma dessas modalidades de decisão possui peculiaridades próprias, e estudá-las em separado é uma maneira mais segura de chegar a conclusões precisas.

Em suma, verifica-se a importância de introduzir, em análises futuras, um debate filosófico a respeito da relevância do elemento emocional para as decisões, principalmente aquelas que envolvem o ambiente social.

REFERÊNCIAS

ALVES, Pedro M.S. **Que Verdade no Erro de Descartes**. Philosophica, v. VII, Departamento de Filosofia, Faculdade de Letras de Lisboa, 1996, p. 171-178.

ARAÚJO, Saulo de Freitas. **Psicologia e Neurociência: Uma Avaliação da Perspectiva Materialista no Estudo dos Fenômenos Mentais**. Juiz de Fora: UFJF, 2011.

AUSTIN, John L. **Quando dizer é fazer**. Porto Alegre: Artes Médicas, 1990.

BELZUNG, Catherine. **Biologia das emoções**. Lisboa: Instituto Piaget, 2007.

BENNETT, M. R.; HACKER, P. M. S. **Fundamentos filosóficos da neurociência**. Lisboa: Instituto Piaget, 2005.

BOYD, Richard. **Materialism without reductionism**. In BLOCK, Ned (Editor). **Readings in Philosophy of Psychology, Volume 1**. USA: Library of Congress Cataloging in Publication Data, 1980.

BRENTANO, Franz. **Descriptive psychology**. New York: Routledge, 1995.

BUSS, David M.; CONFER Jaime C.; EASTON, Judith A.; FLEISCHMAN, Diana S.; GOETZ, Cari D.; LEWIS, David M. G.; PERILLOUX, Carin. **Evolutionary Psychology: Controversies, Questions, Prospects, and Limitations**. American Psychologist, v. 65, n. 2, 2010, p. 110-126.

BUSS, David M. **Evolutionary psychology: the new science of the mind**. USA: Pearson Education, 2008.

CAMPELL, Donald T. **Rationality and Utility from the Standpoint of Evolutionary Biology**. In: Hogarth & Reder, Rational Choice, 1987.

CANDIOTTO, Kleber B. B. **A perspectiva materialista não-reducionista de Dennett**. In: CHITOLINA, Claudinei L. et al (Org.). **A natureza da mente**. Maringá: Humanitas Vivens, 2011.

CANDIOTTO, Kleber B. B.; BASTOS, Cleverson L. **Da psicologia às ciências cognitivas**. Curitiba: Editora CRV, 2011.

CHRISTEN, Yves. **Uma introdução à sociobiologia**. Lisboa: Publicações Dom Quixote, 1981.

CHURCHLAND, Patricia S. **Neurophilosophy: Toward a Unified Science of the Mind-brain**. USA: The MIT Press, 1989.

CHURCHLAND, Paul M. **Eliminative Materialism and the Propositional Attitudes**. The Journal of Philosophy, v. 78, n. 2, 1981, p. 67-90.

CHURCHLAND, Paul M. **Matéria e consciência**. São Paulo: UNESP, 2004.

CHURCHLAND, Paul M. **Matter and Consciousness**. USA: The MIT Press, 1988.

COHEN, Jonathan D. **The Vulcanization of the Human Brain: A Neural Perspective on Interactions Between Cognition and Emotion**. Journal of Economic Perspectives, v. 19, n. 4, 2005, p. 3-24.

COSMIDES, Leda; TOOBY, John; BARKOW, Jerome. **The Adapted Mind: Evolutionary Psychology and the Generation of Culture**. New York: Oxford University Press, 1992.

DAMÁSIO, António. **O erro de Descartes**. São Paulo: Companhia das Letras, 1996.

DAVIDSON, Richard. **Toward a biology of personality and emotion**. Annals New York Academy of Sciences, v. 1, n. 1, 2008, p. 191-207.

DENNETT, Daniel C. **A perigosa ideia de Darwin**. Rio de Janeiro: Rocco, 1998.

DARWIN, Charles. **A origem das espécies.** São Paulo: Editora Escala, 2009.

DESCARTES, René. **Discurso do método.** Introdução, análise e notas de Étienne Gilson. São Paulo: Martins Fontes, 2009.

DESCARTES, René. **Descartes: obras escolhidas**. J. Guinsburg, Roberto Romano e Newton Cunha (Org.). São Paulo: Perspectiva, 2010.

FODOR, Jerry. **The language of thought**. Nova York: Thomas Y. Crowell Company, 1975.

FODOR, Jerry. **The mind-body problem**. In: John Heil (Editor). **Philosophy of mind. A guide and anthology**. Oxford: Oxford University Press, p. 168-82.

FRIEDENBERG, Jay; SILVERMAN, Gordon. **Cognitive science: an introduction to the study of mind**. Sage Publications, Inc., Thousand Oaks, 2006.

GARDNER, Howard. **A nova ciência da mente.** São Paulo: Edusp – Editora da Universidade de São Paulo, 2003.

GOLD, Ian; STOLJAR, Daniel. **A neuron doctrine in the philosophy of neuroscience**. Behavioral and Brain Sciences, v. 22, n. 5, 1999, p. 809-833.

GREENE, Joshua D.; SOMMERVILLE, R. Brian; NYSTROM, Leigh E.; DARLEY, John M; COHEN, Jonathan D. **An fMRI investigation of emotional engagement in moral judgment**. Science, v. 293, 2001, p. 2105-2107.

GRAY, Jeremy; BRAVER, Todd; RAICHLE, Marcus. **Integration of emotion and cognition in the lateral prefrontal cortex.** PNAS, New York, v. 99, n. 6, 2002, p. 4115-4120.

HAGGARD, Patrick. **Human volition: towards a neuroscience of will**. Nature

reviews/neuroscience, v. 9, 2008, p. 934-946.

HAIDT, Jonathan. **The moral emotions. Handbook of affective Sciences.** Oxford: Oxford University Press, 2003, p. 852-870.

HAMILTON, William D. **The Genetical Evolution of Social Behaviour.** Journal of Theoretical Biology, 7, p. 1-52, 1964.

HOFSTADTER, Douglas. **Reflections (on John Searle`s Minds, Brains and Programs).** In: HOFSTADTER, Douglas; DENNETT, Daniel C. (Org.). **The mind's I: fantasies and reflections on self and soul**. New York: Basic Books, 2000.

HOFSTADTER, Richard. **Social Darwinism in American Thought**. Beacon Press; Reprint edition, 1992.

KOENIGS, Michael; YOUNG, Liane; ADOLPHS, Ralph; TRANEL, Daniel; CUSHMAN, Fiery; HAUSER, Marc; DAMASIO, Antonio. **Damage to the prefrontal cortex increases utilitarian moral judgements**. Nature, v. 446, n. 7138, 2007, p. 908–911.

LA TAILLE, Yves de. **Moral e Ética: Dimensões Intelectuais e Afetivas.** Porto Alegre: Artmed, 2007.

LEVINE, Joseph. **Materialism and qualia: the explanatory gap.** Pacific Philosophical Quarterly, v. 64, 1983, p. 354-361.

MATTHEWS, Eric. **Mente: conceitos-chave em filosofia**. Porto Alegre: Artmed, 2007.

MAYR, Ernest. **O desenvolvimento do pensamento biológico**. Brasília: Editora Universidade de Brasília, 1998.

MOORE, Jim. **The Evolution of Reciprocal Sharing**. Ethology and Sociobiology, v. 5, p. 5-14, 1984.

MOLL, Jorge; OLIVEIRA-SOUZA, Ricardo. **Moral judgments, emotions and the utilitarian brain.** Trends in Cognitive Sciences, v. 11, n. 8, 2007, p. 319-21.

NAGEL, Ernest. **The structure of science**. Indiana: Hacket, 1995.

PINKER, Steven. **Como a mente funciona**. São Paulo: Companhia das Letras, 1998.

PUTNAM, Hilary. **Minds and machines**. In: Sidney Hook (Editor). **Dimensions of Mind**. New York: New York University Press, 1964.

PYLYSHYN, Zenon. **The "causal power" of machines**. Behavioral and Brain Sciences, v. 3, n. 3,1980, p. 442-444.

SEARLE, John R. **A Taxonomy of Illocutionary Acts**. In Searle, **Experience and Meaning. Studies in the Theory of Speech Acts**, Cambridge: Cambridge University Press, 1975, 1–29.

SEARLE, John R. **Intencionalidade**. São Paulo: Martins Fontes, 2002.

SEARLE, John R. **Mente, cérebro e ciência**. Lisboa: Edições 70, 1992.

SEARLE, John R. **Minds, brains, and programs**. Behavioral and Brain Sciences, v. 3, n. 3, 1980, p. 417-457.

SEARLE, John R. **Rationality in action.** London: The MIT Press, 2011.

SEARLE, John R. **The construction of social reality**. London: Penguin Books, 1996.

SEARLE, John R. **What is a Speech Act?** Philosophy in America, London: Allen and Unwin, 1965.

SMITH, Barry. **John Searle: From speech acts to social reality**. Cambridge: Cambridge University Press, 2003, p. 1–33.

TEIXEIRA, João de Fernandes. **Filosofia do cérebro**. São Paulo: Paulus, 2012.

TEIXEIRA, João de Fernandes. **Mente, cérebro e cognição**. Petrópolis: Vozes, 2011.

TEIXEIRA, João de Fernandes. **Mentes e Máquinas**. Porto Alegre: Artes Médicas, 1998.

TRIVERS, Robert L. **The evolution of reciprocal altruism**. The Quarterly Review of Biology, v. 46, n. 1, 1971, p. 35-57.

WILLIAMS, George C. **Adaptation and Natural Selection**. Princeton: Princeton University Press, 1966.