

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA
MESTRADO

JOSÉ AFONSO DE PAULA RETTO

**O PROBLEMA DA LIBERDADE DA VONTADE E A PSICOLOGIA
COGNITIVA**

JUIZ DE FORA

2016

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA
MESTRADO

JOSÉ AFONSO DE PAULA RETTO

**O PROBLEMA DA LIBERDADE DA VONTADE E A PSICOLOGIA
COGNITIVA**

Dissertação apresentada ao Programa de Pós-Graduação em Psicologia da Universidade Federal de Juiz de Fora como requisito para a obtenção do título de Mestre em Psicologia.

Orientador: Prof. Dr. Gustavo Arja Castañon

JUIZ DE FORA-MG

2016

TERMO DE APROVAÇÃO

Dissertação defendida e aprovada, em 15 de julho de 2016, pela banca constituída por:

Orientado: Prof. Dr. Gustavo Arja Castañon

Presidente: Prof. Dr. Richard Theisen Simanke

Membro Titular: Prof. Dr. Nilson Guimarães Doria

JUIZ DE FORA

2016

AGRADECIMENTOS

Gostaria de agradecer primeiramente ao meu orientador, Gustavo Arja Castañon, por ter aceito orientar o trabalho e pela paciência e apoio nos momentos difíceis. Agradeço também ao professor Saulo de Freitas Araujo e ao professor Helmuth Kruger pelas sugestões e correções no processo de qualificação. Além dos professores Richard Simanke e Nilson Guimarães Doria por terem aceito participar da banca de defesa. Sou grato também a todos os professores que compõem a linha História e Filosofia da Psicologia pela seriedade e integridade intelectual com que conduzem sua vida acadêmica. Ao professor Saulo em especial, gostaria de agradecer novamente por ter, desde o início de minha graduação, despertado minha curiosidade para as questões mais instigantes da psicologia e filosofia, me inspirando a sempre buscar uma formação intelectual mais sólida. Afetivamente esse trabalho só foi possível pelo apoio de minha mãe, Marisa, e minha amada, Natiely Borges, que foi a pessoa que acreditou na realização desse mestrado, até mesmo quando eu não o fiz, dando apoio emocional indispensável para que viesse a ser finalizado. Não poderia deixar de registrar também minha estima aos amigos de Juiz de Fora e colegas de mestrado, que suportaram meus desabafos quando o processo de escrita não caminhava e o desalento assombrava. Por fim, presto meus agradecimentos a UFJF e à CAPES pela bolsa concedida para a realização da pesquisa.

RESUMO

Esta dissertação aborda o tema da vontade, investigando como esse conceito é tratado na psicologia cognitiva. A partir disso, procuramos determinar se a psicologia cognitiva privilegia algum tipo de explicação para a vontade. Para tanto, inicialmente expusemos uma breve contextualização histórica do problema como abordado pela filosofia para situar, desta maneira, a forma como a psicologia cognitiva herdou a questão. Mostramos como o debate em torno do conceito da vontade se relaciona com o problema do livre arbítrio, a liberdade de ação e a responsabilidade moral assim como possui implicações referentes ao determinismo nomológico. Apresentamos os argumentos a favor e contra a possibilidade de uma vontade livre em um mundo determinista, de acordo com a visão compatibilista, que aceita que essas duas hipóteses sejam possíveis; a incompatibilista, que nega que uma vontade livre seja possível em um mundo determinista; e uma visão pessimista, que nega a possibilidade de uma vontade livre tanto num mundo determinista quanto indeterminista. Em seguida, investigamos os fundamentos da psicologia cognitiva, tais como o funcionalismo, o construtivismo, e a teoria do processamento de informação, a fim de rastrear possíveis filiações filosóficas ligadas ao problema da vontade. Procuramos também oferecer um resumo de como outras disciplinas pertencentes ao projeto de ciência da cognição influenciaram uma possível visão da psicologia cognitiva quanto a esta questão, principalmente a ciência da computação, com Alan Turing, e a neurociência, com os experimentos de Benjamin Libet. Por fim, apresentamos algumas considerações de influentes psicólogos quanto ao tema e vimos como o conceito de vontade está implicado nas investigações sobre a motivação, visto nesta se tratar dos motivos ou causas do comportamento, assim como no conceito de função executiva, relacionado aos atos voluntários, controle cognitivo global, planejamento e execução do comportamento. Concluímos que ainda parece precária uma tomada de posição definitiva quanto ao problema da vontade, como alguns pesquisadores parecem proclamar, mesmo à luz dos recentes avanços científicos.

Palavras-chave: vontade, psicologia cognitiva, filosofia da psicologia.

ABSTRACT

This dissertation approaches the issue of will, investigating how this concept is treated in cognitive psychology. From this, we sought to determine whether cognitive psychology favors some kind of explanation for the will. Therefore, initially we exposed a brief historical contextualization of the problem as discussed by philosophy to find how cognitive psychology inherited the issue. We show how the debate around the concept of the will relates to the problem of free will, freedom of action and moral responsibility as well as have implications for the nomological determinism. We present the arguments for and against the possibility of free will in a deterministic world, according to compatibilist view, accepting these two hypotheses are possible; the incompatibilist, which denies that a free will is possible in a deterministic world; and a pessimistic view, which denies the possibility of free will in either a determinist or indeterministic world. Then we investigated the foundations of cognitive psychology, such as functionalism, constructivism, and the theory of information processing in order to track possible philosophical affiliations linked to the problem of will. We also seek to provide an overview of how other disciplines pertaining to the cognition science project influenced a possible vision of cognitive psychology in this matter, especially computer science, with Alan Turing, and neuroscience, with the experiments of Benjamin Libet. Finally, we present some considerations of influential psychologists on the subject and we have seen how the concept of will is involved in research on motivation, since this is address the reasons or causes of behavior, as well as the concept of executive function, related to voluntary acts, global cognitive control, planning and execution of behavior. We conclude that it still seems precarious for a definitive position on the issue of the will, as some researchers seem to proclaim, even in light of recent scientific advances.

Key-words: will, cognitive psychology, philosophy of psychology.

SUMÁRIO

INTRODUÇÃO.....	01
CAPÍTULO I: O PROBLEMA DA VONTADE NA FILOSOFIA.....	06
1.1 Breve contextualização histórica sobre o conceito de vontade.....	06
1.2 Livre arbítrio, liberdade de ação e responsabilidade moral.....	15
1.3 O debate sobre livre arbítrio e vontade na filosofia contemporânea.....	18
1.3.1 Modelo de vontade como <i>faculdade</i>	18
1.3.2 Modelo hierárquico da vontade.....	18
1.3.3 Modelo da vontade como responsiva a razões.....	20
1.3.4 Livre Arbítrio e Determinismo.....	21
1.3.4.1 Determinismo Causal.....	21
1.3.4.2 Compatibilismo, Incompatibilismo e Pessimismo.....	24
1.3.4.2.1 Argumentos a favor do Incompatibilismo.....	26
1.3.4.2.2 Argumentos a favor do Compatibilismo.....	31
1.4 Considerações sobre o capítulo.....	34
CAPÍTULO II: PSICOLOGIA COGNITIVA: FUNDAMENTOS FILOSÓFICOS.....	36
2.1 Psicologia cognitiva como parte do projeto de uma ciência da cognição.....	38
2.2 Contribuições da ciência da computação à psicologia cognitiva.....	43
2.3 Funcionalismo.....	58
2.3.1 Funcionalismo computacional.....	66
2.3.2 Psico-funcionalismo.....	67
2.3.3 Funcionalismo analítico.....	69
2.3.4 Definições funcionais e a sentença de Ramsey.....	71
2.3.5 Caracterização de estados experienciais.....	72
2.3.6 Caracterização de estados intencionais.....	73
2.3.7 Caracterização dos <i>inputs</i> e <i>outputs</i> de um sistema.....	76
2.3.8 Críticas ao funcionalismo.....	77
2.4 Construtivismo.....	81
2.5 Processamento de Informação.....	83
2.6 Os experimentos de Benjamin Libet e suas repercussões.....	89

CAPÍTULO III: O PROBLEMA DA VONTADE NA PSICOLOGIA COGNITIVA.....	99
3.1 Plans and the Structure of Behavior.....	100
3.2 <i>Cognition and Reality</i>	104
3.3 Vontade e livre arbítrio segundo dicionários e enciclopédias de psicologia cognitiva.....	112
3.4 Motivação.....	122
3.5 Função Executiva.....	136
3.6 É possível falar em vontade livre a partir dos pressupostos da psicologia cognitiva?.....	143
 CONCLUSÃO.....	 149
 REFERÊNCIAS.....	 152

INTRODUÇÃO

No senso comum, ou na *folk psychology*, notamos uma referência constante à vontade quando, por exemplo, dizemos “não estou com vontade de fazer tal coisa”, ou “faça o que tiver vontade”, ou ainda “estou fazendo isso contra minha vontade”. Nos referimos à vontade como uma espécie de atributo de nossa psique que comanda nossas ações, ou melhor, escolhe entre as possíveis linhas de ação àquela que iremos tomar ou não tomar. Ainda assim, também nos defrontamos com situações nas quais olhamos para nosso passado e julgamos serem nossas experiências frutos de fatores que de certa forma nos influenciaram de uma maneira causal. Quando pensamos em nosso futuro, porém, nos defrontamos com diversas representações de uma situação em que teremos muitas opções a escolher. Da mesma maneira pode também ocorrer com uma representação do passado, onde mesmo reconhecendo fatores que determinaram a forma como agimos, temos a impressão de que poderíamos ter agido de forma diferente e, em alguns casos, até lamentamo-nos por escolhas equivocadas.

Nas situações exemplificadas acima, podemos dizer que o que de fato está em jogo é uma experiência subjetiva que aponta para uma inclinação humana de representar-se como dotada de uma vontade livre, ou seja, de acreditar que podemos unicamente através de nossa vontade determinar-nos a nós mesmos, sendo ela causa primeira de nosso agir. Essa impressão de que possuímos, pelo menos minimamente, algum grau de liberdade e controle sobre nossas ações nos é dada em nossa experiência imediata. Posso dizer, pelo menos a um nível pessoal, que, ao fazer uma experiência introspectiva, da forma mais direta possível, percebo que posso voluntariamente ficar em repouso ou escolher me mover (ou não), por exemplo levantando o braço direito, me virando para um lado ou outro, no momento em que me convier. O que quero dizer aqui é que num nível mais básico de ação é difícil me convencer que não sou livre para escolher, de que não ajo segundo minha vontade, ou seja, que não sou um agente que tem efeito causal sobre o mundo e vice-versa.

Seria de se esperar que a noção de vontade fosse central na Psicologia, na investigação e teorização das motivações, processos mentais e comportamento humano. Porém, esse tipo de discussão é bem menos frequente do que esperaríamos. Essa é muito mais frequente no campo da filosofia e, dado o ainda incipiente espaço da filosofia da psicologia em nossa disciplina, a formação do psicólogo não contempla uma discussão mais aprofundada sobre temas como a vontade e o livre arbítrio.

As teorias mais clássicas da psicologia, em sua maioria, apresentam um ponto de vista contrário a essa impressão imediata de que somos movidos por uma vontade livre. Como, por exemplo, a posição de Skinner em seu *O Mito da Liberdade* (1973), ou de Freud com seu determinismo psíquico (Freud, 1900; 1901; 1905; 1938). Encontramos apenas no Humanismo uma teoria que incorpora em sua visão teórica um ser humano possuidor de livre arbítrio (Maslow, 1968).

Por outro lado, em minha prática como psicólogo clínico, no contato com aqueles que buscam serviços psicológicos, pelos mais variados motivos, desenvolvi a crença de que independentemente da abordagem adotada, está se tratando de uma pessoa que deve ser considerada e respeitada como um ser livre, ou pelo menos que se deve acreditar na liberdade como uma potencialidade a ser desenvolvida. Na maior parte das vezes, aquele que tem um sofrimento psicológico busca uma libertação desse sofrimento, uma maior autonomia, ou maior controle de si mesmo. Tomada de consciência, responsabilização sobre suas ações, autonomia, tudo isso parece estar ancorado na noção de liberdade do sujeito, em sua capacidade de fazer escolhas, em tomar as rédeas de sua própria vida.

Como compatibilizar um posicionamento ético do psicólogo frente ao ser humano, com teorias desenvolvidas dentro da própria psicologia, ou mesmo em outras áreas, como filosofia e neurociências, que tomam o ser humano como uma espécie de máquina? Seríamos assim uma engrenagem em um sistema mecânico, determinista, onde teríamos pouco ou mesmo nenhum poder de escapar a essas determinações que nos acometem vindas de todos os níveis possíveis: desde o biológico até o social, cultural, etc., passando muitas das vezes por fatores que nem ao menos tomamos conhecimento, sejam fatores inconscientes, subliminares, ou manipulações veladas. Ainda assim, a própria sociedade parece estar ancorada numa ideia de liberdade do sujeito. Desde um nível jurídico, onde há a necessidade de responsabilização dos atos para que o sujeito possa ser punido; passando por um nível religioso, onde existe a noção de liberdade, livre arbítrio, em muitas religiões das mais diversas culturas e épocas históricas; até num nível médico/psicopatológico, como alterações da vontade: hipobulias, com debilidade da vontade e potencial volitivo, compulsões, manias, etc.

Apesar de tudo isso, vemos hoje uma crescente onda de estudos que se posicionam frente a esse debate da liberdade da vontade humana com um posicionamento negativo. Muitos desses impulsionados pelos experimentos de Benjamin Libet, como o famoso artigo *Unconscious cerebral initiative and the role of conscious will in voluntary action* (1985), apesar de, nesse mesmo artigo, o próprio Libet considerar que suas descobertas não deveriam ser

tomadas como avessas a noção de vontade livre, mas sim como algo que afeta a visão de como a vontade livre opera na vida humana. Onde os processos associados com a responsabilidade individual e com a vontade livre operariam não como uma instância iniciadora dos atos voluntários, mas como uma instância que selecionava esses atos e os controlava. Porém, esses estudos influenciaram visões mais extremistas, como por exemplo de Wegner & Wheatley (1999), que tomaram os achados de Libet como provas para a *tese da ilusão* da vontade livre e, através de experimentos que criavam nos sujeitos um senso de autoria em ações que de fato não haviam realizado, desenvolvem a teoria da “*Aparente Determinação Causal Mental*”. Na mesma linha, vemos também o trabalho de Kiesel & colaboradores (2005) onde se busca demonstrar que certos tipos de escolhas, consideradas pelas pessoas como sendo realizadas livremente, podem em grande parte ser apenas fruto de uma influência causal subliminar.

Nas pesquisas mais recentes dentro das neurociências cognitivas (Gazzaniga, 2006) notamos uma tendência a uma visão mais negativa em relação a liberdade da vontade. Dentro das ciências cognitivas, vemos uma retomada dessas questões, com grande número de publicações sobre o assunto. O que a psicologia cognitiva pode acrescentar a esse debate? Já que uma de suas propostas como um novo empreendimento dentro do estudo científico da cognição diz buscar uma reaproximação com temas psicológicos da filosofia, estando desde seus primórdios muito ligada a uma reflexão filosófica (Gardner, 1996; Penna, 1984). Gardner, aliás, ao falar dos fundamentos da ciência cognitiva, diz que houve uma retomada do estudo de conceitos tais como *vontade* e *propósito*, que haviam sido abandonados pelo behaviorismo por serem conceitos vagos ou muito genéricos. Reeve (2006), em seu manual sobre motivação, diz que o estudo da motivação pela psicologia cognitiva busca oferecer a melhor resposta possível para uma pergunta fundamental, que seria “o que causa o comportamento?”. Para ele, o conceito de vontade seria uma espécie de “grande teoria” explicativa para o que hoje é chamado de motivação humana, e poderia ser considerada a origem filosófica para os conceitos motivacionais estudados na psicologia cognitiva contemporânea, que não mais utilizaria grandes teorias explicativas, mas micro teorias, buscando dar conta de responder por diversas frentes as questões levantadas no debate, já por demais sobrecarregado, do tema na filosofia.

Tendo então, como questão central, o *problema da vontade*, o objetivo do presente trabalho é levantar como é tratado o conceito de vontade na psicologia cognitiva, analisando-o e comparando-o com seu tratamento no debate filosófico. Além disso, procuramos determinar se a psicologia cognitiva privilegia algum tipo de explicação para a vontade e se oferece um posicionamento quanto a questão da vontade ser livre ou não.

Para contextualização do debate atual do tema na filosofia utilizamos principalmente para a busca das fontes primárias autores levantados na pesquisa exploratória na base de dados *PhilPapers*. Posteriormente, realizamos uma busca para tentar cobrir o principal do tratamento do tema na psicologia cognitiva, aí então por meio de pesquisa na base de dados Psycinfo e *Philosopher's Index*.

Em relação à sua estrutura geral, organizamos o presente estudo em três capítulos. No primeiro capítulo, apresentamos inicialmente uma breve contextualização histórica quanto ao problema da vontade e, após isso, partimos então para uma descrição do estado do debate atual em torno do conceito de vontade no cenário filosófico para situar, desta maneira, a forma como a psicologia cognitiva herdou a questão. Dada a dificuldade do tema, e o grande número de autores que trataram sobre a questão da vontade, o objetivo desse primeiro capítulo é apenas o de contextualizar o debate na filosofia. Estabeleceremos, então, os sentidos dos termos vontade e liberdade da vontade, ou livre arbítrio, e a questão do compatibilismo/incompatibilismo, central na discussão sobre a temática. Como veremos, o *problema da vontade* está intimamente atravessado pelos problemas ligados aos conceitos de causalidade, acaso, determinação, indeterminação etc., o que requer, portanto, uma tomada de posição referente ao debatidíssimo problema do determinismo contra liberdade ou livre arbítrio.

No segundo capítulo, apresentamos uma definição do que chamamos de psicologia cognitiva e como ela se enquadra como uma das disciplinas de um projeto mais abrangente de ciência da cognição. Buscamos mostrar também como as outras disciplinas desse projeto, como por exemplo a inteligência artificial, filosofia da mente, linguística, neurociências, antropologia, estão sendo radicalmente influenciadas pelo trabalho uma das outras. Portanto, fizemos uma exposição das contribuições da ciência da computação à psicologia cognitiva, notadamente do trabalho de Alan Turing. Conceitos tais como de computação e controle (esse último advindo da cibernética), segundo o cognitivismo, podem aguçar questões psicológicas já que estes expressariam ideias sobre processos mentais mais claramente do que conceitos verbais poderiam expressar, utilizando-se, para tanto, da linguagem da computação ou teoria da informação como forma de descrevê-los. Da filosofia da mente, como fundamento filosófico para a psicologia cognitiva, destacamos o funcionalismo, e como essa visão permitiria falar sobre a natureza dos estados mentais e fazer proposições explicativas para o comportamento humano. Além disso, apresentamos os experimentos do neurocientista Benjamin Libet e seus colaboradores, que tiveram forte repercussão, principalmente no campo de estudo da consciência, dos atos volitivos e livre arbítrio.

No terceiro capítulo apresentamos uma análise de duas obras que foram fundamentais para a psicologia cognitiva e onde em nossa pesquisa encontramos algo relevante a ser dito sobre o tema do problema da vontade e livre arbítrio. A primeira delas, *Plans and the Structure of Behavior* (1960), de Miller, Galanter e Pribram, que com essa obra estabeleceram um marco importante para o cognitivismo, encontramos um reconhecimento sobre como o problema da vontade humana seria um problema central na também nascente psicologia cognitiva. Na segunda obra que gostaríamos de destacar, *Cognition and Reality* (1976), Ulric Neisser oferece algumas considerações sobre a liberdade humana e controle e predição do comportamento. Posteriormente, investigamos como o conceito de vontade e livre arbítrio foi definido em dicionários e enciclopédias de psicologia e ciência cognitiva, a fim de termos uma visão mais atual da discussão. Fizemos, também, uma incursão mais geral na área de estudo da motivação, devido a abrangência da área e os vários modelos existentes de abordagem dos motivos da ação humana, além de apresentarmos o conceito neuropsicológico de função executiva, muito em voga em estudos recentes da psicologia cognitiva e que é relacionado aos atos voluntários, controle cognitivo global, planejamento e execução do comportamento. Por fim, levantamos a questão de se seria possível falar em uma vontade livre partindo-se de uma visão cognitivista.

CAPÍTULO I

O PROBLEMA DA VONTADE NA FILOSOFIA

Apresentamos neste capítulo uma breve contextualização histórica quanto ao problema da vontade. Onde buscamos mostrar como diferentes filósofos conceituam o termo, além de suas respostas aos problemas envolvidos com essa conceituação. Como veremos, é um debate extenso onde todos os grandes filósofos tiveram ao menos algo a falar sobre o tema, por conseguinte, buscaremos apresentar apenas aqueles de maior impacto e, por motivos de espaço, apenas de uma forma resumida. Mostramos também como o conceito de vontade pode se relacionar ao livre arbítrio, a liberdade da ação e a responsabilidade moral. Após isso, partimos para uma descrição do estado do debate atual em torno do conceito no cenário filosófico e dos vários modelos existentes da vontade, assim como suas implicações referentes ao determinismo nomológico. Por fim, apresentaremos os principais argumentos, a favor e contra, de uma possível visão da vontade livre em relação ao determinismo causal.

1.1. Breve contextualização histórica sobre o conceito de vontade

Segundo Ferrater-Mora (2004), o conceito de vontade foi tratado durante o percurso da história da filosofia por quatro diferentes pontos de vista: 1) psicologicamente, a vontade pode ser entendida como uma faculdade humana, como expressão de um certo tipo de ato; 2) moralmente, a vontade foi tratada em relação com os problemas da intenção e a questão da busca de se alcançar o Bem; 3) teologicamente, como aspecto fundamental e, para alguns autores, básico da realidade ou personalidade divina; e 4) metafisicamente, considerando-se a vontade como um princípio da realidade e motor de toda mudança.

Dado o objetivo proposto neste trabalho, onde se busca uma aproximação com uma visão científica em relação aos problemas envoltos no conceito de vontade, nos limitaremos às questões relativas ao ponto de vista psicológico desse conceito, mesmo sendo em alguns momentos difícil separar tais pontos de vista, já que podem estar intimamente relacionados em alguns autores.

Visto que diversos autores conceberam a vontade como vontade humana, ou seja, um atributo ou característica humana, outros ainda podem falar em vontade referindo-se a atos

voluntários ou volições. Nesse último caso, a palavra vontade se torna algo problemática, já que, se tomada literalmente, leva a objetivar os atos mencionados, com a vontade sendo uma espécie de “entidade”. A fim de evitar tal objetivação, tanto em textos filosóficos como psicológicos, hoje em dia a palavra vontade é usada menos frequentemente ou, se usada, apenas como uma abreviatura mais ou menos cômoda de atos voluntários. Da mesma forma, podemos ler em Abbagnano (2007) que, atualmente, nem a filosofia nem a psicologia fazem interpretações desse tipo relativas à conduta humana já que “as noções de *comportamento* e de *forma*, bem como a tendência *funcionalista* da psicologia não permitem falar de ‘princípios’ da atividade humana” (p. 1007).

Apesar dessa divergência, diversos autores, como veremos, insistiram no caráter irreduzível da vontade e dos atos voluntários. Por um lado, se buscou fazer uma distinção entre a vontade e o desejo ou impulso; e, por outro lado, uma distinção entre a vontade e a inteligência ou a razão. Podemos encontrar também autores que tem a opinião de que a vontade está sempre em uma relação muito estreita ou entrelaçada com outras faculdades ou tipos de atos. Particularmente foram acentuadas as relações existentes entre atos da vontade e razões para se exercer a vontade, ou para querer algo. Aqueles que insistem no caráter irreduzível da vontade tendem a considerá-la como algo “não racional”. Aqueles que relacionam os atos da vontade com outras faculdades ou atos, tendem a considerar que a vontade é dirigida por “razões”, ou preferências, que podem ser objeto de deliberação, ou que a vontade é um elemento num contínuo que pode partir desde impulsos ou instintos, às vezes mecanizados e organicamente instituídos, até atos de execução (Ferrater-Mora, 2004).

Esses exemplos de posições tomadas em relação à vontade, podem ser encontrados por toda história da filosofia. Desde Platão e Aristóteles, que tenderam a classificar potências ou poderes da alma. Como na famosa divisão tripartite da alma – sendo a mesma divisão feita na sociedade ou Estado – Platão, em sua República (2002) considera a vontade como um tipo de divisão. Encontrando-se abaixo da faculdade da razão, que dirige o homem (assim como a sociedade e o Estado), e acima do desejo ou apetite sensível. Aí a vontade não é uma faculdade intelectual, mas tampouco pode-se dizer que seja irracional já que seus atos se executam conforme a razão. Sendo diferenciada do desejo já que ao seguir os desejos cegamente não se está exercendo o poder da vontade, mas se esta apenas sendo dominado pelos apetites sensíveis. Vemos aí já uma diferenciação entre vontade e desejo, estando o último na ordem do sensível ou concupiscível e a primeira pertencendo a ordem do intelecto. Aristóteles (2010) também insistiu no caráter racional da vontade, mas tendo em comum com o desejo apenas o fato de ser

um motor, ou seja, de mover a alma, pois a vontade “apetece”, ou seja, é um apetite, assim como o desejo.

O maior ou menor grau que se possa atribuir à vontade como motora vai depender da relação que for estabelecida entre vontade e razão. E, segundo Ferrater-Mora (2004), esse foi um problema fundamental na filosofia medieval, não só por ordem psicológica mas também teológica, e que de certa forma deu continuidade ao que foi apresentado por Platão e Aristóteles, filósofos que tiveram fundamental influência na teologia cristã. Esse é o caso de São Tomás de Aquino, que teve suas doutrinas sobre a natureza da vontade fortemente assentadas em bases aristotélicas. Contudo, uma nova ideia do homem – ou, como foi chamada, “homem novo” – principalmente como expressão da obra de Santo Agostinho, foi levando a um maior destaque e preeminência do conceito de vontade, não só do homem como em Deus, o que fomentou deste modo o que foi chamado de “voluntarismo”, geralmente em oposição ao “intelectualismo”. Alguns autores que tiveram relevância na corrente voluntarista, durante a Idade Média, foram São Pedro Damiano, Duns Scotus e Guilherme de Ockham, na época moderna, autores como Descartes e, mais posteriormente, porém com um tipo especial e característico de voluntarismo, podemos citar Wundt (Marcellos, 2012).

Comum a essa corrente voluntarista é a ideia de que há na alma humana ações e paixões, figurando entre as ações os atos intelectivos, sendo que todas as ações se encaixam, ou mesmo se reduzem à vontade. Portanto, os próprios atos intelectivos são dirigidos pela vontade. Desta maneira, pode-se dizer que somente através de um ato de vontade se pode ajuizar. Vários psicólogos voluntaristas modernos, por exemplo Wundt, admitem a primazia da vontade no plano anímico, mesmo que rejeitando conceber a vontade como a realidade, característica dos filósofos voluntaristas metafísicos, entre eles Schopenhauer (2005).

Essa contraposição entre um destaque maior da vontade ou um destaque maior do intelecto, mesmo sem detrimento da vontade, nesse último caso, pode ser visto em São Tomás de Aquino e em Duns Scotus. Na *Suma de Teologia* (2001), São Tomás considera que a vontade não se encontra submetida em nenhum de seus atos à necessidade, sendo, desta forma, vontade e livre-arbítrio não duas potências distintas, mas antes uma única potência. Podemos perceber em São Tomás de Aquino uma pequena mudança em relação à noção aristotélica de vontade. Para Aquino, toda ação humana visa, direta ou indiretamente, um fim. Em outras palavras, o comportamento humano é sempre motivado. Pensamos sobre o que queremos realizar e colocamos isso como uma meta. Fazemos isto em virtude de nosso intelecto à luz de seu desejo fundamental para o bem, que é construído na vontade. Em seguida, sentimos uma atração ou

desejo para esse determinado objetivo ou fim, e nossa vontade nos inclina em direção a ele. Em seguida, pensamos em como atingir esse objetivo ou fim, ou seja, nos envolvemos na atividade de deliberação. Fazemos, então, um julgamento final sobre o que fazer e escolhemos o que fazer com base nesse julgamento. Aquino argumenta que a escolha é uma função da vontade à luz de um julgamento pelo intelecto. Em outras palavras, a vontade move o agente para uma determinada ação, ação esta determinada pelo intelecto.

Porém, São Tomás de Aquino acrescenta um outro aspecto que influencia a ação, que ele chama de paixões. As paixões são um pouco parecidas com a nossa concepção de emoções. São sentidas como estados motivacionais, assim como a raiva ou a alegria, podendo ter um efeito positivo ou negativo sobre o que fazemos. No entanto, embora as paixões sejam influências muito poderosas sobre nossas ações, podendo fazer parecer bom o que normalmente não considerariamos como bom, ainda assim as paixões não podem esmagar a inteligência e a vontade (desde que essas estejam funcionando corretamente) e, desta maneira, determinar o que fazemos.

Segundo Aquino, nós não somos necessariamente conscientes de todas essas atividades, mas para que elas ocorram não há a necessidade de sermos conscientes delas. Essa consideração se baseia no princípio de que se o ser humano é capaz de fazer algo tem de haver algum poder ou capacidade que lhe permite fazê-lo. John Duns Scotus, porém, argumenta que se Aquino está correto, os seres humanos não agem livremente. Isso porque, na visão de Scotus, o intelecto é determinado pelo ambiente externo. O conteúdo de nossas crenças e julgamentos é em função do mundo que nos rodeia e não estaria dentro de nosso controle. Essa característica de nossas crenças e sua relação com o mundo significa que o intelecto não é livre. Assim, se Aquino está correto em afirmar que o movimento da vontade é determinado pela atividade do intelecto e, se é verdade que o intelecto não é livre, a vontade também não o será, logo os seres humanos não agiriam livremente (McClusky, 2007).

Para salvaguardar a liberdade da vontade, Scotus nega a forte conexão entre vontade e intelecto, como vista em Aquino, argumentando que a vontade não é determinada por um julgamento do intelecto. Scotus se baseia em nossa experiência comum para defender esta reivindicação. Todos nós já estivemos em situações em que sabemos o que devemos fazer e ainda assim não estamos dispostos a fazê-lo. O aluno sabe que deve estudar para os exames, mas está tão confortável ficar deitado no sofá que ele não se levanta para estudar. Ele só vai se levantar para estudar quando realmente quiser fazer isto e nenhum julgamento vai movê-lo para esse curso de ação enquanto houver uma oposição de seu desejo. Scotus descreve este tipo de

caso como aquele em que a vontade, a fonte de seu desejo de permanecer no sofá, quer (ou, neste caso, não quer), em oposição ao julgamento do intelecto. Assim, a vontade é livre de determinação pelo intelecto. Scotus ainda concorda com Aquino que a vontade depende do intelecto para identificar possíveis cursos de ação, mas rejeita a visão de que o julgamento do intelecto determina a escolha da vontade (McClusky, 2007). As considerações de Tomás de Aquino e de John Duns Scotus são paradigmas úteis para ilustrar algumas das vantagens e desvantagens das abordagens voluntaristas e intelectualistas da ação e sua liberdade. Daí têm-se a visão de que, se a vontade é livre, ela não pode ser determinada, não havendo, portanto, explicações possíveis para o que a move em direção as várias escolhas possíveis, sejam elas dadas pela razão, paixões, etc. Como causa final, existe uma espécie de aleatoriedade no seu modo de agir, que não se sujeita a nenhuma regra imposta.

Para Xavier Zubiri (1992), tanto a posição de São Tomás de Aquino quanto a de Duns Scotus são parciais. Na medida em que a inteligência coloca o homem perante a realidade, torna possível, desta maneira, levar a vontade a se decidir, ou seja, concluir suas tendências. Porém essas tendências são inconclusivas no sentido de não poderem ser concluídas. A vontade não é algo que vem “desde fora” para levar a cabo tal conclusão, mas sim é a própria tendência. Não é, portanto, ou não somente é, apetite racional, como propunha São Tomás e tampouco livre determinação, como queria Duns Scotus, ou ainda atividade espontânea, como Zubiri diz ser declarado por alguns autores modernos. Apetite, determinação e ação seriam dimensões de um mesmo ato, o “querer”. Sendo assim, a essência da vontade, segundo Zubiri, seria o querer. Não somos arrastados por nossas tendências, mas antes nos deparamos em presença delas. Sendo tais tendências aspirações entre as quais podemos ter certas preferências ou não.

Deste modo, podemos não somente querer algo, como da mesma forma podemos querer realizarmo-nos em algo que queremos. Porém, não se pode querer sempre o que se quer, mas o que se “pode querer”. E o que “se pode querer” pode ser cultivado até chegar-se a um “domínio de si”, ou pode estreitar-se até chegar à “escravização de si mesmo”. O domínio de si, resultado do cultivo da possibilidade de querer, é a liberdade da vontade. Liberdade esta que tampouco nos chega “desde fora”, mas sim “desde dentro”, como modalidade de certos atos. Desta maneira, para Zubiri (1992), a liberdade está fundada na volição, porém, como a volição é “tendenciosa”, também está nas tendências, que portanto não constituem um obstáculo para a liberdade, mas antes são condição de possibilidade desta.

No que foi chamado de época moderna clássica, aproximadamente entre os séculos XVII e XVIII, o tema da vontade também ocupou lugar de destaque nos debates filosóficos, em

teologia, psicologia e epistemologia. Segundo Ferrater-Mora (2004), duas grandes tendências opostas de pensamento podem ser notadas referentes à relação que se considerou manterem entre si a vontade e o intelecto. Uma primeira tendência é a dos racionalistas, entre os quais podemos citar como mais destacados, Descartes e Leibniz. O primeiro sendo decididamente voluntarista, considera que a vontade é a faculdade de assentir ou negar o juízo. Em sua relação com o intelecto, a vontade é aquela que decide, ou seja, todo ato intelectual é um ato da vontade. Na verdade, segundo Descartes (1997) todos os atos ou ações, ao contrário das paixões, são atos da vontade. Portanto, pode-se dizer que Descartes seja voluntarista e ainda racionalista, porém não intelectualista. “A liberdade da nossa vontade reconhece-se sem provas, apenas pela experiência que temos dela” (p. 41) e ainda, “a vontade tem mais extensão do que o entendimento, derivando disso os nossos erros” (p. 39). Já Leibniz, com respeito ao voluntarismo, se opõe a Descartes, embora não se oponha com respeito ao racionalismo. Para ele, segundo Ferrater-Mora (2004), a vontade se origina no pensamento e não se pode admitir que os erros dependam mais da vontade do que do intelecto, em suma, só queremos o que nos aparece ao intelecto.

Uma segunda tendência é a dos empiristas, de Hobbes a Hume. De acordo com Ferrater-Mora (2004), é comum a eles a ideia de que não há apetite racional e, mesmo o ato voluntário sendo o começo de uma ação que resulta de uma deliberação anterior, nem por isso podemos dizer que o ato voluntário seja “intelectualizado”. Desse ponto de vista, os empiristas ainda podem ser considerados voluntaristas, não porque considerem que a vontade ou os atos voluntários sejam predominantes, mas negam que sejam algo racional ou intelectual. Os empiristas tendem a igualar o ato voluntário à própria execução da ação. Não há, para eles, uma determinação do intelecto ou um ato de vontade que atue como motor da ação, mas antes, o que se chama ação já inclui em si o ato voluntário, logo não faria sentido pensar em um ato de vontade sem uma ação correspondente.

Immanuel Kant, por sua vez, destaca o aspecto moral da vontade. Assim como sua filosofia conseguiu ser um meio termo para o empirismo e o racionalismo, mostrando que ambas as teorias do conhecimento estavam em parte corretas, o mesmo pode ser dito sobre suas considerações referentes ao conceito de vontade, à questão do livre arbítrio e do determinismo. Kant não opõe determinismo e liberdade da vontade. Ele afirma que o determinismo pertence à ordem dos fenômenos, enquanto a liberdade pertence à ordem do *númeno*. O fenômeno pode ser definido em oposição a coisa em si, “são-nos dadas coisas como objetos dos nossos sentidos e a nós exteriores, mas nada sabemos do que elas possam ser em si mesmas; conhecemos

unicamente os seus fenômenos, isto é, as representações que em nós produzem, ao afetarem os nossos sentidos” (Kant, 1987, p.58). É a nossa razão que organiza os fenômenos segundo a lei da causalidade, portanto, para Kant, a diferença entre fenômeno e númeno é que vai nos permitir resolver a antinomia de determinismo e liberdade. O homem, enquanto fenômeno, é determinado no tempo pela lei da causalidade, porém, enquanto númeno, permanece livre, não sendo desta maneira determinado por essas mesmas leis.

Cabe aqui uma explicação do que vem a ser *númeno* em Kant. De forma breve, podemos dizer que é a realidade em si mesma, que está para além do que pode ser conhecido. Pode apenas ser pensada, mas, ainda assim, permanece, e permanecerá sempre, dada a própria estrutura de nossa razão, um objeto incognoscível. Podemos, por analogia, considerá-la a “causa” externa da possibilidade de conhecimento, sem esta não teríamos segurança alguma de que existe algo para além de nós mesmos, o que levaria, em última análise, ao solipsismo. Assim, o que move minha vontade, ou minha vontade mesma, está para além do mundo dos fenômenos, daquilo que pode ser conhecido, só podemos conhecer a manifestação dessa vontade depois que ela se dá no mundo fenomênico.

A Crítica da Razão Pura ensina a tomar o objeto numa dupla significação, qual seja, como *fenômeno* e como *coisa em si*. O princípio da causalidade incide sob o objeto tomado como fenômeno, aí sim a vontade pode ser pensada conforme a lei natural, sendo determinada e não-livre. Ao mesmo tempo a vontade pode ser pensada enquanto coisa em si mesma, logo não sendo submetida à lei natural, sendo, portanto, livre. O que não implica necessariamente uma contradição. Para Kant não podemos conhecer a alma mediante a razão especulativa e muito menos conhecê-la empiricamente, da mesma forma a liberdade, podemos apenas pensá-las. Ao estudar a sucessão temporal, Kant afirma: “todas as mudanças acontecem segundo o princípio da ligação de causa e efeito” (2001a, p.217), mas anuncia na *Antinomia da Razão Pura* a tese de que “a causalidade segundo as leis da natureza não é a única de onde podem ser derivados os fenômenos do mundo no seu conjunto. Há ainda uma causalidade pela liberdade que é necessário admitir para os explicar” (p.406).

Afirma que “a lei da natureza consiste precisamente em nada acontecer sem uma causa suficiente determinada *a priori*” (Kant, 2001a, p.406), mas se buscarmos as causas de estados antecedentes, seremos conduzidos a uma cadeia infinita de “causas”, não havendo um primeiro início, nenhuma completude da série das causas precedentes. É preciso admitir “uma *espontaneidade absoluta* das causas, espontaneidade capaz de dar início *por si* a uma série de fenômenos que se desenrola segundo as leis da natureza e, por conseguinte, uma liberdade

transcendental” (p.408). Assim, Kant afirma que a ideia transcendental de liberdade constitui “o conceito da absoluta espontaneidade da ação, como fundamento autêntico da imputabilidade dessa ação” (p.408), tendo que ser admitida uma faculdade de iniciar espontaneamente uma série de coisas sucessivas ou de estados.

O conceito prático de liberdade se funda justamente sobre a ideia transcendental da mesma, como a independência do arbítrio em relação à coerção de impulsos da sensibilidade. Um arbítrio é sensível ao ser afetado patologicamente por motivações da sensibilidade, mas o homem tem uma faculdade de determinar-se por si mesmo independentemente da coação de impulsos sensíveis. Kant denominará, então, *inteligível* o que no objeto dos sentidos não é fenômeno, sendo possível, neste objeto do mundo sensível, ter uma faculdade que não é objeto da intuição sensível, podendo assim considerar a causalidade deste ser por duas perspectivas: “como *inteligível* quanto à sua *ação*, considerada a de uma coisa em si, e como *sensível* pelos seus *efeitos*, enquanto fenômeno no mundo sensível” (Kant, 2001a, p.466). Desta maneira, formar-se-ia um conceito empírico e um conceito intelectual em relação à causalidade de um sujeito, que teria lugar juntamente num só e mesmo efeito.

Assim, teríamos no sujeito do mundo dos sentidos um *caráter empírico*, onde suas ações como fenômenos estão conectadas de forma causal com outros fenômenos, segundo leis constantes da natureza; e ao mesmo tempo um *caráter inteligível*, que embora seja causa de suas ações como fenômenos, não se subordina ele mesmo a quaisquer condições da sensibilidade, não sendo fenômeno, mas uma *coisa em si* mesma. Desta maneira, Kant afirma que a liberdade e a natureza, cada qual com seu significado pleno, se encontrariam, ao mesmo tempo e sem conflito, nas mesmas ações, conforme referência à sua causa sensível ou inteligível.

De fato, a liberdade em Kant é fundamental para sua filosofia prática, “prático é tudo aquilo que é possível pela liberdade” (2001a, p.636), e nesta existiriam leis morais puras que determinam *a priori* o uso da liberdade. Em 1785, Kant publica a *Fundamentação da Metafísica dos Costumes*, onde desenvolve os conceitos de vontade, liberdade, autonomia e moralidade. Ao se referir à vontade, afirma: “Tudo na natureza age segundo leis. Só um ser racional tem a capacidade de agir *segundo a representação* das leis, isto é, segundo princípios, ou: só ele tem uma *vontade*.” (Kant, 2001b, p.47) Prosseguindo, podemos ler: “a *vontade* é uma espécie de causalidade dos seres vivos, enquanto racionais, e *liberdade* seria a propriedade desta causalidade, pela qual ela pode ser eficiente independentemente de causas estranhas que a

determinem, assim como a *necessidade natural* é a propriedade da causalidade de todos os seres irracionais de serem determinados à atividade pela influência de causas estranhas” (p. 93).

Em muitos casos, o estudo do conceito de vontade levou a considerações sobre o conceito de liberdade, ou a ilusão ou ausência de liberdade. Ao se conceber a vontade como um “motor”, pode-se em princípio afirmar que há vontade sem que haja necessariamente liberdade ou livre arbítrio, sendo o que se chama “vontade” algo como um movimento natural. De acordo com Ferrater-Mora (2004), uma visão determinista da vontade pode estar baseada não só numa concepção do que se qualificou como “causalidade da Natureza”, mas também na ideia de que um ato voluntário pode ser determinado por “razões”. Em muitos casos, como vimos, autores que buscam compreender em que consiste, ou como funciona, a vontade, não admitiram o determinismo. Kant, entretanto, admitia um determinismo na forma de um rigoroso encadeamento causal do reino natural que conhecemos apenas em seu aspecto fenomênico. Porém, não há determinismo no reino moral ou dos fins morais, desta maneira podemos falar sem rodeios de vontade e de liberdade. Posteriormente a Kant, houveram outros importantes teóricos, como Fichte e Schopenhauer, que continuaram a investigação sobre o tema, porém, sobretudo no sentido metafísico.

Vários filósofos no final do século XIX e começos do XX, trataram a vontade como uma noção central, Exemplos são Wundt e William James. Tanto a psicologia como a metafísica de Wundt são declaradamente voluntaristas. William James também faz um exame da vontade de um ponto de vista psicológico, desenvolvendo a ideia de que a vontade está ligada a alguma representação que não é obstaculizada por nenhuma outra que lhe seja contrária (Ferrater-Mora, 2004).

No pensamento filosófico da atualidade as questões relativas ao conceito de vontade são abordadas de forma bem diferente das “metafísicas da vontade” das quais o século XIX foi generoso. Em termos gerais, há uma maior tendência a examinar não tanto “a vontade”, mas os atos volitivos ou as volições. Onde os autores que investigam tal tema tendem a considerar que o exame desses atos pressupõe levantar a questão de eles serem ou não voluntários, no sentido de se partem realmente de uma escolha consciente do agente, ou mesmo se essa escolha de fato existe ou pode existir. O *problema da vontade* está, de acordo com isso, intimamente atravessado pelos problemas ligados aos conceitos de causalidade, acaso, determinação, indeterminação etc., o que requer, portanto, uma tomada de posição sobre o problema do determinismo contra liberdade ou livre arbítrio.

1.2. Livre arbítrio, liberdade de ação e responsabilidade moral

Por que razão deveríamos nos preocupar se temos ou não livre arbítrio? Um dos motivos se deve a estreita relação entre livre arbítrio e outros dois temas filosóficos importantes: a liberdade de ação e a responsabilidade moral. Porém, apesar da relação íntima entre esses conceitos, eles não devem ser confundidos, sendo assim, é importante que uma diferenciação entre eles seja feita.

Sem uma reflexão mais apurada, consideramos que, pelo menos na maior parte das situações, nossas ações são livres pois estamos a exercer nosso livre arbítrio. Para facilitar a exposição, consideremos um exemplo proposto por Timpe (2006) de uma mulher, Alice, que está ponderando se deve ou não sair para passear com seu cachorro. Alice pensa consigo mesma: “Eu deveria passear com o cachorro, ele precisa se exercitar. Porém não quero sair pois está muito frio lá fora. Mas acho que a melhor decisão a tomar seria levá-lo para passear...”. Como vimos, o livre arbítrio está diretamente relacionado a liberdade de ação. Antes de Alice sair ou não para passear com o cachorro, primeiro ela deve decidir ou escolher que ação tomar. Assumindo que ações humanas resultam de suas capacidades racionais, podemos ver que a possibilidade de ações livres depende da possibilidade de possuir livre arbítrio: dizer que um agente humano age de forma livre é minimamente dizer que esse agente pode exercer sua vontade ou escolha de forma livre.

Vários filósofos definiram livre arbítrio de forma muito semelhante a essa. Thomas Hobbes, no seu *Leviatã* (1973), sugeriu que liberdade consiste em, não havendo impedimentos externos, um agente fazer o que ele quer fazer: “por liberdade entende-se, conforme a significação própria da palavra, a ausência de impedimentos externos, impedimentos que muitas vezes tiram parte do poder que cada um tem de fazer o que quer, mas não podem obstar a que use o poder que lhe resta, conforme o que seu julgamento e razão lhe ditarem” (p.133). Hume, em *Uma Investigação sobre o Entendimento Humano* (2003), ponderando sobre a liberdade, diz que ela é o poder de agir ou não agir de acordo com a determinação da vontade, isto é, se escolhermos ficar estáticos, iremos. Se escolhermos nos mover, também iremos. Essa liberdade hipotética é universalmente permitida a todos que não são prisioneiros.

Segundo esses autores, isto sugere que a liberdade é a capacidade de selecionar um curso de ação, sendo um agente livre aquele que não é impedido por nenhum obstáculo externo de concluir seu curso de ação. Assim, Hobbes e Hume sustentariam que Alice é livre para caminhar com seu cachorro desde que nada a impeça de levar a cabo sua decisão de caminhar com ele.

Da mesma forma, ela é livre para não passear com seu cachorro, desde que nada a obrigue a acompanhá-lo caso ela decida não o fazer.

No entanto, ainda podemos argumentar que a explanação anterior é falha por não fazer uma distinção importante entre dois tipos possíveis de liberdade: liberdade da vontade e liberdade de ação. Essa distinção é válida já que é possível possuir livre arbítrio sem no entanto possuir liberdade de ação. Suponhamos que antes de Alice ter feito a escolha por passear com seu cachorro ela tenha tirado um cochilo. Enquanto isso acontecia, começou a nevar e a neve bloqueou as portas e janelas de sua casa, sendo assim, mesmo que ela quisesse, não poderia. Nesse caso, Alice ainda teria livre arbítrio, já que escolheu livremente passear com o cachorro, porém não teria liberdade de ação, já que estaria impossibilitada de fazê-lo por fatores externos a sua vontade.

Partindo desses pressupostos, a liberdade de ação seria dependente de uma tomada de decisão interna livre, ou de um livre arbítrio. Portanto, se se pode ou não presumir que alguém tenha liberdade de ação vai depender primeiro se essa pessoa tem ou não livre arbítrio. Assim, se hipoteticamente Alice sofresse uma lavagem cerebral durante seu cochilo fazendo-a inevitavelmente querer sair para passear com o cachorro, mesmo que não houvessem impedimentos externos, sair para passear com o cachorro não seria uma ação livre. A razão para isso é que a decisão de Alice não teria surgido de seu livre arbítrio, mas, no caso, de lavagem cerebral. Desta maneira, o livre arbítrio seria condição necessária para uma ação livre, mesmo presumindo serem os dois conceitos distintos. No que se segue, a frase “agindo por livre vontade” significa engajar-se em uma ação como o resultado da utilização do livre arbítrio. O uso da frase não nega a distinção entre livre arbítrio e ação livre (Timpe, 2006).

Um segundo motivo para nos preocuparmos sobre o livre arbítrio é que ele parece ser necessário para a responsabilidade moral. Mesmo existindo diversas considerações sobre o que seja exatamente responsabilidade moral, é amplamente aceito que podemos diferenciá-la de responsabilidade causal. Como exemplo, considere um galho que cai de uma árvore em cima de um carro estacionado logo abaixo, amassando sua lataria. Enquanto podemos dizer que o galho é causalmente responsável pelo carro amassado, não é moralmente responsável por isso, pois galhos não são agentes morais. Dependendo da cadeia causal, é possível alguém ser moralmente responsável por algo, mesmo não sendo diretamente responsável no sentido causal. Para presente propósito, vamos apenas considerar que um agente moral é responsável por um evento ou estado de coisas apenas se ele é passível de louvor ou culpa moral por esse mesmo evento ou estado de coisas. De acordo com a visão dominante sobre a relação entre

responsabilidade moral e livre arbítrio, da mesma forma que no caso da ação livre, o livre arbítrio seria condição necessária para a responsabilidade moral. Por exemplo, alguém que é coagido ou forçado a agir contra sua própria vontade não teria responsabilidade moral sobre esse ato (Timpe, 2006).

Porém, alguns filósofos não acreditam que o livre arbítrio seja necessário para a responsabilidade moral. De acordo com John Martin Fischer (1994), agentes humanos não têm livre arbítrio, mas eles ainda assim são moralmente responsáveis por suas escolhas e ações. De forma bem resumida, Fischer acha que o tipo de controle necessário para a responsabilidade moral é mais fraco do que o tipo de controle necessário para o livre arbítrio. Além disso, ele acha que a verdade do determinismo causal impediria o tipo de controle necessário para o livre arbítrio, mas que não impediria o tipo de controle necessário para a responsabilidade moral. Muitos dos filósofos que acreditam não possuímos tal coisa como livre arbítrio, podem, apesar disso, ainda assim advogarem a favor da necessidade, mesmo que apenas prática, da responsabilidade moral.

A questão de se somos realmente livres ao fazermos nossas escolhas, como podemos ver, está estreitamente relacionada com importantes questões no campo da moral e ética, e também em aspectos jurídicos. Psicólogos, hoje em dia, podem ser requisitados para dar um parecer sobre se uma pessoa pode ou não ser responsabilizada por suas ações, por exemplo por apresentar algum distúrbio psicológico ou outro fator que influencie em seu julgamento. Pressupõe-se, aí, que aspectos cognitivos que por ventura viessem a estar prejudicados pudessem interferir no comportamento, inclusive de forma a que essa pessoa não pudesse ser mais responsabilizada por suas ações. Quando essa mesma pessoa se encontra consciente de suas ações, admite-se que ela tem controle de si e de suas capacidades cognitivas e, portanto, é responsável por seus atos.

No entanto, questões relativas ao tema do livre arbítrio não se limitam a sua necessidade para a ação livre e responsabilidade moral. Vários filósofos sugerem que o livre arbítrio também é um requisito para a própria racionalidade, a autonomia e dignidade das pessoas, assim como a criatividade, a cooperação e o valor da amizade e do amor (Kane, 1998; e Ekstrom, 1999). Vemos, portanto, que o livre arbítrio é fundamental para muitas dessas questões.

1.3. O debate sobre livre arbítrio e vontade na filosofia contemporânea

Com o desenvolvimento dessas diferentes linhas de pensamento referentes ao problema da vontade, foram propostos modelos que buscam integrar essas diferentes visões em modelos explicativos da vontade e o que a permite ser considerada livre. Apresentamos a seguir os principais modelos da vontade como entendida num debate mais contemporâneo na filosofia – vontade como *faculdade*, vontade como “malha” hierárquica, vontade como responsiva a razões.

1.3.1. Modelo de vontade como *faculdade*

O modelo de vontade como uma faculdade tem origem nos escritos de filósofos antigos tais como Platão e Aristóteles e foi a visão dominante na filosofia medieval e moderna, como pudemos ver anteriormente na contextualização histórica, tendo ainda alguns defensores na literatura contemporânea. Todos os seres vivos, segundo esse modelo, possuem algumas capacidades tais como, por exemplo, a capacidade de crescimento e reprodução. Da mesma forma, a inteligência e a volição também são capacidades, ou, dizendo de outra maneira, inteligência e vontade são faculdades, no caso, faculdades presentes nos seres humanos que os caracterizam como agentes livres, ou possuidores de livre arbítrio. Segundo esse modelo, é em virtude de termos essas faculdades adicionais, e da interação entre elas, que possuímos livre arbítrio (Timpe, 2006).

O intelecto apresenta a vontade coisas descritas como boas. Sendo a vontade, ou a faculdade volitiva, um apetite para o bem, ela é naturalmente atraída para tais coisas. A vontade também é capaz de comandar as outras faculdades; pode comandar o corpo para se mover ou o intelecto para considerar algo. Assim, através da interação entre o intelecto e vontade, um agente tem livre arbítrio para buscar algo que ele vê como bom.

1.3.2. Modelo hierárquico da vontade

Segundo O'Connor (2014) e Timpe (2006), um relato contemporâneo que teve grande influência é o modelo hierárquico da vontade de Harry Frankfurt (1969), também conhecido como concepção “estruturalista” ou “malha” da vontade. Neste modelo, o agir humano é considerado livre se tem uma certa estrutura interna ou “malha” entre os vários níveis de desejos e vontades. De acordo com esse modelo, os agentes podem ter diferentes tipos de desejos.

Frankfurt sugere que uma diferença central entre a atividade humana e a animal é nossa capacidade de refletir sobre nossos próprios desejos e crenças e formar desejos e julgamentos sobre esses primeiros desejos. Por exemplo, posso querer comer um doce (desejo de primeira ordem), mas também posso querer *não querer* o doce (desejo de segunda ordem), pois posso estar de dieta. A diferença entre desejo de primeira ordem e desejo de segunda ordem, ele argumenta, fornece a chave para entender tanto o livre arbítrio quanto a liberdade de ação.

Nem todos os nossos desejos resultam em ação. Na verdade, se tenho desejos conflitantes, então é impossível satisfazê-los todos. Um desejo que se torna ação, para Frankfurt, é o que pode ser chamado de vontade, ou seja, vontade é o desejo que move o agente para a ação, o desejo que é efetivado. Minha vontade é livre quando posso tornar qualquer desejo de primeira ordem um desejo sobre o qual eu ajo, da mesma forma quando posso tornar um desejo de segunda ordem em ação. Pois, segundo Frankfurt, um desejo de segunda ordem refletiria meu verdadeiro eu. De acordo com a visão hierárquica da vontade, o livre arbítrio consistiria, então, em ter volições de segunda ordem. Em outras palavras, um agente tem uma vontade livre se ele é capaz de ter o tipo de vontade que ele quer ter. Um exemplo que pode ser dado é o de um dependente químico, normalmente ele age segundo um desejo com o qual ele não quer estar de acordo. Sua vontade está dividida e suas ações procedem de desejos com os quais ele não se identifica reflexivamente. Portanto, segundo esse modelo, ele não agiria livremente.

A objeção central feita por comentaristas sobre esse modelo é: o que há de tão especial sobre desejos de ordem superior? Por que supor que eles refletem inevitavelmente o meu verdadeiro eu contra os desejos de primeira ordem? Frankfurt é explícito quando diz que os desejos de ordem superior não precisam ser enraizados na perspectiva moral ou mesmo na personalidade de uma pessoa (Frankfurt, 1971). Assim, parece que, em alguns casos, um desejo de primeira ordem pode ser muito mais um reflexo do meu verdadeiro eu (mais “interno para mim”, na terminologia de Frankfurt) do que um desejo mais fraco, que busca ser um tipo de pessoa que quer de certa forma ser diferente.

Em escritos posteriores, Frankfurt primeiro responde a esta preocupação apelando para “decisões tomadas sem reservas” e, em seguida, à uma ordem superior de desejos com os quais alguém se sente “satisfeito” ou realizado, de tal forma que não se tenha inclinações para fazer mudanças nesse desejo. Todavia, a ausência de uma inclinação para mudar o desejo, obviamente, não atinge a condição de identificação que confere liberdade. Parece que um estado

tão negativo de satisfação seja apenas um estado no qual se possa encontrar, e não um estado que precise ou não de aprovação (Pettit, 2001).

1.3.3. Modelo da vontade como responsiva a razões

Fischer e Ravizza (1998) apresentam esse modelo, que tem um espaço importante na discussão contemporânea sobre o tema envolvendo o livre arbítrio, e que toma como ponto de partida a afirmação de que a “*agency*” (termo sem correspondente exato em português, aproximadamente equivaleria a auto-orientação, livre arbítrio, autonomia) envolve uma sensibilidade a certas razões. Um agente age com livre arbítrio se ele é sensível às considerações racionais apropriadas, e não age com livre arbítrio se não tem tal capacidade de resposta. Como exemplo, considere novamente o caso de Alice e sua decisão de caminhar com seu cachorro. Uma visão da vontade como responsiva a razões diz que Alice teria livre arbítrio para caminhar com seu cachorro se, caso ela tivesse boas razões para não caminhar com o cachorro, então ela decidiria não sair para caminhar com ele. Imagine se Alice tivesse ligado a televisão depois de acordar de sua soneca e soubesse da nevasca antes de decidir caminhar com o cachorro. Se soubesse da nevasca, ela teria tido uma boa razão para decidir não passear com o cachorro. Mesmo que tais razões nunca tivessem ocorrido a ela (isto é, se ela não tivesse tomado conhecimento da nevasca antes de sua decisão), sua disposição para que tais razões influenciem suas volições mostra que ela é sensível às razões.

Nesse ponto de vista, a coerção e a manipulação podem minar o livre arbítrio em virtude de fazer os agentes não responsivos a razões. Se for feita uma lavagem cerebral para Alice sair para passear com o cachorro em um determinado momento, então, mesmo que ela tivesse bons motivos para não o fazer, por exemplo ligar o noticiário e ver que está nevando, ela iria tentar passear com o cachorro da mesma forma. Assim, os agentes manipulados não são responsivos a razões, em virtude dessa falta de liberdade na vontade. Esse modelo tem sido muito utilizado também para se entender questões relativas a comportamentos compulsivos e de dependentes químicos, como um tipo de comportamento onde a vontade é tida como “enfraquecida”, ou não livre, por não responder a razões (Poland and Graham, 2011; Levy, 2013).

1.3.4. Livre Arbítrio e Determinismo

A afirmação de que a vontade humana é livre fere à visão de um mundo regido por leis naturais? Para buscar responder a esse questionamento, iremos analisar nos subitens a seguir o determinismo causal e se seria possível aceitar uma concepção de vontade livre admitindo-se que o mundo em que vivemos seja um mundo determinista. Ou seja, se a vontade livre é compatível ou não com o determinismo. Para tanto, apresentamos a visão compatibilista, que aceita que essas duas hipóteses sejam possíveis; a incompatibilista, que nega que uma vontade livre seja possível em um mundo determinista; e uma visão pessimista, que nega a possibilidade de uma vontade livre tanto num mundo determinista quanto indeterminista.

1.3.4.1. Determinismo Causal

Grande parte dos estudos contemporâneos sobre o livre arbítrio se concentram em se ele é ou não compatível com o determinismo causal, que também, por vezes, pode ser chamado de *determinismo nomológico*. É importante distinguir o determinismo causal de outros tipos de determinismo, como determinismo lógico ou determinismo teológico. Determinismo causal (doravante simplesmente “determinismo”) é a tese de que o curso do futuro é inteiramente determinado pela conjunção do passado e as leis da natureza. De acordo com Ferrater-Mora (2004) “numa acepção geral, o determinismo sustenta que tudo o que houve, há e haverá, e tudo o que aconteceu, acontece e acontecerá está de antemão fixado, condicionado e estabelecido, não podendo haver nem acontecer senão o que está de antemão fixado, condicionado e estabelecido” (p. 690). Esse condicionamento prévio abrangeria todos os fenômenos do universo, não se limitando unicamente as ações humanas, como a ideia de destino ou predestinação, sendo por isso universal e portanto uma tese metafísica. Além disso, o determinismo está quase sempre, mas não necessariamente, associado à ideia de uma causalidade que rege todo o universo.

Timpe (2006) propõem que imaginemos uma proposição que descreve completamente a maneira como o universo inteiro foi em dado momento do passado, digamos, 100 milhões de anos atrás. Vamos chamar essa proposição de “P”. Imaginemos também uma proposição que expressa o conjunto de todas as leis da natureza; chamaremos essa proposição de “L”. O determinismo é, então, a tese de que a conjunção de P e L implicam um futuro único. Dado P e L, só há um futuro possível, uma forma possível para que as coisas aconteçam. Dizendo de outra forma, o determinismo é a tese de que todos os estados de coisas que aconteceram em

algum momento no passado, quando conjugadas com as leis da natureza, implicam que o mundo possível é o mundo real. Uma vez que um mundo possível inclui os estados de coisas que irão ocorrer, a verdade do determinismo equivale à tese de que o passado e as leis da natureza implicam os estados de coisas que irão ocorrer no futuro, e que apenas os estados de coisas decorrentes da passado e as leis de fato ocorrem.

Outro fator importante é que um sistema ser determinado é diferente de ser previsível. É possível que o determinismo seja verdadeiro e que mesmo assim ninguém seja capaz de prever o futuro. O fato de que nenhum agente humano sabe ou é capaz de saber as verdades futuras não tem qualquer influência sobre se há verdades futuras decorrentes da conjunção do passado e das leis da natureza. No entanto, existe uma ligação mais fraca entre a tese do determinismo e a previsibilidade do futuro. Se o determinismo for verdadeiro, então um ser com um conhecimento completo de P e L, e com capacidades intelectivas suficientes, deve ser capaz de prever infalivelmente o caminho que o futuro vai seguir. No entanto, dado que os seres humanos não têm nem o conhecimento relevante e nem as capacidades intelectivas necessárias, o fato de que não somos capazes de prever o futuro não é evidência para a falsidade do determinismo (Timpe, 2006).

Se o determinismo é ou não é verdadeiro isto seria uma questão contingente; ou seja, o determinismo não é necessariamente verdadeiro, nem é necessariamente falso. Dito de outra forma, seria possível tanto um mundo determinista quanto um mundo não determinista. Sendo assim, a verdade ou falsidade do determinismo torna-se uma questão empírica, a ser descoberta investigando-se a forma como o mundo é. Isto não é negar que a verdade do determinismo tenha implicações metafísicas. Por um lado, a verdade do determinismo implica que as leis da natureza não são meramente probabilísticas – pois se assim fossem, então o conjunto do passado e as leis da natureza não implicariam em um único futuro possível. Além disso, como veremos adiante, nossa preocupação central relativa ao tema se concentra nas implicações que a verdade do determinismo teria para o conceito de vontade e livre arbítrio.

O ponto a ser observado no momento é que, se a verdade do determinismo é uma verdade contingente sobre a forma como o mundo realmente é, a investigação científica deve nos dar uma visão sobre este assunto. Assim como a psicologia – ou, para propósitos do presente trabalho, a psicologia cognitiva – deve nos dar uma visão sobre este assunto referente ao agir humano. Digamos que um mundo possível é determinista se o determinismo causal é verdade nesse mundo. Há pelo menos duas maneiras que poderíamos refutar uma visão determinista de mundo. Como já referido, se as leis da natureza em um determinado mundo fossem

probabilísticas, então tal mundo não seria determinista. Em segundo lugar, se houvessem entidades nesse mundo que não fossem totalmente regidas por leis da natureza, então, mesmo que essas leis fossem elas próprias deterministas, o mundo não seria determinista.

Alguns filósofos e cientistas hoje em dia sugerem que certas descobertas na física contemporânea nos dão boas razões para duvidar da verdade do determinismo. Por exemplo, a interpretação padrão da teoria quântica, a Interpretação de Copenhague, desenvolvida por Neils Bohr e Werner Heisenberg, sustenta que as leis que regem a natureza são indeterministas e probabilísticas. De acordo com essa interpretação, se uma partícula, como um quark, deve ou não desviar em uma direção específica em um determinado momento, isso é descrito apenas adequadamente por equações probabilísticas. Sendo essas equações irreduzíveis, ou seja, não são um mero reflexo do desconhecimento de variáveis hipotéticas escondidas, como geralmente acontece em outros usos de equações probabilísticas. Embora essas equações possam prever a probabilidade de um quark se desviar para um lado ou outro, em um determinado momento, se ele realmente desvia ou não isso é algo indeterminista ou totalmente aleatório (Leite & Simon, 2010).

Há também interpretações deterministas da Teoria Quântica, como a Interpretação de Everett, ou a Interpretação de muitos mundos (IMM), que postulam a existência de “universos paralelos”. Felizmente, por se tratar de um assunto de grande complexidade e que demanda conhecimentos muito especializados, o resultado do debate sobre se a Teoria Quântica é mais corretamente interpretada determinística ou indeterministicamente deve ser evitado para os nossos atuais propósitos. Como argumentado por Blanshard (1964), mesmo se sistemas de micro partículas como quarks se comportam de forma indeterminista, pode ser que sistemas que envolvam objetos físicos maiores, como carros, cães e pessoas sejam deterministas. É possível que o único indeterminismo seja na escala de micro partículas e que macro objetos obedeçam às leis deterministas. Se for este o caso, então o determinismo causal, como o definido acima, rigorosamente falando, seja falso, mas é “quase” verdadeiro. Ou seja, poderíamos substituir o determinismo rigoroso pelo “quase determinismo” (*near-determinism*), a tese de que, apesar da indeterminação quântica, o comportamento de todos os outros objetos num nível dos macros objetos, incluindo aí as nossas ações físicas, obedeceriam a leis deterministas.

Quais seriam as implicações da verdade de qualquer determinismo? Ou, mais especificamente, quais seriam as implicações para o debate da vontade humana e livre arbítrio? Uma maneira de pensar sobre essas implicações seria fazer a seguinte pergunta: Será que ainda

poderíamos ser livres mesmo sendo o determinismo causal (ou quase determinismo) verdadeiro?

1.3.4.2. Compatibilismo, Incompatibilismo e Pessimismo

No final da seção anterior nos deparamos com a seguinte questão: poderíamos ter livre arbítrio mesmo que o determinismo seja verdadeiro? Essa questão é uma maneira útil para pensarmos o problema da possibilidade de uma vontade livre mesmo em um mundo regido por leis deterministas, além de possibilitar uma diferenciação entre as principais posições em relação ao livre arbítrio. Chamamos de *compatibilistas* aqueles que acreditam que podemos ter livre arbítrio mesmo que o determinismo causal seja verdadeiro (ou mesmo que um quase-determinismo seja verdadeiro. No que se segue, iremos omitir essa qualificação). Em outras palavras, para esses autores, a existência do livre arbítrio em um mundo possível é compatível com um mundo determinista. De acordo com o compatibilista é possível que um agente seja determinado em todas as suas escolhas e ações e ainda assim algumas de suas escolhas serem feitas de forma livre (O'Connor, 2014).

Já de acordo com os *incompatibilistas*, a existência do livre arbítrio é incompatível com a verdade do determinismo. Se um determinado mundo possível é determinístico, então nenhum agente nesse mundo tem livre arbítrio. Há pelo menos dois tipos de incompatibilistas. Os que pensam que o determinismo é verdadeiro para o mundo real e, portanto, nenhum agente no mundo real tem livre arbítrio: tais incompatibilistas são frequentemente chamados de “deterministas rígidos” (Pereboom, 2001). Outros incompatibilistas pensam que o mundo real não é determinista e que pelo menos alguns dos agentes do mundo real têm livre arbítrio: estes são referidos como “libertários” (Kane, 2014; Palmer, 2014). No entanto, estas duas posições não são as únicas, já que, por exemplo, é possível que alguém seja incompatibilista, ache que nosso mundo não seja determinista, e ainda assim pense que não temos livre arbítrio. Embora seja menos claro como se possa categorizar tal posição, talvez “negadores do livre arbítrio”, ela ilustra que o determinismo rígido e o libertarianismo não esgotam os tipos possíveis de incompatibilistas. Uma vez que todos incompatibilistas, independentemente do tipo, concordem que a falsidade do determinismo é uma condição necessária para o livre arbítrio, e os compatibilistas não concordem com esta afirmação, falaremos simplesmente de incompatibilistas e compatibilistas, seguindo a sugestão de Van Inwagen (2008), já que muitas

dessas terminologias acabam por causar uma confusão conceitual desnecessária relativa a um debate já por demasiado complexo.

Outro ponto importante é ter em mente que tanto o compatibilismo como o incompatibilismo são afirmações sobre possibilidade. De acordo com o compatibilista, é possível que alguém seja completamente determinado e ainda assim livre. O incompatibilista, por outro lado, sustenta que um tal estado de coisas é impossível. Mas nenhuma posição por si só está fazendo uma declaração sobre a existência ou não de agentes que na verdade possuem o livre arbítrio. Suponhamos por um momento que o incompatibilismo seja verdadeiro. Se a verdade do determinismo é uma questão contingente, então, agentes serem ou não moralmente responsáveis dependerá se o mundo real é ou não determinista. Além disso, mesmo que o mundo real seja indeterminista não se segue imediatamente que o indeterminismo presente seja do tipo necessário para que se admita o livre arbítrio (vamos retornar a um ponto semelhante abaixo quando se considera uma objeção ao incompatibilismo). Da mesma forma, assumindo que, tanto o compatibilismo seja verdadeiro e que o determinismo causal seja verdadeiro no mundo real, não se segue daí que agentes do mundo real, inevitavelmente, possuem livre arbítrio.

Finalmente, existe ainda a visão pessimistas sobre o livre arbítrio (Broad, 1952; Strawson, 1994). Os pessimistas concordam com os incompatibilistas que o livre arbítrio não é possível se o determinismo for verdadeiro. No entanto, ao contrário dos incompatibilistas, os pessimistas não acham que uma visão de mundo indeterminista ajudaria. Na verdade, dizem eles, em vez de ajudar a apoiar o livre-arbítrio, o indeterminismo o debilita. Consideremos, novamente, como exemplo, Alice deliberando sobre levar seu cão para passear. De acordo com os pessimistas, se Alice é determinada, ela não pode ser livre. Mas se o determinismo for falso, então haverá indeterminação em algum momento antes de sua ação. Exatamente onde se localiza essa indeterminação dependerá da visão particular sobre a natureza do livre arbítrio. Vamos supor que a indeterminação está localizada nas razões que ocorrem para Alice. É difícil ver, o pessimista argumenta, como essa indeterminação poderia favorecer um livre arbítrio, pois, já que a ocorrência dessas razões seja algo indeterminista, ter razões não estaria, da mesma forma, dentro do controle de Alice. Já que, se Alice decide com base em quaisquer razões que ela tenha, então sua vontade é baseada em algo fora de seu controle. Baseia-se, em vez disso, no acaso. Assim, os pessimistas pensam que a adição do indeterminismo faz com que não tenhamos o tipo de controle necessário para o livre arbítrio. Enquanto o pessimismo pode parecer a mesma posição defendida pelos negadores de livre arbítrio, o pessimismo é uma

afirmação mais forte. Negadores do livre arbítrio, enquanto pensam que o livre arbítrio é possível, dizem simplesmente que ele não é real: os agentes de fato não têm livre arbítrio. Os pessimistas, no entanto, têm uma posição mais forte, argumentam que o livre arbítrio é impossível. Não só não temos livre arbítrio como não há maneira alguma em que poderíamos tê-lo (Strawson, 1994). Assim, a única maneira de preservar a responsabilidade moral, para o pessimista, é, portanto, negar que o livre arbítrio seja uma condição necessária para a responsabilidade moral.

Como nos mostra os pessimistas, mesmo uma resolução para o debate entre compatibilistas e incompatibilistas não vai por si só resolver o debate sobre se temos ou não temos realmente livre arbítrio. No entanto, é para este debate que agora nos voltamos.

1.3.4.2.1. Argumentos a favor do incompatibilismo

Incompatibilistas dizem que o livre arbítrio é incompatível com o determinismo. Nem todos os argumentos para o incompatibilismo podem ser considerados aqui; vamos nos focar em duas variedades principais dos argumentos correntes no debate contemporâneo em torno do tema, segundo O'Connor (2014) e Timpe (2006). A primeira variedade é construída em torno da ideia de que ter o livre arbítrio é uma questão de ter uma escolha sobre algumas de nossas ações, e que ter uma escolha é uma questão de ter opções genuínas ou alternativas sobre o que se faz. A segunda variedade de argumentos é construída em torno da ideia de que a verdade do determinismo significa que nós não causamos nossas ações de forma propriamente dita, ou seja, nós não originamos nossas ações de uma forma significativa, portanto nossas ações não são controladas por nós. Em outras palavras, falta-nos a capacidade de autodeterminação. Vamos considerar um argumento representante de cada variedade.

A. Argumento Consequência ou “modelo do jardim de caminhos que se bifurcam”

O argumento mais conhecido e influente a favor do incompatibilismo desse primeiro conjunto de argumentos é chamado de “Argumento da Consequência”, e tem sido defendido principalmente por Carl Ginet (1966) e Peter van Inwagen (1983; 2008). É baseado em uma distinção fundamental entre o passado e o futuro. Consideremos, primeiro, uma apresentação informal desse argumento, conforme vemos em Timpe (2006). Com base na direção do fluxo do tempo, parece haver uma profunda assimetria entre o passado e o futuro e, da mesma

maneira, na direção da causalidade. O passado parece ser fixo e inalterável. Não obstante, parece que o mesmo não acontece em relação ao futuro, no geral acreditamos que podemos ter uma influência sobre o que está por vir através de nossas vontades e ações subsequentes.

A assimetria entre passado e futuro é ilustrada pelo fato de que nós não deliberamos sobre o passado da mesma forma que deliberamos sobre o futuro. Enquanto podemos deliberar sobre se uma ação passada foi realmente a melhor ação que poderíamos ter feito, deliberamos sobre o futuro de uma maneira diferente. Podemos questionar se nossas ações passadas foram de fato as melhores ações a serem tomadas, já quanto as ações futuras, podemos tanto nos questionar quais seriam as melhores ações possíveis bem como que atos futuros podemos executar. Assim, parece que o futuro está aberto de uma maneira que o passado não está. Em outras palavras, quando um agente está usando seu livre arbítrio, o que ele está fazendo é escolher entre uma variedade de opções diferentes para o futuro, cada uma das quais é possível, dado o passado e as leis da natureza. Por esta razão, este ponto de vista do livre arbítrio é muitas vezes chamado de “Modelo do jardim de caminhos que se bifurcam” (Timpe, 2006).

Uma maneira natural para pensarmos o livre arbítrio é em termos do futuro de um agente como um jardim de caminhos que se bifurcam e se ramificam a partir de um único caminho, já que agir com vontade livre requer possibilidades alternativas. Um *locus* de ação livre surge quando as ofertas atuais, a partir de um passado singular de um agente, abre vários caminhos possíveis para vários futuros possíveis. Se tornou, então, célebre utilizar o mote em referência ao conto do fabulista argentino Jorge Luis Borges. Digamos, como o Jardim de caminhos que se bifurcam sugere que, quando uma pessoa age por sua própria vontade, ela poderia ter agido de outro modo. Sua capacidade de ter agido de outra forma é subscrita pela sua capacidade de ter escolhido entre cursos alternativos de ação (McKenna, 2009).

Porém, essa visão do livre arbítrio entra em conflito direto com o determinismo, já que, entendido em sentido estrito como caracterizado acima, ele nos diz que a qualquer momento, dado os fatos do passado e as leis da natureza, apenas um futuro é possível. Entretanto, o modelo do jardim de caminhos que se bifurcam sugere o contrário, que um agente livremente disposto poderia ter agido de uma forma diferente da que agiu, portanto, que mais de um futuro é possível. Temos aqui, então, como podemos ver em McKenna (2009), um argumento incompatibilista que codifica as considerações acima expostas:

- A. Um agente qualquer, x , executa um ato qualquer a de livre e espontânea vontade se e somente se x tem o controle sobre a .

- B. x tem o controle sobre a somente se x tem a capacidade de escolher entre cursos alternativos de ação para agir de forma a .
- C. Se x tem a capacidade de escolher entre cursos alternativos de ação para agir de forma a , então existem cursos alternativos de ação para agir de forma a abertos para x (ou seja, x poderia ter agido de outra forma que a)
- D. Se o determinismo é verdadeiro, então somente um futuro é possível dado o passado factual e as leis da natureza.
- E. Se apenas um futuro é possível, dado o passado factual e as leis da natureza, então não há cursos alternativos de ação a qualquer ato aberto a qualquer agente (ou seja, nenhum agente poderia ter agido de outra forma além da que realmente agiu).
- F. Portanto, se o determinismo é verdadeiro, não é o caso que qualquer agente, x , execute qualquer ato, a , de sua livre e espontânea vontade.

Para facilitar a apresentação, vamos simplificar o argumento acima da forma que se segue:

1. Se uma pessoa age segundo seu livre arbítrio, então ela poderia ter feito de outra forma (A-C).
2. Se o determinismo é verdadeiro, ninguém pode fazer o contrário do que faz (D-E).
3. Portanto, se o determinismo é verdadeiro, ninguém age segundo seu livre arbítrio (F).

Chame esse argumento simplificado de argumento incompatibilista clássico. De acordo com o argumento, se o determinismo é verdadeiro, ninguém tem acesso a alternativas na forma exigida pelo modelo dos jardins de caminhos que se bifurcam, ou seja, ninguém tem ou já teve uma escolha sobre qualquer aspecto do futuro, inclusive o que normalmente consideramos ser nossas ações livres. Assim, se o determinismo é verdadeiro, segundo essa argumentação, não temos livre arbítrio.

B. Argumento da Origem

O segundo conjunto geral de argumentos para a incompatibilidade do livre arbítrio e determinismo se baseia na importância da fonte de uma vontade para o livre arbítrio. Começaremos com uma apresentação informal do argumento antes de considerar uma apresentação formal do mesmo. De acordo com essa linha de pensamento, um agente tem livre arbítrio quando suas vontades emanam de um agente de uma forma em particular (por exemplo,

de suas crenças, desejos, etc.). O que é importante para o livre arbítrio, como alegam os defensores desse argumento, não é simplesmente que a cadeia causal das vontades de um agente passa por esse agente, mas que *se origina* com o agente. Em outras palavras, um agente age com livre arbítrio somente se suas ações se originam dele, ou seja, se ele é a fonte última ou *causa primeira* de sua ação (Kane, 1998; Vihvelin, 2011).

Consideremos, novamente, a afirmação de que o livre arbítrio é uma condição necessária para a responsabilidade moral. Assim como a ponderação sobre casos de coerção e manipulação nos sugere, mesmo que um agente coagido ou manipulado esteja agindo segundo suas crenças e desejos, isso não é suficiente para que seja considerado moralmente responsável por seus atos. Normalmente assumimos que a coerção e certas formas de manipulação minam a responsabilidade moral de um agente precisamente porque um agente coagido ou manipulado não é o criador, ou seja, não é a origem de sua ação coagida. Se Alice é coagida a passear com seu cachorro via lavagem cerebral, então a ação de ter ido passear com o cachorro é originária da lavagem cerebral, e não de Alice. Consideremos, a partir disso, as semelhanças entre os casos de coerção e manipulação por um lado e as implicações da verdade do determinismo, por outro. Se o determinismo for verdadeiro, talvez seja verdade que Alice decida caminhar com o cachorro por causa de suas crenças e desejos, mas essas crenças e desejos elas mesmas são produtos inevitáveis de cadeias causais que não podemos traçar um começo ou origem. Segundo os defensores desse tipo de argumento para o incompatibilismo, a verdade do determinismo significaria que os agentes não causam suas ações da maneira necessária para o livre arbítrio e, em última instância, a responsabilidade moral (Timpe, 2006).

Podemos representar uma versão formal do argumento, conforme o modelo apresentado por Vihvelin (2011), o chamado “argumento da Origem”, da seguinte forma:

1. Um agente age com vontade livre se e somente se ele é o criador (ou melhor, fonte) de suas ações.
2. Se o determinismo é verdadeiro, então tudo o que qualquer agente faz é, em última análise, causado por eventos e circunstâncias fora de seu controle.
3. Se tudo o que um agente faz é, em última análise, causado por eventos e circunstâncias além de seu controle, então o agente não é o criador (ou melhor, fonte) de suas ações.
4. Portanto, se o determinismo é verdadeiro, nenhum agente é o autor (ou melhor, fonte) de suas ações.
5. Portanto, se o determinismo é verdadeiro, nenhum agente tem livre arbítrio.

O argumento da origem parece ser válido. Para avaliar sua solidez, é preciso avaliar a verdade de suas três premissas. A premissa 3 é claramente verdadeira, uma vez que para que um agente seja um criador ele não pode ser, em última análise, determinado por qualquer coisa fora de si mesmo. A premissa 2 do argumento também é verdadeira dada a definição de determinismo. Para rejeitar a conclusão do argumento, deve-se, portanto, rejeitar a premissa 1.

Anteriormente vimos uma consideração sobre o livre arbítrio que negava implicitamente a premissa 1, qual seja, o modelo hierárquico do livre arbítrio. De acordo com este modelo, um agente age com livre arbítrio desde que a cadeia causal para sua ação passe por desejos de 1ª e de 2ª ordem do agente. Uma forma de enfatizar a necessidade de originação diante de tal modelo hierárquico é adotar a visão da causação do agente. Se a premissa 1 é verdadeira, então a vontade do agente não pode ser o produto de uma cadeia causal determinista que está para além do agente. Que outras alternativas poderiam ser pensadas? Duas outras alternativas seriam: ou volições são sem causa ou são causadas indeterministicamente. É difícil ver como um agente poderia ser o autor ou fonte última de volições se volições fossem sem causa. Da mesma forma, por razões que vimos acima quando discutimos a visão pessimista do livre arbítrio, parece que a causalidade indeterminista minaria, ao invés de promover, o controle de um agente sobre suas vontades. Por estas razões, alguns incompatibilistas defendem olhar para a causa envolvida em volições sob uma nova luz. Em vez de manter uma vontade que é causada por um evento anterior (seja determinística ou indeterministicamente), estes incompatibilistas defendem que volições são causadas diretamente pelos agentes (O'Connor, 2000). Nesse ponto de vista, podemos encontrar uma distinção entre dois tipos diferentes e irreduzíveis de causalidade: causalidade de evento (*event-causation*) e agente de causalidade (*agent-causation*), estando o último o único envolvido no livre arbítrio. Os defensores do agente de causalidade propõem que agentes são substâncias duradouras que possuem diretamente o poder de causar volições. Apesar de muitos filósofos questionam se é coerente falar de um agente de causação, se for coerente, então isso poderia fornecer um suporte para a premissa 1 do argumento da origem.

C. A relação entre os argumentos

De acordo com Timpe (2006), a forma que os argumentos de consequência e origem foram delineados acima podem sugerir que os dois tipos de argumentos são mais independentes um do outro do que realmente são. Um incompatibilista pode argumentar que os agentes originam suas ações na forma exigida pela premissa 1 do “argumento de origem” se e somente

se eles têm uma escolha sobre suas ações na forma sugerida pelo “argumento de consequência”. Em outras palavras, se as minhas futuras vontades não fossem do tipo em que eu pudesse ter uma escolha sobre, então eu não seria a origem dessas vontades. E, como os argumentos acima afirmam, a verdade do determinismo ameaça tanto o nosso controle sobre nossas ações e vontades quanto nossa capacidade de originar essas mesmas ações e vontades. Porque, se o determinismo causal é verdadeiro, então o passado, conjuntamente com as leis da natureza, é suficiente para cada volição que um agente tem ou possui, e as cadeias causais que levam a essas volições não começariam do próprio agente. Assim, a maioria dos incompatibilistas pensam que ter uma escolha e ser auto-determinante são duas coisas que caminham lado a lado. Robert Kane (1998; 2014), por exemplo, argumenta que, se os agentes têm “responsabilidade final” (seu termo para o que é aqui chamado de “originação” ou “auto-determinação”), então esses agentes também terão possibilidades alternativas abertas para ele. De acordo com essa linha de argumentação, o poder de causar suas próprias ações não é um poder distinto do poder de escolher e de fazer o contrário. Sendo assim, os dois tipos de argumentos para o incompatibilismo podem ser simplesmente dois lados da mesma moeda.

1.3.4.2.2. Argumentos a favor do Compatibilismo

Tendo situado representantes dos dois argumentos mais importantes a favor do incompatibilismo, vamos agora considerar os argumentos a favor do compatibilismo. Ao considerar esses argumentos, é bem útil valer-se dos argumentos a favor do incompatibilismo para abordá-los e, assim, começaremos por considerar as maneiras como os compatibilistas responderam aos argumentos apresentados anteriormente.

A. Rejeitando os argumentos incompatibilistas

Como observado acima, o argumento da origem para o incompatibilismo parece ser válido, sendo que a única maneira do compatibilista rejeitar a conclusão do argumento seria rejeitando a sua primeira premissa. Ou seja, dada a definição de determinismo, compatibilistas devem rejeitar que o livre arbítrio requer que um agente seja o autor ou fonte última de suas ações. Mas como isso pode ser feito? Na maioria das vezes, compatibilistas propõem uma rejeição à “*ultimacy condition*” de livre e espontânea vontade, apelando para uma visão hierárquica ou “responsiva a razões” da vontade, por exemplo, respectivamente, Frankfurt

(1971) e Fischer e Ravizza (1998). Se tudo o que é necessário para o livre-arbítrio é uma certa “malha” entre volições de 1ª ordem e desejos de 2ª ordem, então não é necessária essa consideração de que um agente seja o originador desses desejos. Além disso, uma vez que a verdade do determinismo não implica que os agentes não tenham desejos e vontades de primeira e segunda ordem, uma consideração hierárquica da vontade é compatível com a verdade do determinismo. Da mesma forma, um agente pode ter livre arbítrio se ele tem o nível necessário de “responsividade a razões” de tal forma que ele teria agido de forma diferente se tivesse tido razões diferentes. Desta maneira, a “ultimacy condition” novamente não é obrigatória. Assim, se são adotadas determinadas considerações da vontade, tem-se razões válidas para rejeitar a premissa central do “argumento da origem”.

B. Argumento de Frankfurt contra o "princípio de possibilidades alternativas"

Em seu artigo de 1969, *“Moral Responsibility and Alternate Possibilities”*, Harry Frankfurt desenvolveu um argumento que possibilitou aos compatibilistas utilizarem-se de uma estratégia alternativa para atacar a primeira premissa do clássico argumento incompatibilista (se uma pessoa age por sua própria vontade, pressupõe-se que ela poderia ter agido de outra forma). Temos aqui, então, o argumento de Frankfurt dirigido contra o princípio de possibilidades alternativas: “uma pessoa é moralmente responsável pelo que ela faz apenas se ela pode fazer o contrário”.

O Ponto central ao ataque de Frankfurt a esse princípio é um exemplo dado por ele onde um agente é moralmente responsável apesar de, no momento da ação pertinente, não poder agir de outra forma, ou fazer o contrário. Um resumo aproximado para o exemplo que Frankfurt apresentou em seu artigo original seria:

Jones decidiu atirar em Smith. Black descobriu o plano de Jones e resolveu ajudá-lo a realizar o plano. Porém, Black preferiria que Jones atirasse em Smith por conta própria. No entanto, preocupado que Jones pudesse vacilar em sua determinação de atirar em Smith, Black secretamente arruma as coisas de modo que, se Jones mostrasse algum sinal de que não iria atirar em Smith (pressupondo que Black tivesse todos os recursos necessários para detectar tais sinais), Black poderia manipular Jones de tal forma que Jones iria atirar em Smith. No decorrer da ação, Jones segue com seus planos e atira em Smith por suas próprias razões. Ninguém de nenhuma forma ameaçou ou coagiu Jones, ofereceu suborno, ou mesmo sugeriu que ele atirasse em Smith. Jones disparou em Smith por sua própria intenção. Black nunca interveio.

Neste exemplo, Jones atirou em Smith por sua própria conta, e fê-lo sem nenhum ônus ou impedimento – o fez livremente. Mas, dada a presença de Black no cenário, Jones não poderia ter agido de outra maneira a não ser atirando em Smith. Assim, temos um contra exemplo para o princípio de possibilidades alternativas. De acordo com McKenna (2009), esse argumento fez com que os compatibilistas tivessem de buscar um outro modelo como base intuitiva para o tipo de controle pertinente à *agency* moralmente responsável. Pois, se o argumento de Frankfurt está correto, muito do debate em torno do livre arbítrio tem sido mal posto através de grande parte da história da filosofia. Se o determinismo ameaça o livre arbítrio e a responsabilidade moral, não é porque ele seja incompatível com um tipo de liberdade que envolva a capacidade de fazer o contrário e, mesmo que seja, não é do tipo de liberdade necessária para a responsabilidade moral. Independentemente disso, o que é mais relevante é que o argumento de Frankfurt vem inspirando muitos compatibilistas a pensarem novas considerações da liberdade ou de controle que se desviam, ou evitam propositalmente, por exemplo, de modelos como o dos jardins de caminhos que se bifurcam.

C. Atitudes reativas de Strawson

Para Peter Strawson (1982) muito dos debates tradicionais entre compatibilistas e incompatibilistas estão equivocados, pois, segundo ele, deveríamos nos focar mais no que ele chama de *atitudes reativas* – são atitudes que temos em relação às outras pessoas, baseando-nos em suas atitudes diante a como a tratamos. Strawson diz que a característica das atitudes reativas é que elas são “reações humanas essencialmente naturais a boa ou má vontade ou indiferença dos outros em relação a nós, conforme exibido em suas atitudes e ações” (p. 67). Exemplos de atitudes reativas incluem gratidão, ressentimento, perdão e amor. Portanto, são atitudes cruciais para as interações interpessoais e fornecem a base para a realização de indivíduos moralmente responsáveis. Strawson, então, vai utilizar-se desse ponto de partida para fazer duas argumentações. A primeira delas é que as atitudes reativas de um agente não seriam afetadas por uma crença de que o determinismo seja verdade ou não:

O compromisso humano à participação nas relações interpessoais comuns é, penso eu, profundamente enraizada para que levemos a sério a ideia de que uma convicção teórica geral possa mudar tanto o nosso mundo que, nele, não haveriam mais coisas como as inter-relações pessoais tais como normalmente a compreendemos... A objetividade sustentada de atitudes interpessoais, e o isolamento humano que isso implicaria, não parece ser algo

de que os seres humanos seriam capazes, mesmo se alguma verdade geral tivesse uma base teórica por isso (p. 68, tradução nossa)

Além disso, Strawson também defende uma reivindicação normativa: a verdade do determinismo não deve comprometer as nossas atitudes reativas. Ele acha que existem dois tipos de casos em que é conveniente suspender nossas atitudes reativas. Uma envolve agentes, tais como crianças ou deficientes mentais, que não são agentes morais, assim, segundo ele, não devemos ter atitudes reativas em relação a agentes não morais. O segundo tipo de caso em que é conveniente suspender nossas atitudes reativas são aquelas em que, enquanto o agente é um agente moral, suas ações em relação a nós não estão conectadas à sua agência no caminho correto. Por exemplo, enquanto eu poderia ter a atitude reativa de ressentimento em relação a alguém que esbarra em mim e me faz derramar minha cerveja, se eu descobrisse que essa pessoa foi empurrada em minha direção, eu não teria mais justificativa em me ressentir com essa pessoa. A verdade do determinismo, no entanto, não implica necessariamente nem que possam haver agentes que são agentes morais, nem que nenhuma das ações de um agente estão conectadas à sua agência moral. De acordo com Strawson, a verdade do determinismo não deve comprometer as nossas atitudes reativas. Desde que a responsabilidade moral é baseada nessas mesmas atitudes, Strawson pensa que a responsabilidade moral é compatível com a verdade do determinismo. Então, se o livre arbítrio é um requisito para a responsabilidade moral, o argumento de Strawson dá suporte para um compatibilismo.

1.4. Considerações sobre o capítulo

Como explicitado acima, podem ser encontrados diversos significados para o conceito de vontade e, ao mesmo tempo, vários outros conceitos podem estar interligados a esta discussão, como a questão da responsabilidade moral, liberdade de ação, determinismo e livre arbítrio, compatibilismo e incompatibilismo. Além disso, é um debate que está em aberto e vem apresentando renovada efervescência, com grande número de publicações relacionadas ao assunto na atualidade, principalmente no campo da filosofia. Já não bastasse o enorme volume de tratados sobre a temática, que, de acordo com Kane (2005), seria o debate mais volumoso dentre todos os problemas filosóficos em toda a história da filosofia e, como apontado por David Hume (2003), a questão mais controversa da metafísica e das ciências. Assim, dada a complexidade e extensão do tema, nos foi possível apresentar apenas essa visão mais geral, a fim de contextualização, para entrarmos, então, em uma investigação de como o assunto é

tratado na psicologia cognitiva e o que ela pode acrescentar ao debate, dada sua especificidade investigativa.

CAPÍTULO II

PSICOLOGIA COGNITIVA: FUNDAMENTOS FILOSÓFICOS

É importante, antes de começarmos a definir melhor o que iremos chamar aqui de psicologia cognitiva, nos precaver contra um equívoco historiográfico bastante comum na literatura. Conforme podemos ver em Castañon, Justi e Araujo (2014), com o grande sucesso da obra de Thomas Kuhn, *A Estrutura das Revoluções Científicas* (Kuhn, 1962), as ideias kuhnianas passaram a ser uma referência muito forte na tentativa de se interpretar o estatuto da psicologia como ciência. A partir disso, foi surgindo a ideia de que a psicologia estava vivendo uma revolução, nos moldes do que Kuhn chamou de revolução científica, como um processo de mudança de um paradigma a outro, onde ocorreria, a partir de uma crise no paradigma vigente, uma nova concepção a respeito de um determinado fenômeno. Implicaria, então, a derrubada dramática de algo anteriormente dominante, seja política ou intelectualmente. No caso da psicologia, “revolução cognitiva”, como veio a ser conhecida (Baars, 1986; Gardner, 1996; Miller, 2003), referir-se-ia a derrubada do behaviorismo, que supostamente teria dominado a psicologia no início do século 20, e onde se diz que os estudos dos processos cognitivos teriam sido largamente ignorados. Porém, as aplicações das ideias de Kuhn à psicologia, segundo diversos historiadores (Driver-Linn, 2003; Leahey, 1992; O’Donohue, 1993), foram feitas de uma forma superficial e em algumas vezes equivocada, o que gerou não só uma simplificação do cenário em suas múltiplas complexidades, como também acarretou em uma distorção histórica que vem sendo perpetuada em muitos manuais de psicologia.

Embora o behaviorismo tenha tido realmente muitos adeptos da década de 1920 em diante, outras abordagens (como por exemplo a gestalt e a psicanálise) floresceram ao lado dele. Nas palavras de Hobbs (Hobbs & Burman, 2009), não foi um movimento monolítico: como testemunham as teorias de aprendizagem rivais de Guthrie, Tolman e Hull. Lovie (1983) e Dewsbury (2000) mostram que a investigação sobre a atenção e cognição em animais eram comuns na época. E, mesmo sendo verdade que desde os anos 1950 muitos psicólogos passaram a ver a si mesmos como “cognitivistas”, isso não pode razoavelmente ser chamado de uma “revolução” no sentido kuhniano já que continuam a haver diferentes perspectivas teóricas que florescem lado a lado dentro da psicologia. Sendo uma das abordagens “não-cognitivistas” mais relevantes neste contexto o próprio behaviorismo radical, um movimento que tem crescido desde os anos 1950, mesmo depois que supostamente a revolução cognitiva ocorreu.

Behavioristas radicais continuam a existir em relativo isolamento de outras abordagens psicológicas, como atestam Coleman & Mehlman (1992).

Leahey (1992), então, por esses motivos, chamou a revolução na psicologia americana de “mítica”. E, se perguntando por que alguns psicólogos espalhariam mitos sobre sua própria área de pesquisa científica, Hobbs (Hobbs & Burman, 2009) busca nos oferecer algumas hipóteses a esse respeito. Segundo ele, muitos de nós gostaríamos de um paradigma unificado acordado para pesquisa e teorização em psicologia. Porém, como isso não ocorreu, se referir a uma “revolução cognitiva” poderia permitir que os proponentes do cognitivismo fingissem que é deles tal paradigma dominante. Mesmo que não isso não tenha sido planejado ou feito de má-fé, já que muitos nesse grupo possam realmente acreditar nessa dita revolução devido a seu envolvimento pessoal no projeto. Algumas pessoas têm mitos de origem que são de grande importância, mesmo que simbolicamente, para a existência de sua tribo particular. Assim como Samelson (1986) argumentou que apresentar Comte como o fundador da psicologia social seria um tal mito de origem, talvez a noção de uma revolução cognitiva seja igualmente um mito da mesma natureza.

No entanto, muitos autores têm argumentado que algo definitivamente aconteceu, mesmo que isso satisfaça ou não os critérios de Kuhn. Por exemplo, Greenwood (1999) concluiu que a mudança que rotulamos como uma “revolução” ocorreu como resultado da substituição de variáveis operacionalizadas (descrições behavioristas) para construções hipotéticas (causas cognitivas). Mandler (2002) sugeriu que a “revolução” era realmente “um retorno” a atitudes e tendências de pré-behavioristas, apoiado por importações provenientes da psicologia da Alemanha, Grã-Bretanha, França e Suíça (incluindo Piaget). Para Burman (Hobbs & Burman, 2009), obviamente nada disso resolve os problemas historiográficos subjacentes, apenas mostra que a questão em si é mais complicada do que a crença bipolar de que houve ou não houve uma tal revolução nos moldes kuhnianos. Porém, é inegável que há algo mais na posição cognitivista do que um grito unificador, “conosco ou contra nós!”. Parece existir, realmente, um projeto razoavelmente bem estabelecido. Desta forma, o termo “revolução” nesse sentido pode ser algo favorável, na medida em que chama a atenção para uma mudança na meta-teoria dominante, orientando assim o que significava “fazer ciência” em um lugar e tempo específicos (no contexto norte-americano da época) e, portanto, destaca mudanças nas implicações para os futuros tradutores (e intérpretes) dos significados dos resultados dessa ciência. Seu uso também oferece oportunidades para fazer histórias das disciplinas (por

exemplo, histórias sociais e institucionais, análises de influência, etc.), ao invés de apenas histórias de fatos psicológicos descobertos em um vácuo social e político.

Para os propósitos do presente trabalho, apesar da controvérsia em torno do termo “revolução cognitiva”, este ainda, como vimos, pode ser útil para descrever a ocorrência de um momento importante na psicologia, mesmo que tal termo tenha apenas um sentido mais retórico e cerimonial do que o de efetivamente demarcar um fenômeno histórico preciso. No mais, efetivamente, a partir desse momento, houve uma aproximação maior de alguns pesquisadores da psicologia em torno de uma ideia geral e mesmo um projeto do que seria uma psicologia científica, principalmente entendida como uma nova abordagem para a investigação da cognição baseada na ideia de processamento de informação (Boden, 2006; Gardner, 1996; Miller, 2003).

Outro ponto importante, que merece esclarecimentos, é o reconhecimento de que nem sempre houve consenso ou mesmo uma maior precisão na definição do conceito de cognição. Assim, o adjetivo “cognitivo” muitas das vezes pode ser utilizado de uma forma mais vaga e abrangente, em geral para designar quaisquer psicologia que defina a mente como objeto de estudo, simplesmente se contraponto ao behaviorismo (Green, 1996). Desta maneira, “mente” e “cognição” podem ser vistos quase que como sinônimos e poder-se-ia dizer que toda psicologia mentalista pertenceria também à psicologia cognitiva. Seguiremos aqui a delimitação apresentada por Castañon, Justi e Araujo (2014) e, a fim de se evitar alguma confusão conceitual, o que tratamos como “psicologia cognitiva” deve ser entendido somente como uma abordagem psicológica bem específica onde o termo cognição é visto como um sistema de regras de processamento de informação e, desta forma, apresenta-se como uma ponto de vista particular do que seja a mente humana. Esta abordagem teria então como representantes o movimento que se formou a partir de um momento específico no tempo e que ficou conhecido como “revolução cognitiva”.

2.1. Psicologia cognitiva como parte do projeto de uma ciência da cognição

Essa abordagem que define a cognição como objeto de estudo podemos dizer que faz parte de um projeto mais amplo, abarcando não só a psicologia como também várias outras áreas de estudo, como por exemplo a inteligência artificial, filosofia da mente, linguística, neurociências, antropologia, etc., e foi associada ao surgimento da assim chamada *ciência cognitiva*. Como foi apontado por Boden (2006), uma das maneiras de se entender a proposta

do que pretende a ciência cognitiva é dar alguns exemplos das amplas questões estudadas por ela. Essas questões envolvem, principalmente, o enigma da mente humana e estão centradas, por conseguinte, na natureza da mente e no problema mente-corpo; a evolução da mente; liberdade e propósito; assim como seriam possíveis as diversas psicopatologias. A maioria dos tópicos estudados na ciência cognitiva cai em uma dessas categorias mais gerais. E aquelas que não o fazem, tais como a natureza da computação, estão intimamente relacionadas a estas questões levantadas. Um importante campo de estudo assinalado é a consciência. Para a ciência cognitiva pode-se dizer que existem fortes correlações entre os eventos cerebrais e estados conscientes, mas por que é assim? A resposta usual parece ser que nossos cérebros geram nossa consciência. Mas como é que eles fazem isso, na prática? E, ainda mais intrigante, como eles podem fazer isso, em princípio?

Que poderes mentais o cérebro humano fornece, e como ele consegue fazer isso? Como é possível o livre arbítrio? E a criatividade? São as ideias criativas imprevisíveis e se assim são, por quê? O que são as emoções – elas entram em conflito com a racionalidade, ou a fornecem apoio? Nossas habilidades são inatas, ou determinadas pela experiência? (Boden, 2006, p. 2, tradução nossa)

Nenhuma dessas questões são novas. Tendo sido originalmente levantadas e amplamente discutidas por filósofos durante toda a história da filosofia. Algumas ainda são exaustivamente discutidas, como a diferença entre a mente e o cérebro, por exemplo, e o livre arbítrio, como vimos no capítulo anterior. A ciência cognitiva, portanto, busca considerar essas questões à luz dos dados científicos disponíveis e investigar através dos métodos peculiares a cada uma de suas disciplinas. Obviamente, essas mesmas questões também foram anteriormente estudadas pela pesquisa científica tradicional em psicologia, antropologia, neurofisiologia, ou biologia, etc. Porém, segundo Boden (2006), “desde os anos 1940, no entanto, cada uma [dessas questões] tem sido ainda mais aguçadas pelos trabalhos em ciência cognitiva” (p.3).

Como vimos, a psicologia cognitiva faz parte de um grupo de disciplinas dentro de um projeto maior de ciência cognitiva. Penna (1984), em sua *Introdução à psicologia cognitiva*, define a psicologia cognitiva tanto como um movimento doutrinário, movimento esse compartilhado com a ciência cognitiva, quanto como uma área de pesquisa. Como observado por Baars (1986), existe uma ambiguidade na expressão “psicologia cognitiva”, essa podendo também se referir à metateoria que defende que por meio de observações empíricas podemos fazer a inferência de constructos teóricos não observáveis. Ambiguidade proveniente do fato de que esta metateoria teria surgido de uma disciplina também denominada como cognitiva,

pertencente ao tradicional campo de estudo da memória, percepção, etc. Assim, seguindo a denominação apresentada em Castañon (2010), quando tivermos nos referindo a metateoria, abordagem ou movimento, usaremos o termo cognitivismo; e, ao utilizarmos a expressão psicologia cognitiva, estaremos nos referindo a esta área de pesquisa específica. Dando prosseguimento, ainda segundo Castañon:

A psicologia cognitiva como área de pesquisa bem estabelecida foi um produto do cognitivismo, mas, ao mesmo tempo, seu sucesso foi o grande motor desse movimento, sendo ambos, portanto, indissociáveis. Como afirma Baars (1986), o estudo da cognição humana provê o domínio empírico no qual o sucesso ou o fracasso da metateoria cognitiva (cognitivismo) pode ser demonstrado. (...) Apesar de ser disciplina central na articulação desses esforços, a psicologia cognitiva não perde a especificidade de seu domínio ou de seus métodos. (p. 234)

Já apresentamos alguns questionamentos que são temas importantes de interesse da pesquisa em ciência cognitiva. Uma outra questão que deve ser levantada agora é como, ou qual método, seria utilizado para se responder, ou investigar, tais questões. De acordo com Boden (2006), A ciência cognitiva tenta responder a essas perguntas de duas formas, que estão intimamente relacionadas. Ambas se relacionam ao conceito de máquinas. Especificamente, a ciência cognitiva utiliza de conceitos abstratos (lógicos/matemáticos) extraídos da *inteligência artificial* (A.I.) e da *teoria do controle* (*control theory*), mais conhecida como cibernética. A A.I. tenta fazer com que computadores façam os tipos de coisas que a mente poderia fazer. Essas coisas podem variar entre interpretar a linguagem ou o *input* de uma câmera, passando por fazer um diagnóstico médico ou a construção de mundos imaginários (virtuais), até o controle dos movimentos de um robô. Já a teoria do controle analisa o funcionamento de sistemas de auto-regulação. Estes sistemas incluem tanto fábricas automáticas de produtos químicos como células e organismos vivos.

Esses conceitos (de computação e controle), segundo o cognitivismo, podem aguçar questões psicológicas já que eles podem expressar ideias sobre processos mentais mais claramente do que conceitos verbais poderiam: ou seja, utilizam-se da linguagem da computação ou teoria da informação para descrever processos mentais. Além disso, quando implementados em modelos de computador, poderiam testar a coerência e as implicações dessas ideias de uma forma mais rigorosa. Essa forma de se fazer ciência pode ser remetida a tradição do *verum factum*, ou “conhecimento do fabricante”. Essa tradição sustenta que a melhor forma de se compreender algo adequadamente seria se a pessoa estivesse habilitada a fazer ou construir aquilo que se quer compreender. Em outras palavras, observação e argumentação abstrata não seriam suficientes.

Um dos principais proponentes da *verum factum* era Giambattista Vico (1668-1744), famoso por argumentar que apenas as ciências humanas podem nos fornecer conhecimento genuíno, porque nelas estudamos as criações, não de Deus, mas dos próprios seres humanos. Especificamente, a história, que para Vico é a chave para a compreensão da mente e cultura humana, envolve uma recriação ativa dos pensamentos dos povos que estão sendo estudados. Com sua declaração do princípio de que *verum et factum convertuntur*, onde “aquilo que é verdadeiro e o que é feito são conversíveis”, ou *verum esse ipsum factum*, onde “verdade é precisamente o que é feito”; Vico enfatiza que a ciência deve ser concebida como o gênero ou modo como algo é feito ou construído. A verdade, portanto, se determina como o reconhecimento interior das causas daquilo que nós mesmos produzimos. Como não produzimos o mundo natural, porque não produzimos interiormente suas causas; nos resta, diante da natureza, nos limitarmos ao registro do ocorrido ou nos restringirmos a uma descrição anatômica do mundo natural. Já o mundo humano, ao contrário do natural, é fruto direto das vontades e ações humanas e possui seus mecanismos de causalidade dentro do próprio homem. Dado que “a norma da verdade é ter feito aquilo”, Vico argumenta que o famoso primeiro princípio de Descartes que diz que as ideias claras e distintas são a fonte da verdade deve ser rejeitado: “Para a mente não fazer a si mesma conforme começa a conhecer-se”¹, Vico observa, “e desde que não faz a si mesma, ela não sabe o gênero ou modo pelo qual ela se faz”. Assim, para Vico, as verdades da moral, ciências naturais e matemática não necessitam de justificação metafísica, como os cartesianos sustentavam, mas exigem antes uma análise das causas – a “atividade” – através das quais as coisas são feitas (Costelloe, 2014; Silva Neto, 2001).

Boden (2006) então nos lembra que, assim como os filósofos românticos do século XVIII e XIX usaram argumentos essencialmente semelhantes para priorizar a arte sobre a ciência, outros aplicaram o *verum factum* às ciências naturais também. Para muitos dos primeiros cientistas modernos um experimento científico foi visto como uma construção, e tecnologia baseada em teoria era uma justificativa intelectual dessa nova ciência. Em suma: se você pode construí-lo, você pode compreendê-lo. Muitos cientistas cognitivos concordariam e suas construções incluem não apenas teorias, mas também modelos de computadores que simulam, ou tentam reconstruir, processos cognitivos humanos.

¹ “For the mind does not make itself as it gets to know itself, and since it does not make itself, it does not know the genus or mode by which it makes itself” (Vico, DA, 52)

Dando prosseguimento, considerando o escopo da ciência cognitiva, para Boden (2006), ela pode ser vista como um campo universal de três maneiras: Em primeiro lugar, porque abrange todos os aspectos da mente e do comportamento (como foi ilustrado pelo grande número de questões acima listadas.). Em segundo lugar, baseia-se em muitas disciplinas diferentes que a estudam. E em terceiro lugar, ela se baseia em mais de um tipo de teoria. Em termos gerais, Boden defende que é o estudo da mente como máquina – uma definição que abrange vários tipos de explicação, como veremos mais adiante.

Na verdade, portanto, a ciência cognitiva se propõe a lidar com todos os tipos de processos mentais. Cognição (linguagem, memória, percepção, resolução de problemas...) estão incluídas, é claro. Mas também motivação, emoção e interação social – e o controle da ação motora, que é em grande medida para o que a cognição se desenvolveu, evolutivamente falando (Boden, 2006). Esses tipos de processos psicológicos, por outro lado, não são claramente distinguíveis. A crença holística de que esses processos estão intimamente entrelaçados, não é algo novo.

Dado que a ciência cognitiva não está focada apenas na cognição, o rótulo pode ser altamente enganoso. Por que, então, foram estas palavras escolhidas em primeiro lugar? George Miller, um dos principais responsáveis para que essas palavras tenham sido popularizadas, e se firmado como nome oficial do campo, deu a seguinte resposta para essa pergunta:

Olhando de volta para a palavra “cognição”, eu não acho que alguém intencionalmente estava excluindo “volição” ou “conação” [aka motivação] ou “emoção”. Eu acho que estavam apenas olhando de volta para o senso comum. Ao usar a palavra “cognição”, estávamos definindo-nos fora do behaviorismo. Nós queríamos algo que remetesse ao mental – mas, “psicologia mental” soava terrivelmente redundante. (Miller 1986, p.210, tradução nossa)

Em suma, eles pretendiam que uma ciência cognitiva abordasse não só a cognição como mais além. Um olhar através de *Plans and the Structure of Behavior* (Miller, G., Galanter, E., & Pribram, K., 1960) confirma isso. Nesse livro, considerado com um dos marcos fundadores do movimento, foi abordado não só o comportamento animal, instinto e aprendizagem, como também a memória humana, a linguagem, resolução de problemas, a personalidade, a doença mental, e hipnose. Questões sociais e culturais também foram abordadas. Nenhum aspecto da vida mental foi excluído. Segundo Boden (2006), praticamente todos os pais fundadores da ciência cognitiva (Noam Chomsky excetuado) tinham se perguntado como motivações e emoções interagem com a cognição, e vários tinham também mencionado psicopatologia. Todas estas questões eram frequentemente discutidas nos primeiros dias. Porém isso não durou muito. Motivação, emoção e interação social (seja em pequenos grupos ou em sociedades) são

ainda mais difíceis de estudar e simular do que a cognição é, e esses aspectos, por conseguinte, foram em breve colocados em segundo plano. Vastas quantidades de pesquisa foram feitas sobre a percepção, linguagem, resolução de problemas, conceitos, crenças, memória e aprendizagem. O nome do campo reflete bem isso.

De qualquer maneira, o nome ciência cognitiva acabou se estabelecendo, e é usado na maioria das vezes. Mesmo assim, os editores da recente *The MIT Encyclopedia of the Cognitive Sciences* (Wilson, R. A.; Keil, F. C., 1999) escolheram a versão no plural, no intuito de destacar o fato de que várias disciplinas muito diferentes estão envolvidas no campo. Porém, a forma singular, por outro lado, destaca os laços intelectuais entre esses diversos campos envolvidos. Porque, como veremos, tem havido inúmeros casos de trabalhos em uma disciplina que estão sendo radicalmente influenciados pelo trabalho de outras. Para buscar entender a mente (mente/cérebro) corretamente, não só necessita-se olhar para ela de todas as direções possíveis, mas é preciso também integrar esses diversos pontos de vista.

Como podemos atestar em Boden (2006), a história da ciência cognitiva é marcada por uma profunda e contínua interdisciplinaridade. Segundo ela, este é um relacionamento mais íntimo do que mera multidisciplinaridade intelectual. Uma ou outra vez, os investigadores em uma área tem emprestado ideias teóricas, e não apenas dados, de um ao outro. Podemos mesmo dizer que uma das ideias fundamentais, decisivas para a fundamentação da ciência cognitiva, e da psicologia cognitiva em si, foram os avanços na área da ciência da computação.

2.2. Contribuições da ciência da computação à psicologia cognitiva

Em 1936, o famoso matemático Alan Turing formaliza a noção de algoritmo, identificando as operações fundamentais que poderiam servir como base ao cálculo matemático. Desenvolveu, então, um dispositivo imaginário capaz de executar essas operações segundo regras bem definidas – a *Máquina de Turing* (Turing, 1937). Esta foi assim concebida para ser um modelo de computação, formalizando então um conjunto de operações básicas às quais se pode reduzir qualquer computação. A máquina de Turing seria portanto um autômato que possui tanto uma unidade de controle quanto um dispositivo especial que funcionaria simultaneamente como entrada (onde faz a leitura), armazenamento, e saída (onde escreve). O armazenamento seria feito em uma fita unidimensional que contém um número ilimitado de células cada uma das quais pode conter um único símbolo. A fita da Máquina de Turing prolonga-se indefinidamente e por isso pode conter teoricamente uma quantidade infinita de

informação. Esta informação pode ser lida e alterada em qualquer ordem e daí seu potencial. Associada a essa fita está uma cabeça de leitura-escrita que pode se mover sobre a fita para a esquerda ou para a direita, podendo escrever ou ler um único símbolo em cada movimento, podendo então, graças à unidade de controle, interpretar uma lista de instruções simples. A regra executada determina o que se convencionou chamar de estado da máquina. A ideia por trás do dispositivo era de se criar um aparelho que seria capaz de manipular símbolos em uma fita de acordo com uma série de regras para armazenar informações.

Sendo assim, o conceito de máquina de Turing é semelhante ao de uma fórmula ou equação. De tal modo, poder-se-ia pensar em uma infinidade de possíveis máquinas de Turing, cada uma correspondendo a um método definido ou algoritmo. Turing propôs que cada algoritmo, formalizado como um conjunto finito de instruções bem definidas, pudesse ser interpretado e executado por um processo mecânico. Na atualidade é fácil relacionar um programa de computador com uma máquina de Turing e a tarefa mecânica de interpretação e execução obedecendo a algoritmos. Desta maneira, a Máquina de Turing Universal incorpora o princípio essencial do computador: uma máquina simples que poderá executar qualquer tarefa bem definida, desde que especificada como um programa apropriado. Sendo Turing, por isso, considerado “pai” da computação.

Como vimos, as ideias de Turing, e uma outra tese, também relacionada diretamente a Máquina de Turing, que ficou conhecida como tese Church-Turing, foram de grande influência também para a psicologia cognitiva e ciência cognitiva como um todo. Porém, segundo Copeland (2015), um outro mito parece ter surgido a respeito do artigo de 1936 de Turing, qual seja, de que ele não deu um tratamento dos limites do mecanismo que desenvolveu, o que gerou, assim, uma crença infundada de que a máquina de Turing universal pode simular o comportamento de qualquer máquina. Esse mito teria passado para a filosofia da mente e gerado um efeito pernicioso. Vamos portanto tentar esclarecer alguns pontos, já que, má interpretada, esta questão pode influenciar uma visão relativa à relação mente/cérebro humanos, que por sua vez teria consequências relativas as interpretações que poderíamos dar sobre o problema central de interesse, a vontade humana. Inicialmente iremos tentar colocar de forma mais detalhada e clara do que se trata a tese Church-Turing.

Alonzo Church (1903-1995) foi um lógico matemático que desenvolveu, ao mesmo tempo que Turing, uma resposta ao problema da indecibilidade de Hilbert (*Entscheidungsproblem*), utilizando porém um caminho diferente, um sistema matemático formal que investiga funções e aplicações de funções e que ficou conhecido como *cálculo*

lambda. A partir de suas formulações, obteve resultados muito semelhantes aos de Turing. E, a partir dos avanços alcançados por ambos, chegou-se a formulação da tese Church-Turing. A tese diz respeito à noção de um método *efetivo* ou *mecânico* em lógica e matemática. De acordo com Copeland (2015), “efetivo” e “mecânico”, nesse contexto, são termos sinônimos que devem ser entendidos como jargões ou “termos da arte” destas disciplinas: eles não carregam o seu significado habitual. Assim entendido, um método ou procedimento, que aqui chamaremos de *M*, para alcançar algum resultado desejado é chamado de efetivo ou mecânico apenas no caso:

1 – se *M* é definido em termos de um número finito de instruções exatas (cada instrução está sendo expressa por meio de um número finito de símbolos);

2 – se *M* irá, levado a cabo sem erro, produzir o resultado desejado num número finito de passos;

3 – se *M* pode (na prática ou em princípio) ser realizado por um ser humano sem a ajuda de qualquer tipo de equipamento, apenas utilizando-se de papel, lápis e borracha;

4 – se *M* não exige nenhum *insight* por parte do ser humano na sua realização.

Um exemplo bem conhecido, segundo Copeland (2015), de um método eficaz é o teste da tabela-verdade para tautologia. Na prática, é claro, este teste é impraticável para as fórmulas que contêm um grande número de variáveis proposicionais, mas, em princípio, poder-se-ia aplicá-lo com sucesso para qualquer fórmula do cálculo proposicional, dado suficiente tempo, obstinação, papel, lápis e borracha. Declarações de que não existe um método efetivo para alcançar tal e tal resultado são comumente expressas dizendo que não há um método efetivo para a obtenção dos valores de uma função matemática tal e tal.

A noção de um método eficaz, portanto, é uma noção informal, e tentativas de caracterizar eficácia carecem de rigor principalmente para o requisito fundamental de que o método não exige nenhum insight ou processo criativo, visto que este ponto é deixado sem explicações mais precisas. De acordo com Copeland (2015), uma das conquistas de Turing em seu artigo de 1936 foi apresentar um predicado formalmente exato com o qual o predicado informal que pode ser calculado por meio de um método efetivo pôde ser substituído. Church fez o mesmo (1936a). A substituição dos predicados que Turing e Church propuseram foi, em face disso, muito diferentes um do outro, mas acabaram por ser equivalentes, no sentido de que cada um escolhe o mesmo conjunto de funções matemáticas. A tese de Church-Turing é a

afirmação de que este conjunto contém todas as funções cujos valores podem ser obtidos por um método que satisfaça as condições acima para eficácia. É claro que, se não houvesse funções das quais o predicado informal, mas não o predicado formal, fossem verdadeiras, então o último seria menos geral do que o anterior e por isso não poderia razoavelmente ser empregado para substituí-lo. Portanto, quando a tese é expressa em termos do conceito formal proposto por Turing, é considerado adequado também se referir à tese como “a tese de Turing”; e, *mutatis mutandis*, no caso de Church (Copeland, 2015).

O conceito formal proposto por Turing é o de computabilidade pela máquina de Turing. Ele defendeu a reivindicação (tese de Turing) de que sempre que houver um método eficaz para a obtenção dos valores de uma função matemática, a função pode ser computada por uma Máquina de Turing. A alegação inversa é facilmente estabelecida, um programa da máquina de Turing é por si só uma especificação de um método eficaz: sem exercer qualquer *insight*, um ser humano pode trabalhar com as instruções do programa e realizar as operações necessárias. Se a tese de Turing é correta, então, falar sobre a existência e não existência de métodos eficazes, pode ser substituída (em matemática e lógica) por falar sobre a existência ou não existência de programas da Máquina de Turing. Esta tese foi declarada por Turing em diversos lugares, com diferentes graus de rigor. A formulação que se segue é uma das mais acessíveis.

LCMs [máquinas de computação lógicas: expressão de Turing para máquinas de Turing] pode fazer qualquer coisa que poderia ser descrito como “regra de ouro” ou “puramente mecânico” (...) Isto é suficientemente bem estabelecido que agora está acordado entre os lógicos que “calculável por meio de um LCM” é a interpretação mais precisa e correta de tais frase. (Turing 1948, p. 7, tradução nossa).

Essa tese foi introduzida por Turing durante sua argumentação de que o *Entscheidungsproblem*, ou problema de decisão para o cálculo de predicados, é insolúvel. Eis as considerações de Church sobre o *Entscheidungsproblem*:

É aqui entendido por Entscheidungsproblem de um sistema de lógica simbólica o problema de encontrar um método efetivo através do qual, dada qualquer expressão Q na notação do sistema, pode-se determinar se Q é ou não demonstrável no sistema. (Church, 1936b, p. 41, tradução nossa)

Desta maneira, Turing demonstrou que, dada sua tese, não pode haver tal método para o cálculo de predicados. Ele provou formalmente que não existe uma máquina de Turing que possa determinar, em um número finito de passos, se ou não qualquer dada fórmula do cálculo de predicados é um teorema do cálculo. Assim, dada a sua tese de que, se existe um método efetivo, ele pode ser levado a cabo por uma das suas máquinas e, portanto, segue-se que não

existe tal método a ser encontrado para o referido problema. Em contrapartida, alguns meses antes, Church tinha alcançado os mesmos resultados negativos empregando o conceito de definibilidade por lambda no lugar de computável pela Máquina de Turing. Church e Turing haviam descoberto o resultado de forma totalmente independente uma da outra. Porém, o método de Turing se mostrou bem mais satisfatório do que o de Church, como o próprio Church reconheceu em uma revisão do trabalho de Turing: “computabilidade por uma máquina de Turing ... tem a vantagem de fazer a identificação com eficácia no (não explicitamente definido) sentido comum imediatamente evidente” (1937a, p 43).

Turing, depois de apreender a proposta de Church, rapidamente reconheceu que o conceito de definibilidade por lambda e seu próprio aparato de computabilidade seriam de fato equivalentes (1937, pp. 263). Assim, segundo Copeland (2015), na proposta de Church, as palavras “função recursiva de inteiros positivos” podem ser substituídas pela expressão “função de inteiros positivos computáveis por uma Máquina de Turing”. Post (1936), referiu-se à identificação de Church de calculabilidade efetiva com recursividade como uma “hipótese de trabalho” e, posteriormente, fez a seguinte crítica: “mascarar essa identificação como uma definição ... nos cega para a necessidade de sua contínua verificação” (p. 105).

Essa, portanto, é a “hipótese de trabalho” que Church propõe e que ficou conhecida como tese de Church: a função de inteiros positivos é efetivamente calculável apenas se recursiva. Logo, a implicação inversa, de que cada função recursiva de inteiros positivos é efetivamente calculável, é comumente referida como o inverso da tese de Church. E, se a atenção é restrita a funções de inteiros positivos, a tese de Church e tese de Turing são equivalentes. Já o termo “tese de Church-Turing” parece ter sido introduzido pela primeira vez por Kleene (1967): “Assim, as teses de Turing e de Church são equivalentes. E nós geralmente referimo-nos a ambas como tese de Church ou, em conexão com esta, uma versão que trata das “máquinas de Turing”, como a tese de Church-Turing” (p.232).

Muita evidência foi acumulada, como podemos ver em Kleene (1952), para a “hipótese de trabalho”, proposta por Church e Turing em 1936. Apresentaremos aqui um resumo dessas evidências: (1) Cada função efetivamente calculável que foi investigada a este respeito acabou por ser computável pela máquina de Turing. (2) Todos os métodos ou operações conhecidas para a obtenção de novas funções efetivamente calculáveis a partir de dadas funções efetivamente calculáveis são paralelos para métodos de construção de novas máquinas de Turing a partir da máquina de Turing original. (3) Todas as tentativas de dar uma análise exata da noção intuitiva de uma função efetivamente calculável acabaram por ser equivalentes, no

sentido de que em cada análise oferecida foi comprovado selecionar a mesma classe de funções, ou seja, aquelas que são computáveis pela máquina de Turing. Devido à diversidade das análises esta é geralmente considerada uma evidência forte em favor da tese aqui tratada. E, de acordo com Copeland (2015), embora de tempos em tempos tentativas tenham sido feitas para testar a tese de Church-Turing, o resumo da situação dado por Turing em 1948 não é menos verdadeiro hoje em dia: “é suficientemente bem estabelecido que agora está acordado entre os lógicos que ‘calculável por meio de um LCM’ é a interpretação mais precisa e correta” (da noção informal em questão).

Agora que já ficou mais ou menos bem estabelecido do que se trata a tese de Church-Turing, vamos buscar apresentar uma amostra de como ela é tratada na filosofia da mente por alguns autores (de grande influência na ciência e psicologia cognitiva) de forma a fazer parecer que a tese de Church-Turing seria uma comprovação de que o cérebro/mente funcionariam de forma análoga a um computador que, por sua vez, seria uma encarnação da Máquina de Turing. Por exemplo, no *Oxford Companion to the Mind* pode-se ler a seguinte afirmação: “Turing mostrou que sua máquina muito simples ... pode especificar os passos necessários para a solução de qualquer problema que pode ser resolvido por instruções, regras explicitamente declaradas ou procedimentos” (Gregory, 1987, p.784). Dennett defende que “Turing tinha provado – e esta é provavelmente sua maior contribuição – que sua máquina de Turing Universal pode calcular qualquer função que qualquer computador, com qualquer arquitetura, pode calcular” (1991, p. 215); defende ainda que “toda tarefa para a qual há uma receita clara composta de passos simples pode ser executada por um computador muito simples, uma máquina de Turing universal, o seguidor de receitas universal” (1978, p. XVIII). Paul e Patricia Churchland afirmam que “os resultados de Turing implicam algo notável, ou seja, que um computador digital padrão, dado apenas o programa certo, uma memória grande o suficiente e tempo suficiente, pode computar qualquer função *input-output* governada por regras. Ou seja, ele pode exibir qualquer padrão sistemático de respostas ao meio ambiente que seja” (1990, p. 32).

Estas várias citações são típicas da escrita atual sobre os fundamentos da teoria computacional da mente e, aparentemente, parece improvável que esses autores tenham a intenção de restringir as noções gerais de “regra explicitamente declarada”, “receita clara composta de passos simples”, “computador com qualquer arquitetura”, “função governada por regras” e “padrão sistemático de respostas”, de modo a aplicarem-se apenas às coisas que podem ser obedecidas, simuladas, calculadas, ou produzidas por uma máquina que implementa

métodos “eficazes” ou “mecânicos” no sentido original de Turing. Mas, como adverte Copeland (2015), a menos que essas noções estejam restritas desta forma desde o início, devemos rejeitá-las.

Para Copeland (2015), Turing não demonstrou que suas máquinas podem resolver qualquer problema que pode ser resolvidos “por instruções, regras explicitamente declaradas ou procedimentos”, nem provou que a máquina de Turing universal “pode calcular qualquer função que qualquer computador, com qualquer arquitetura, pode calcular”. Ele provou que sua máquina universal pode calcular qualquer função que qualquer máquina de Turing pode calcular; e apresentou e desenvolveu argumentos filosóficos em apoio a isso, a tese aqui chamada tese de Turing, como vimos acima. Mas uma tese sobre a extensão dos métodos eficazes – o que quer dizer, a respeito da extensão dos procedimentos de uma determinada espécie que um ser humano sem a ajuda de máquinas é capaz de executar – não carrega nenhuma implicação quanto à extensão dos procedimentos que as máquinas são capazes de realizar, até mesmo máquinas que agem de acordo com “regras explicitamente declaradas”. Por exemplo, dentro do repertório de uma máquina de operações atômicas, ou um computador quântico, podem haver operações que nenhum ser humano sem a ajuda de máquinas seria capaz de executar.

A proposição que se segue, muito diferente da própria tese de Turing, de que uma máquina de Turing pode calcular qualquer coisa que pode ser calculada por qualquer máquina trabalhando com dados finitos em conformidade com um programa finito de instruções, é por vezes também referida como uma versão da tese de Church-Turing ou tese de Church. Por exemplo, Smolensky diz: “modelos conexionistas ... possivelmente podem até desafiar a forte interpretação da tese de Church como a alegação de que a classe de computações bem definidas se esgotam por aquelas das máquinas de Turing” (1988, p. 3).

Segundo Copeland (2015), há vários exemplos do que ele chama “afrouxamento da terminologia estabelecida”, e nem Church nem Turing endossariam tal afrouxamento, muito menos o teriam formulado. Entre outros exemplos desse afrouxamento da terminologia na literatura, os seguintes são típicos:

Que existe uma formulação mais geral da máquina e que ela conduz a um conjunto exclusivo de funções de input-output passou a ser chamada tese de Church. (Newell, 1980, p. 150; apud Copeland, 2015, tradução nossa)

É difícil ver como qualquer linguagem que poderia realmente ser executada em um computador

físico poderia fazer mais do que Fortran² pode fazer. A ideia de que não há tal linguagem é chamada tese de Church. (Geroch & Hartle, 1986, p. 539; apud Copeland, 2015, tradução nossa)

Ou, mais distante ainda de qualquer coisa que Church ou Turing realmente escreveram:

Agora, posso afirmar a versão física do princípio de Church-Turing: “Cada sistema físico realizável finitamente pode ser perfeitamente simulado por um modelo de máquina de computação universal que opera por meios finitos”. Esta formulação é tanto mais bem definida e mais física do que a própria maneira de Turing de expressá-la. (Deutsch, 1985, p. 99; apud Copeland, 2015, tradução nossa)

Gandy (1980) é um dos poucos escritores a distinguir explicitamente entre a tese de Turing e a proposição mais forte de que tudo o que pode ser calculado por uma máquina pode ser calculado por uma máquina de Turing. Tomando emprestado a terminologia de Gandy, iremos chamar a proposição mais forte de “Tese M”. E iremos utilizar expressões como “a tese de Church-Turing propriamente dita” para a proposição que Church e Turing realmente defenderam.

Tese M – o que quer que possa ser calculado por uma máquina (trabalhando com dados finitos em conformidade com um programa finito de instruções) é computável por uma máquina de Turing.

A Tese M em si admite duas interpretações, de acordo com se a frase “pode ser gerada por uma máquina” é tomada em um sentido mais estrito, de que “pode ser gerada por um computador que está em conformidade com as leis físicas (se não houver limitações de recursos) do mundo real”, ou em um sentido mais amplo, onde a máquina imaginária em questão existe em qualquer mundo possível. Sob a última interpretação, a tese M é falsa. É simples descrever máquinas fictícias, ou *hypercomputers* (Copeland & Proudfoot, 1999a) que geram funções não computáveis pela máquina de Turing (ver por exemplo, Copeland, 2000a; Copeland e Proudfoot, 2000; e Stewart, 1991). E, ainda é uma questão empírica em aberto se a versão mais estrita da tese M é verdadeira ou não. Como mostramos mais acima, é apenas uma hipótese de trabalho que necessita ser constantemente averiguada. A especulação de que pode haver processos físicos cujo comportamento está de acordo com funções não computáveis por uma máquina de Turing remonta pelo menos ao período ao longo das últimas cinco décadas (Copeland & Proudfoot, 1999b).

² Família de linguagens de programação desenvolvida pela IBM a partir da década de 1950 e que continua a ser utilizada hoje em dia.

Assim, segundo Copeland (2015), o erro de confundir a tese de Church-Turing propriamente dita com a tese M levou a algumas reivindicações notáveis nos fundamentos da psicologia cognitiva. Por exemplo, é frequente encontrar o ponto de vista de que a psicologia deve ser capaz de ser expressa, em última análise, em termos da máquina de Turing (por exemplo: Fodor, 1981, p. 130; Boden, 1988, p. 259). Nestes, parece não haver nenhum espaço para modelos mecânicos da mente que não sejam equivalentes à Máquina de Turing. No entanto, é certamente possível que a psicologia encontre a necessidade de empregar modelos de cognição humana que transcendam as máquinas de Turing, ou mesmo que não possam ser reduzidos a modelos mecânicos da mente. Ou seja, esta ainda é uma hipótese em aberto, e não há até o momento nenhuma lei que categoricamente impeça essa possibilidade, como alguns parecem fazer crer.

Note ainda que, em alguns casos, o aparente endosso de algum autor pela tese M é apenas aparente. Neste contexto, é importante lembrar que na literatura técnica a palavra “computável” é muitas vezes ligada, por definição, a calculabilidade efetiva (Copeland, 2015). Portanto, uma função é dita computável se e apenas se existe um processo efetivo para a determinação de seus valores. Assim, uma formulação comum da tese de Church-Turing na literatura técnica e nos livros didáticos é: *todas as funções computáveis são computáveis por Máquina de Turing*. Corolários, como o a seguir, são algumas vezes oferecidos: “certas funções são incomputáveis em um sentido absoluto: incomputável mesmo por [máquina de Turing], e, portanto, incomputável por qualquer máquina real do passado, presente ou futuro” (Boolos e Jeffrey, 1980, p. 55).

Deste modo, dada à definição de “computável” como “efetivamente calculável”, a tese de Church-Turing implica que se uma função ‘f’ não é computável por máquina de Turing, então não é computável por qualquer máquina. No entanto, para um leitor casual da literatura técnica, essas declarações podem parecer dizer mais do que elas de fato dizem. E, mesmo com a decisão de amarrar o termo “computável” com o conceito de efetividade não liquida o valor de verdade da tese M. Aqueles que respeitam esta decisão terminológica são simplesmente prevenidos de descrever uma máquina que falsifica a tese M de computar a função que ela gera. A palavra “mecânica”, também, no uso da linguagem técnica, está ligada à efetividade e, como já se observou, os termos “mecânico” e “efetivo” são utilizados de forma intercambiáveis. Gandy (1988) descreve a história deste uso da palavra “mecânica”. Assim, declarações como a seguinte podem ser encontrados na literatura técnica: “Turing propôs que

uma certa classe de máquinas abstratas poderiam realizar qualquer procedimento de computação mecânica” (Mendelson, 1964, p. 229).

Ainda de acordo com Copeland (2015), entendida corretamente, essa observação atribui a Turing não a tese M, mas a tese de Church-Turing. Esse uso de “mecânico” tende a obscurecer a possibilidade de que podem haver máquinas, ou órgãos biológicos, que calculem (ou computem, em sentido lato) funções que não são computáveis pela Máquina de Turing. A pergunta “Pode uma máquina executar um procedimento que não seja mecânico?” pode aparecer auto-respondida, contudo esta é precisamente a interrogação feita se a tese M for questionada. No entanto, esta não é a única tese problemática que está ligada à tese de Church-Turing. Um erro que, segundo Copeland, infelizmente é muito comum na literatura moderna sobre computabilidade e o cérebro é sustentar que os resultados de Turing de alguma forma implicam que o cérebro, assim como qualquer sistema biológico ou físico que seja, possa ser simulado por uma máquina de Turing.

Como exemplo, podemos citar uma passagem do verbete de Turing na Oxford “*A Companion to the Philosophy of Mind*” onde, partindo da seguinte afirmação, “podemos depender da existência de uma Máquina de Turing que captura as relações funcionais do cérebro”, chega tão distante quanto, “essas relações entre *input* e *output* são funcionalmente bem-comportadas o suficiente para serem descritíveis por ... relações matemáticas ... nós sabemos que alguma versão específica de uma máquina de Turing será capaz de imitá-las” (Guttenplan, 1994, p.595). Searle, em “*The Rediscovery of the Mind*”, escreve de uma forma semelhante:

Podem as operações do cérebro serem simuladas por um computador digital? ... A resposta parece-me ... comprovadamente ‘Sim’ ... Isto é, se naturalmente interpretada, a questão significa: Existe alguma descrição do cérebro de tal forma que sob essa descrição você poderia fazer uma simulação computacional das operações do cérebro. Mas, dada a tese de Church de que qualquer coisa que possa ser dada uma caracterização bastante precisa como um conjunto de passos pode ser simulada em um computador digital, segue-se trivialmente que a pergunta tem uma resposta afirmativa. (1992, p. 200, tradução nossa)

Da mesma forma, podemos ver em Johnson-Laird (1987):

Se você assumir que [a consciência] é cientificamente explicável ... [e] se é garantido que a tese de Church-Turing está correta, então ... se você acredita que [o funcionalismo] é falso ... então ... você [deve] sustentar que a consciência pode ser modelada em um programa de computador da mesma forma que, digamos, o clima pode ser modelado ... [e se] você aceitar o funcionalismo ... você deve acreditar que a consciência é um processo computacional. (p. 252, tradução nossa)

E em Churchland e Churchland (1983), lemos:

A tese de Church diz que tudo o que é computável é Turing computável. Assumindo que, com alguma segurança, o que a mente-cérebro faz é computável, então isso pode ser, em princípio, simulado por um computador. (p. 6, tradução nossa)

A tese de Church-Turing não implica que o cérebro (ou a mente, ou a consciência) possa ser modelado por um programa da Máquina de Turing, nem mesmo em conjunto com a crença de que o cérebro (ou mente, etc.) seja cientificamente explicável, ou exiba um padrão sistemático de respostas ao meio ambiente, ou ainda, seja “governado por regras” (etc.). Cada um dos autores citados parece estar assumindo a verdade de um primo próximo da tese M, que Copeland (2015) chama de *tese S*: “qualquer processo onde possa ser apresentada uma descrição matemática (ou que seja cientificamente descritível ou cientificamente explicável) pode ser simulado por uma máquina de Turing”. Tal como acontece com a tese M, nem a tese de Church-Turing propriamente dita, nem qualquer resultado comprovado por Turing ou Church implica a tese S, e é assim mesmo que a tese seja tomada num sentido mais estrito, no tocante a processos que estão em conformidade com a física do mundo real (tomada num sentido geral seria falsa da mesma forma que a tese M, pelos mesmos motivos apresentados, de acordo com as referências dadas anteriormente).

Assim, qualquer dispositivo ou órgão cujos processos internos possam ser descritos completamente por meio de funções efetivamente calculáveis podem ser simulados por um programa de máquina de Turing, desde que o *input* para o dispositivo ou os órgãos seja em si mesmo computável por uma máquina de Turing, o que é o mesmo que dizer que seja finito ou expresse como um número computável, no sentido de Turing (o qual é explicado a seguir); porém, qualquer dispositivo ou órgão cuja descrição matemática envolva funções que não sejam efetivamente calculáveis não pode ser desse modo simulado. Como Turing (1937a, 1937b) mesmo mostrou, tais funções existem e são incontáveis seus exemplos, conforme apresentado nestes últimos artigos que citamos. Portanto, ainda é uma questão em aberto se a neurociência concluída vai empregar funções que não são efetivamente calculáveis (Copeland, 2015).

Turing introduz suas máquinas com a intenção de fornecer uma descrição idealizada de uma determinada atividade humana, a atividade tediosa da computação numérica que, até o advento de máquinas de computação automáticas, era a ocupação de milhares de pessoas em estabelecimentos de negócios, governamentais e de pesquisa. Portanto, podemos dizer que a máquina de Turing é um modelo, idealizado em certos aspectos, de um ser humano calculando de acordo com um procedimento eficaz. Wittgenstein colocar esse ponto de forma bem direta:

“As ‘Máquinas’ de Turing. Estas máquinas são humanos que calculam” (Wittgenstein, 1980, p. 1096)

Esse é um ponto que Turing enfatizava, de várias formas, uma ou outra vez, como por exemplo quando diz “um homem dotado de papel, lápis e borracha, e sujeito a uma disciplina rigorosa, é em efeito uma máquina universal” (Turing, 1948, p. 9). Ao falar dos computadores digitais, para os quais a Máquina de Turing Universal foi um modelo técnico, cada um deles, são computacionalmente equivalentes a uma máquina de Turing e, portanto, eles também são, em um certo sentido, modelos de seres humanos envolvidos em computação. Turing escolheu as seguintes palavras para enfatizar isso ao explicar essas máquinas eletrônicas, ainda em seu estágio inicial de desenvolvimento, de uma maneira adequada para um público leigo: “A ideia por trás dos computadores digitais pode ser explicada por dizer que estas máquinas se destinam a realizar quaisquer operações que poderiam ser feitas por um humano calculando” (Turing, 1950, p. 436). Também podemos ver, em um documento técnico, contendo seu projeto para o Computing Engine ou ACE Automático³, como Turing reforça, de forma mais precisa, esse ponto:

A classe de problemas capazes de solução por parte da máquina [ACE] pode ser definida bem especificamente. Eles são justamente o subconjunto de problemas que podem ser resolvidos pelo exaustivo trabalho de escritório humano, trabalhando com regras fixas e sem entendimento. (Turing, 1946, p. 38-9, tradução nossa).

Como nos esclarece Copeland (2015), não era alguma deficiência de imaginação ou preguiça que levaria Turing a formular suas máquinas de computação para fazer algo que poderia ser alcançado por um humano calculando, a finalidade para a qual a máquina de Turing fora inventada assim o exigia. O “Entscheidungsproblem” é o problema de saber se um certo tipo de procedimento seria humanamente executável, e o objetivo de Turing foi precisamente mostrar que não existe tal procedimento no caso da lógica de predicados. Ele argumentou ainda que o seu modelo de computação humana é suficientemente geral, no sentido de que não existe nenhuma função intuitivamente computável (ou seja, efetivamente calculável) que as máquinas de Turing não sejam capazes de computar.

Esta última afirmação é, naturalmente, a tese de Turing. Apresentaremos aqui, outras duas formulações adicionais da tese, do seu artigo de 1936.

Os “números computáveis” [números cujas representações decimais podem ser geradas

³ O ACE foi um computador eletrônico construído no Laboratório Nacional de Física, em Londres. Uma versão piloto concorreu pela primeira vez em 1950 e na época era o computador mais rápido do mundo. O modelo comercial foi chamado de Deuce.

progressivamente por uma máquina de Turing] incluem todos os números que naturalmente são considerados como computáveis. (1937a, p. 249, tradução nossa)

É minha asserção que estas operações [as operações primitivas de uma máquina de Turing] incluem todas aquelas que são utilizadas no cálculo de um número. (1937a, p. 232, tradução nossa)

Para compreender estas afirmações, como Turing teria pretendido, é essencial ter em mente que quando ele usa as palavras “computador”, “computável” e “computação”, ele as emprega não em seu sentido moderno como pertencentes a máquinas, mas como pertencentes a humanos na atividade de calcular. Muitas passagens tornam isso óbvio.

Computadores sempre despendem um longo tempo escrevendo números e decidindo o que fazer a seguir depois de fazer as multiplicações atuais, e é a mesma coisa com o ACE ... ele [o ACE] fará o trabalho de cerca de 10.000 computadores ... Os computadores ainda serão empregados em pequenos cálculos ... (Turing, 1947, p. 120, tradução nossa)

Assim, quando Turing afirma que todo número ou função que “seria naturalmente considerada computável” pode ser calculada por uma máquina de Turing, ele não está afirmando a tese M, mas uma tese sobre a extensão dos números e funções efetivamente calculáveis. Da mesma forma, quando Church escreveu (em uma revisão de Post (1936)): “definir eficácia como computabilidade por uma máquina arbitrária, sujeita as restrições de finitude, parece ser uma representação adequada da noção usual” (Church 1937b, p. 43), ele deve ser entendido não como favorável a alguma forma da tese M, mas como favorável a identificação das funções calculáveis de forma eficaz com aquelas funções que podem ser calculados por uma máquina arbitrária, cujos princípios de funcionamento são tais como para imitar as ações de um humano calculando.

É igualmente importante notar também que quando Turing usa a palavra “máquina” ele muita das vezes não dá o sentido de uma máquina-em-geral, como diríamos hoje, mas o de uma máquina de Turing. Em um ponto ele chama explicitamente a atenção para esse uso idiossincrático: “A expressão ‘processo de máquina’ significa, é claro, o processo que poderia ser realizado pelo tipo de máquina que eu estava pensando [em Turing (1936)]” (Turing 1947, p. 107).

É interessante frisar também que, com os avanços da teoria da computação, novas máquinas de Turing foram surgindo e, apesar de o princípio básico ser o mesmo, e essas novas máquinas de Turing poderem ser reduzidas a ideia original de Turing, essas novas versões são muito mais poderosas e de especial interesse para a moderna I.A. e a simulação de processos mentais superiores. Uma delas, conhecida como Máquina de Turing Não-Determinística, difere de uma forma muito particular da Máquina de Turing padrão, que seria considerada

determinística. Essa última possui uma função de transição que, dado um estado e um símbolo na posição de execução da fita, isso vai especificar três coisas: um novo símbolo a ser escrito na posição de execução da fita, a direção para o qual a fita deve mover-se e o seu novo estado. Por exemplo, um X na fita no estado A pode fazer a máquina determinística escrever um Y na fita, mover a cabeça uma posição para a direita e mudar para o estado B. Uma máquina de Turing Não-Determinística vai diferir da primeira pois um estado e um símbolo de fita não mais definem estas três coisas de forma única – mais de uma ação pode ser aplicável dado um estado e um símbolo. Como essa máquina Não-Determinística “sabe” qual dessas ações ela deve tomar? Há duas maneiras de olhar esta questão. Uma é supor que a máquina sempre escolherá uma transição que eventualmente leve a um estado de aceitação. A outra maneira é imaginar que a máquina se ramifica em muitas cópias, cada qual leva a diferentes possíveis transições. Enquanto uma Máquina de Turing Determinística possui apenas um único “caminho de computação” a ser seguido, uma Máquina de Turing Não-Determinística possui uma “árvore de computação”. Se qualquer ramo da árvore para em uma condição de aceitação, dizemos que a Máquina de Turing Não-Determinística aceita a entrada e, portanto, emite sua resposta tomando como consideração o caminho que levou a uma resposta aceita (Sipser, 2007).

No mais, o próprio Turing, em *“Computing Machinery and Intelligence”* (1950), examina a questão de saber se as máquinas poderiam “pensar”. De acordo com ele, pensada dessa forma a questão seria por si só sem sentido para merecer discussão (p. 442). No entanto, se a considerarmos de uma forma mais precisa, e de algum modo relacionada a questão primeiramente levantada, de se um computador digital pode fazer bem um certo tipo de jogo que Turing descreve como “jogo da imitação”, então, pelo menos a seus olhos, temos uma questão que se admite uma discussão satisfatória. Para tanto, ou seja, para testar se um computador pode fazer esse “jogo de imitação”, Turing desenvolve o que ficou conhecido como “Teste de Turing”. Essa expressão é por vezes usada de forma mais geral para se referir a alguns tipos de testes comportamentais que buscam demarcar a presença de mente, pensamento ou inteligência em entidades supostamente possuidoras de tais atributos. Assim, por exemplo, às vezes é sugerido que o teste de Turing é prefigurado no *Discurso do Método* de Descartes. Copeland (2000a, p. 527) encontra uma antecipação do teste nos escritos cartesianos de 1668. No *Discurso*, Descartes diz:

Se houvesse máquinas que tivessem os órgãos e o aspecto de um macaco ou de qualquer outro animal sem razão, não teríamos nenhum meio de reconhecer que elas não seriam, em tudo, da mesma natureza desses animais; ao passo que, se houvesse algumas que se assemelhassem a nossos corpos e imitassem as nossas ações tanto quanto moralmente é possível, teríamos

sempre dois meios muito certos para reconhecer que, mesmo assim, não seriam homens verdadeiros. O primeiro é que nunca poderiam servir-se de palavras nem de outros sinais, combinando-os como fazemos para declarar aos outros nossos pensamentos. Pois pode-se conceber que uma máquina seja feita de tal modo que profira palavras, e até profira algumas a propósito das ações corporais que causem alguma mudança em seus órgãos, como por exemplo ela perguntar o que lhe queremos dizer se lhe tocarmos em algum lugar, se em outro, gritar que a machucamos, e outras coisas semelhantes, mas não é possível conceber que as combine de outro modo para responder ao sentido de tudo quanto dissermos em sua presença, como os homens mais embrutecidos podem fazer. E o segundo é que, embora fizessem várias coisas tão bem ou talvez melhor do que algum de nós, essas máquinas falhariam necessariamente em outras, pelas quais se descobriria que não agiam por conhecimento, mas somente pela disposição de seus órgãos. Pois, enquanto a razão é um instrumento universal, que pode servir em todas as circunstâncias, esses órgãos necessitam de alguma disposição particular para cada ação particular; daí ser moralmente impossível que haja numa máquina a diversidade suficiente de órgãos para fazê-la agir em todas as ocorrências da vida da mesma maneira que nossa razão nos faz agir. (2001, pp. 63-64)

Parece aqui que Descartes dá uma resposta negativa à questão de saber se as máquinas podem pensar; e, além disso, parece que o motivo para essa resposta negativa está ligado justamente a uma confiança de que nenhuma máquina poderia passar num teste do mesmo tipo do proposto por Turing: uma máquina não poderia falar e agir da mesma maneira que humanos adultos o fazem. Uma vez que Descartes explicitamente diz que existem “dois meios muito certos” pelos quais podemos descartar que alguma coisa é uma máquina – já que é, de acordo com Descartes, inconcebível que uma mera máquina poderia produzir diferentes arranjos de palavras, de modo a dar uma resposta adequadamente significativa a tudo o que é dito em sua presença; e é para todos os propósitos práticos impossível para uma máquina ter órgãos diferentes o suficiente para fazê-la agir em todas as contingências da vida da maneira na qual nossa razão nos faz agir – parece que ele deve concordar com a outra alegação que nada do que pode produzir diferentes arranjos de palavras de modo a dar uma resposta adequadamente significativa para tudo o que é dito na sua presença pode ser uma simples máquina. Dada a suposição adicional, que suspeita-se que Descartes estaria preparado a conceder, que somente as coisas que pensam poderiam produzir diferentes arranjos de palavras, de modo a dar uma resposta adequadamente significativa para o que for dito em sua presença, parece seguir-se daí que Descartes teria concordado que o teste de Turing seria um bom teste para sua suposição de que não pode haver máquinas pensantes. Logo, dado o conhecimento de que algo é realmente uma máquina, e evidências de que essa coisa pode produzir diferentes arranjos de palavras, de modo a dar uma resposta adequadamente significativa para tudo o que é dito em sua presença, este seria um indício de que podem haver máquinas pensantes.

Podemos ver também a expressão “Teste de Turing” como ligado a certos tipos de condições logicamente suficientes, alegadamente de natureza puramente comportamental, para

a presença de mente, ou pensamento, ou inteligência, em entidades supostamente possuidoras de tais atributos. Visto desta forma, por exemplo, o experimento mental do “estúpido”, oferecido por Ned Block (1981), é frequentemente apontado como uma forte objeção ao teste de Turing. De forma bem abreviada, o que um representante desta visão tem em mente é a ideia de que é *logicamente possível* para uma entidade passar pelos tipos de testes que Descartes e Turing tem em mente, para usar as palavras e, talvez, até mesmo agir exatamente da mesma maneira que os seres humanos fazem e, ainda assim, serem inteiramente deficientes em inteligência, não possuidor de uma mente, etc.

Visto isso, no mais, estas importantes conquistas tecnológicas – as contribuições de Turing e o desenvolvimento da computação – foram de grande inspiração para o surgimento de uma nova forma de se pensar o problema mente-cérebro, ideia esta que abordaremos a seguir e que foi decisiva para a fundamentação da psicologia cognitiva: o funcionalismo.

2.3. Funcionalismo

Podemos dizer que funcionalismo em filosofia da mente é a doutrina segundo a qual o que faz algo um pensamento, desejo, dor (ou qualquer outro tipo de estado mental) não depende da sua constituição interna ou essência, mas sim unicamente de sua função, ou o papel que desempenha, no sistema cognitivo do qual é uma parte. Esta doutrina têm antecedentes em Hobbes na sua concepção da mente como uma “máquina de calcular”, tornando-se totalmente articulada (e popularmente apoiada) no último terço do século XX. Embora o termo “funcionalismo” seja usado também para designar uma variedade de posições em uma variedade de disciplinas, o que inclui psicologia, sociologia, economia e arquitetura, nos concentraremos aqui exclusivamente no funcionalismo como uma tese filosófica sobre a natureza dos estados mentais.

Teorias funcionalistas assumem a identidade de um estado mental a ser determinado por suas relações causais com estímulos sensoriais, outros estados mentais e comportamento. Dando um exemplo simplista, uma teoria funcionalista pode caracterizar *dor* como um estado que tende a ser causado por um lesão corporal, que produz a crença de que algo está errado com o corpo e o desejo de estar fora desse estado, o que produz ansiedade e, na ausência de qualquer força maior, ou desejos conflitantes, retração do corpo ou lamento. De acordo com esta teoria, toda e qualquer criatura com estados internos que cumpram estas condições, ou que desempenhem essas funções, são capazes de estar com dor (Levin, 2013).

Suponhamos que, nos seres humanos, existe uma espécie característica de atividade neuronal (estimulação das fibras C, por exemplo) que satisfaça estas condições. Se assim for, então de acordo com a teoria funcionalista, os seres humanos podem estar em dor simplesmente submetendo-se a estimulação das fibras C. Porém, além disso, a teoria permite que criaturas com constituições físicas bem diferentes da constituição humana possuam estados mentais: se, hipoteticamente, existissem alienígenas baseados em silício ou androides inorgânicos que também atendessem a essas condições, então essas criaturas, da mesma maneira, poderiam estar com dor. Como funcionalistas muitas vezes colocam, a dor pode ser *realizada* por diferentes tipos de estados físicos em diferentes tipos de criaturas ou, dito de outra forma, pode ser *multiplamente realizada*. De fato, uma vez que as descrições fazem referência explícita apenas às relações causais de um estado com estímulos, a comportamentos, e o que veio a ser conhecido como “tópico neutro” (Smart, 1959) – que é uma imposição de não haverem quaisquer restrições lógicas sobre a natureza dos itens que satisfaçam as descrições – então também é logicamente possível que estados *não*-físicos desempenhem essas funções e, portanto, realizem estados mentais noutros sistemas. Desta maneira, segundo Levin (2013), o funcionalismo é compatível com o tipo de dualismo que leva estados mentais a causar, e serem causados por, estados físicos.

Ainda assim, embora o funcionalismo seja oficialmente neutro entre o materialismo e o dualismo, ele tem sido particularmente atraente para os materialistas, uma vez que muitos materialistas acreditam, segundo Lewis (1966), que é altamente provável que quaisquer estados capazes de reproduzir as funções em questão são unicamente os estados físicos. Se assim for, então o funcionalismo pode ser visto como uma alternativa materialista à tese da identidade psicofísica (Place, 1956; Smart, 1959; e, mais recentemente, Polger, 2011), que afirma que cada tipo de estado mental é idêntico com um tipo particular de estado *neural*. Esta tese parece implicar que nenhuma criatura com um cérebro diferente do nosso possa compartilhar nossas sensações, crenças e desejos, não importando o quão semelhante seja o seu comportamento e organização interna do nosso próprio e, portanto, o funcionalismo, com a sua alegação de que os estados mentais podem ser multiplamente realizados, tem sido considerado como fornecedor de uma teoria da mente que é ao mesmo tempo compatível com o materialismo e mais inclusiva ou, conforme Block (1980), com um menor “chauvinismo de espécie”. No entanto, mais recentemente, alguns filósofos têm sustentado que a tese da identidade pode ser mais abrangente do que os funcionalistas estão dispostos a assumir (como podemos ver em Bickle, 2012; Shapiro & Polger, 2012).

Assim, um outro ponto que queremos chamar atenção é a relação do funcionalismo com a chamada teoria da identidade mente/cérebro. Essa teoria afirma que a mente e seus processos são idênticos aos estados e processos do cérebro. Estritamente falando, não é necessário considerar que a mente seja idêntica ao cérebro já que, no uso corriqueiro da linguagem, podemos falar “ele não está bom da cabeça/cérebro” e “ele não tem uma mente boa” de forma intercambiáveis, mas não se diria “sua mente pesa 1,5kg”. Aqui, tomamos a identidade mente e cérebro como sendo uma questão de identificar processos e talvez estados da mente e do cérebro. Considere uma experiência de dor, ou de ver algo, ou de ter uma imagem mental. A teoria da identidade da mente em efeito diz que essas experiências são apenas processos cerebrais, e não que estejam meramente correlacionados com processos cerebrais (Levin, 2013).

Tem sido comum pensar que a teoria da identidade foi substituída pela teoria do funcionalismo. Porém, pode-se argumentar que os funcionalistas exageraram muito a sua diferença em relação a teóricos da identidade. De fato, alguns filósofos, como Lewis (1972 e 1994) e Jackson, Pargetter e Prior (1982), veem o funcionalismo como uma rota em direção a uma teoria da identidade. Como Lewis, alguns funcionalistas definem estados e processos mentais em termos de suas relações causais com o comportamento mas param de identificá-los com as suas realizações neurais. É claro que o termo “funcionalismo” tem sido usado vagamente e de diferentes maneiras. A palavra “funcionalista” tem afinidades com a de “função” em matemática e também com a de “função” na biologia. Em matemática uma função é uma relação entre dois ou mais conjuntos, estabelecida por um lei de formação, isto é, uma regra geral. Da mesma forma, se os processos mentais são definidos direta ou indiretamente por conjuntos de pares de estímulo-resposta as definições poderiam ser vistas como “funcionais” no sentido matemático. No entanto, há provavelmente uma ligação mais estreita com o termo como é usado na biologia, como se poderia definir “olho” pela sua função, embora o olho de uma mosca e olho de um cão sejam anatomicamente e fisiologicamente muito diferentes. Funcionalismo identifica estados e processos mentais por meio de seus papéis causais e, como mencionado acima, sabemos que os papéis funcionais são possuídos por estados e processos neurais. No entanto, um dualista interacionista, como o neurofisiologista Sir John Eccles (2000) negaria que todos os papéis funcionais sejam assim tão bem encarnados.

Pensar em termos causais sobre as crenças e desejos se encaixa muito bem não só com a *folk psychology*, mas também com as ideias humeanas sobre os motivos da ação. De acordo com Smart (2014), embora este ponto de vista tenha sido criticado por alguns filósofos, ele pode parecer correto, como pode ser visto, se considerarmos um possível avião robô projetado

para encontrar o seu caminho de Melbourne para Sydney. Segundo esse exemplo proposto por Smart, o designer teria de incluir uma versão eletrônica de algo como um mapa do sudeste da Austrália, o que proporcionaria uma aproximação do que chamamos “crença”. Também seria programado um equivalente eletrônico a “ir para Sydney”. Este programa proporcionaria o lado “desejo”. Se alguma tempestade colocasse o avião para fora do curso, em seguida, uma reação iria empurrar o avião de volta para o caminho certo para Sydney. A existência de mecanismos intencionais tem, finalmente, mostrado aos filósofos que não há nada de misterioso sobre a teleologia. Também não existem grandes problemas semânticos com o termo intencionalidade. Uma consideração holística não precisaria supor que a individuação de crenças e desejos seja algo preciso, ainda que seja boa o suficiente para a *folk psychology*. Como argumenta Smart, a forma na qual o cérebro representa o mundo pode não ser como um idioma. A representação pode ser mais como um mapa. Um mapa relaciona todos os recursos que estão nele para coisas fora dele. No entanto, mapas contêm uma quantidade finita de informação e podemos pensar em crenças como expressando os diferentes *bits* de informação que poderiam ser extraídos a partir do mapa. Pensando dessa maneira, crenças corresponderiam mais aproximadamente ao que consideramos serem as crenças individuais características da *folk psychology*.

Dentro desta ampla caracterização do funcionalismo, no entanto, uma série de distinções podem ser feitas. Uma, de particular importância, é a distinção entre teorias nas quais as caracterizações funcionais dos estados mentais se propõem a fornecer análises dos significados das condições dos estado mental (ou, de outra forma, restringirem-se as informações a priori) e teorias que permitem a caracterização funcional de estados mentais para recorrer a informações provenientes de experimentação científica (Rey, 1997; Shoemaker, 1984). Há outras diferenças importantes entre as teorias funcionalistas. Estas diferenças e as motivações para tais, podem ser melhores apreciadas ao se examinar as origens do funcionalismo e ao se traçar sua evolução em resposta tanto às críticas explícitas a tese quanto a mudanças nos pontos de vista sobre a natureza da explicação psicológica. Obviamente, tal feito não pode ser aqui aprofundado, visto o objetivo do presente trabalho, e o limite de tempo e espaço para tal. Porém, pretendemos fazer mais algumas observações relativas ao funcionalismo que podem vir a ser proveitosas e nos ajudar a melhor entender o posicionamento da psicologia cognitiva, e das ciências cognitivas como um todo, quanto à questão mente-cérebro e, mais especificamente, ao problema da liberdade da vontade e em como se poderia pensar na vontade, como algo mental, tendo efeito causal sobre o comportamento.

A visão mais antiga que pode ser considerada um antepassado distante do funcionalismo é a teoria da alma de Aristóteles (2006). Em contraste com a reivindicação de Platão de que a alma pode existir sem o corpo, Aristóteles argumentou (2006, pp.71-73) que a alma humana é a *forma* do corpo humano, o conjunto de poderes ou capacidades que lhe permitem expressar aquilo o que realmente é, sua essência, o que, para Aristóteles, é uma questão de cumprir a função ou finalidade que a define como o tipo de coisa que é. Assim como a forma de um machado é tudo o que lhe permite cortar, e a forma de um olho é tudo o que lhe permite ver, a alma humana deve ser identificada com todos aqueles poderes e capacidades que permitem um corpo humano naturalmente organizado cumprir sua função, que, segundo Aristóteles, é sobreviver e florescer como a vida, agir, perceber e ser racional. Assim, Aristóteles argumenta, a alma é inseparável do corpo, e compreende todas as capacidades que são requeridas para que um corpo possa viver, perceber, raciocinar e agir.

Um outro importante antepassado do funcionalismo contemporâneo são as considerações feitas em 1651 por Hobbes em seu *Leviatã*, onde o raciocínio é visto como uma espécie de computação que se processa por princípios mecânicos comparáveis às regras da aritmética. Raciocínio, Hobbes argumenta, é “nada além de certos *cálculos*, que está a somar e subtrair, das consequências dos nomes em geral acordados para a *marcação* e *significação* dos nossos pensamentos” (1973, p. 113). Além disso, Hobbes sugere que o raciocínio, juntamente com a imaginação, sensação e deliberação sobre ação, todos esses procedem de acordo com princípios mecanicistas, podendo ser realizados por sistemas de variados tipos físicos. Como ele coloca em sua Introdução do *Leviatã*, onde compara uma comunidade com um indivíduo humano, “por que não podemos dizer que todos os autômatos (maquinas que se movem a si mesmas por meio de molas e rodas, tais como um relógio) possuem uma vida artificial? Pois o que é o coração, senão um espécie de mola; e os nervos, senão outras tantas cordas; e as articulações, senão outras tantas rodas” (1973, p. 09). Não foi até meados do século XX, no entanto, que tornou-se comum especular que o pensamento pode ser nada mais que uma computação governada por regras que pode ser realizados por criaturas de constituições físicas variadas.

Como vimos anteriormente, Turing (1950) propôs que a pergunta: “Podem as máquinas pensar?” fosse substituída pela questão: “É teoricamente possível que um computador digital de estados finitos, provido com uma extensa mas finita tabela de instruções, ou programa, possa fornecer respostas às perguntas de forma a enganar um interrogador ignorante [de que se trata de uma máquina] fazendo-o pensar que se trata de um ser humano?”. Agora, em referência ao

seu autor, esta questão é mais frequentemente expressa como “é teoricamente possível que um computador digital de estados finitos (adequadamente programado) possa passar no teste de Turing?”. Ao argumentar que esta pergunta é um substituto legítimo para a original e especulando que a sua resposta seja “sim”, Turing identifica pensamentos com os estados de um sistema definido unicamente por seus papéis na produção de outros estados internos e *outputs* verbais, uma visão que tem muito em comum com as teorias funcionalistas contemporâneas. E, segundo Levin (2013), o trabalho de Turing foi explicitamente invocado por muitos teóricos durante os estágios iniciais do funcionalismo do século XX, além de ter sido uma inspiração declarada para uma classe de teorias, as teorias “*machine state*” mais comumente associadas com Hilary Putnam (1975), e que tiveram um papel importante no desenvolvimento inicial da doutrina.

Ainda segundo Levin (2013), um outro antecedente importante do funcionalismo foram as teorias comportamentais que surgiram em meados do início do século XX. Estas incluem tanto as teorias psicológicas empíricas associadas principalmente a Watson e Skinner, e o behaviorismo lógico e analítico de filósofos como Malcolm (1968), Ryle (1949) e, notadamente, Wittgenstein (1953). Embora o funcionalismo seja significativamente diferente do behaviorismo no que este último tenta explicar o comportamento sem qualquer referência a estados e processos mentais, o desenvolvimento de duas importantes linhagens do funcionalismo, o *psico-funcionalismo* e o *funcionalismo analítico*, ambas podem ser proveitosamente vistas como tentativas para corrigir as dificuldades, respectivamente, do behaviorismo empírico e lógico, mantendo certos *insights* importantes dessas teorias.

Como uma teoria psicológica que busca uma fundamentação empírica, o behaviorismo considera que o comportamento de seres humanos (e outros animais) poderia ser explicado apelando exclusivamente a disposições comportamentais, ou seja, a tendências similares a leis, ou regras, dos organismos se comportarem de certas maneiras, dados certos estímulos ambientais. Essas disposições comportamentais, ao contrário de pensamentos, sentimentos e outros estados internos que só podem ser diretamente observados por introspecção, são objetivamente observáveis e fazem, indiscutivelmente, parte do mundo natural. Assim, comportamentos pareciam ser o objeto legítimo na ciência emergente da psicologia. Além disso, as teorias behavioristas prometiam evitar uma potencial regressão que parecia ameaçar as explicações psicológicas ao se invocarem representações internas, ou seja, que para explicar como tais representações produzem os comportamentos em questão, se apelasse para um agente

inteligente interno (um “homúnculo”), que interpreta essas representações – habilidade esta que também teria de ser explicada.

A promessa do behaviorismo, de acordo com Levin (2013), estava em sua convicção de que poderia haver uma ciência do comportamento humano tão objetiva e explanatória como outras ciências consideradas “de alto nível”, como química e biologia. O behaviorismo de fato teve grande sucesso, mesmo que de início em um contexto muito específico, especialmente no domínio da aprendizagem animal, e muitos de seus princípios ainda são usados, pelo menos para propósitos heurísticos, em diversas áreas da psicologia. Mas muitos psicólogos (e outros, como por exemplo, Chomsky) argumentaram que o sucesso do behaviorismo parecia depender principalmente de um controle implícito de certas variáveis que, quando feito de forma mais explícita, envolveria uma referência a estados mentais que dificilmente poderiam ser excluídos de consideração. Por exemplo, ratos são normalmente colocados em uma situação experimental com uma certa fração de seu peso normal e, portanto, pode ser assumido que *sentem fome* e *querem* as recompensas de alimentos que os condicionam a se comportar de determinadas maneiras. Da mesma forma, presume-se que os seres humanos, em situações experimentais análogas, *querem* cooperar com os experimentadores, e *entendem* como seguir suas instruções.

Parecia, portanto, aos críticos do behaviorismo, que as teorias que apelavam explicitamente às crenças de um organismo, seus desejos e outros estados mentais, bem como a estímulos e comportamento, proporcionariam uma consideração mais completa e acurada dos motivos pelos quais os organismos se comportam como se comportam. E isso poderia ser feito, além disso, sem comprometer a objetividade da psicologia, desde que os estados mentais a que essas teorias se referem fossem apresentados como estados que juntos desempenham uma função na produção de comportamento, ao invés de estados identificáveis apenas por introspecção. Assim, para Levin (2013), esse trabalho foi iniciado com uma série de teorias psicológicas “cognitivas” que refletiam estes pressupostos. Uma importante força do funcionalismo contemporâneo, que chamamos aqui de *psico-funcionalismo* (Fodor 1968, Block & Fodor 1972), pode ser vista como um suporte filosófico dessas novas teorias cognitivas da mente.

O behaviorismo lógico, em contraste com o behaviorismo como uma teoria psicológica, é uma tese sobre os significados dos nossos conceitos ou termos mentais. De tal modo, de acordo com o behaviorismo lógico, todas as afirmações sobre estados e processos mentais são equivalentes em sentido a declarações sobre disposições comportamentais. Podemos dar como

exemplo, mais uma vez de forma bem simplista, a declaração “Alice tem uma dor de dente” como equivalente em sentido a uma declaração do tipo “Alice está disposta (estando todas as condições iguais) para gritar ou gemer e esfregar o queixo”. Ou, a declaração “Alice está com sede” como equivalente a uma declaração como “se for oferecida a Alice um pouco de água, ela vai estar disposta (com todas as condições estando iguais) a beber”. Além disso, os behavioristas lógicos argumentam que, se afirmações sobre estados mentais são equivalentes em sentido a declarações sobre disposições comportamentais, poderia haver uma explicação não tão problemática de como termos mentais podem ser aplicados tanto quanto se fala de si mesmo quanto aos outros, e como eles podem ser ensinados e aprendidos.

Entretanto, como alguns filósofos têm apontado (Chisholm, 1957; Geach, 1957), o behaviorismo lógico fornece uma explicação não muito plausível dos significados dos termos referentes aos nossos estados mentais, uma vez que, intuitivamente, um sujeito pode ter os estados mentais em questão, sem as disposições comportamentais relevantes, e vice versa. Por exemplo, um ator pode simular perfeitamente um comportamento da dor, sob certas condições, mesmo ele realmente não sentindo dor. O problema, segundo Putnam (1965), é que nenhum estado mental, por si só, pode ser assumido de forma plausível para dar origem a qualquer comportamento particular a menos que também se pressuponha que o sujeito possui outros estados mentais adicionais de vários tipos. E assim, ao que parece, não é de fato possível dar um sentido de preservação a traduções de declarações que invocam dores, crenças e desejos em termos puramente behavioristas. No entanto, a ideia de que os conceitos de senso comum dos estados mentais revelam um laço essencial entre estados mentais e suas expressões comportamentais típicas é mantido, e elaborado, nas teorias funcionalistas “analíticas” contemporâneas.

Tendo em conta essas influências históricas, e utilizando uma divisão mais didática, como a proposta por Levin (2013), é útil pensar em teorias funcionalistas como pertencentes a uma das três principais forças no funcionalismo, que chamaremos aqui de *funcionalismo computacional* (“machine state functionalism”), *psico-funcionalismo* e *funcionalismo analítico*, e vê-las como emergentes, respectivamente, das teorias A.I., do behaviorismo empírico e do behaviorismo lógico. É importante reconhecer, contudo, que existem pelo menos algumas sobreposições nas linhagens dessas diferentes forças do funcionalismo, e também que há teorias funcionalistas, tanto anteriores quanto mais recentes, que podem ser classificadas em algum lugar entre as apresentadas.

2.3.1. Funcionalismo computacional

Em seu início, as teorias funcionalistas de Putnam (1975) podem ser vistas como uma resposta às dificuldades enfrentadas pelo behaviorismo como uma teoria científica psicológica e como uma defesa das novas teorias computacionais da mente, que se apresentavam, cada vez mais, como importantes rivais a teoria behaviorista. De acordo com o *funcionalismo computacional* de Putnam, qualquer criatura com uma mente pode ser considerada como uma máquina de Turing, cuja operação pode ser totalmente definida por um conjunto de instruções (ou programa) cada uma tendo a forma:

Se a máquina está em um estado de S_i , e recebe o *input* I_j , ele vai entrar em um estado de S_k e produzir um *output* O_l (para um número finito de estados, *inputs* e *outputs*).

Um programa desse tipo descreve a operação de uma máquina de Turing determinística, mas a maioria dos funcionalistas computacionais (e.g., Putnam, 1975) consideram que um modelo mais adequado para a mente seria o de um autômato *probabilístico*: aquele em que o programa especifica, para cada estado e conjunto de entradas, a *probabilidade* com que a máquina vai entrar em algum estado posterior e produzir alguma saída particular (Levin, 2013).

Em ambos os modelos, no entanto, os estados mentais de uma criatura devem ser identificados com estados computacionais da máquina ($S_1, \dots, S_n$). Sendo que esses estados não são meramente disposições comportamentais, uma vez que eles são especificados não apenas em termos das suas relações de *inputs* e *outputs*, mas também para o estado da máquina no momento. Por exemplo, se *acreditar que irá chover* é considerado como um estado computacional da máquina, não será considerado como uma disposição para pegar um guarda-chuva apenas ter-se olhado a previsão do tempo, mas sim como uma disposição para pegar um guarda-chuva, se tiver olhado a previsão do tempo e se estiver no estado de querer continuar seco (o *querer* como um estado da máquina). Assim, o funcionalismo computacional pôde evitar o que muitos pensavam ser uma grande dificuldade do behaviorismo. Além disso, máquinas deste tipo ofereceriam um modelo simples de como estados internos cujos efeitos sobre o *output* ocorrendo por meio de processos mecânicos podem ser vistos como *representações* (embora a questão sobre o *que*, exatamente, eles representam tem sido um tema constante de discussões, como veremos mais a frente). Finalmente, estados computacionais da máquina não estão vinculados a qualquer realização física particular; já que, o mesmo programa, afinal, poderia ser executado em diferentes tipos de *hardware* de computador (Levin, 2013).

É fácil ver, portanto, como as máquinas de Turing forneceram um modelo frutífero para teorias funcionalistas, pelo menos em seu início. No entanto, dado que os estados computacionais de uma máquina são nada menos que todo o estado de um sistema, a equação funcionalista, oferecida em seus primórdios, dos estados mentais com estados computacionais de uma máquina perdeu importância como um modelo para a caracterização funcional do complexo de estados internos diferentes que podem ser realizados simultaneamente em um sujeito humano, ou outro qualquer (Block & Fodor, 1972; Putnam, 1975). No entanto, a ideia de que os estados internos podem ser totalmente descritos em termos de suas relações com *inputs*, *outputs*, e *entre si*, e podem figurar nas descrições de regras e previsões de um sistema de *outputs*, foi uma ideia rica e importante que foi mantida pela teorias funcionalistas contemporâneas. Alguns funcionalistas, como por exemplo Rey (1997), argumentam que os estados mentais são melhor caracterizados como estados computacionais (apesar de haver discordâncias, para tanto, veja Piccinini, 2004).

2.3.2. Psico-funcionalismo

Uma segunda força no funcionalismo, o psico-funcionalismo, deriva principalmente da reflexão sobre os objetivos e metodologias de teorias psicológicas cognitivas. Em contraste com a insistência dos behavioristas de que as leis da psicologia deveriam apelar unicamente para disposições comportamentais, psicólogos cognitivos argumentam que teorias empíricas do comportamento mais fidedignas deveriam tomar como base um complexo de estados e processos mentais, estabelecidos e individuados em termos de regras que desempenhem um papel na produção do comportamento a ser explicado. Por exemplo (como mostrado por Fodor, 1968, Cap. 3), um psicólogo pode começar a construir uma teoria da memória, postulando a existência do declínio do “traço de memória” – um processo cuja ocorrência ou ausência é responsável por efeitos como a perda de memória ou sua retenção e que é afetada pelo estresse ou emoção – de maneiras distintas (Levin, 2013).

Em uma teoria deste tipo, o que faz de algum processo neural uma instância de declínio de traço de memória é uma questão de como ele funciona, ou o papel que desempenha, em um sistema cognitivo. As propriedades neurais e químicas desse dado processo são relevantes na medida em que permitem que haja o declínio do traço de memória, à forma como foi hipotetizado. De forma similar, assim também o é para outros estados e processos mentais invocados por teorias psicológicas cognitivas. Desta maneira, a psicologia cognitiva, como um

todo, dispõe-se a ser uma ciência de “nível elevado” assim como a biologia por exemplo o é, e, portanto, ter sua autonomia das ciências de “nível inferior”, como a neurofisiologia. Da mesma forma que, como na biologia, entidades fisicamente distintas podem ter corações, que são identificados como tais pois funcionam para fazer circular o sangue num determinado organismo vivo, e outras entidades fisicamente distintas podem ter olhos, das mais diferentes formas, mas que são igualmente identificados como tais enquanto permitem a um organismo vivo ver. Diferentes estruturas ou processos físicos podem ser instancias para a deterioração do traço de memória – ou mesmo fenômenos tais como pensamentos, sensações e desejos – contanto que eles desempenhem os papéis descritos pela teoria cognitiva relevante (Levin, 2013).

Assim sendo, como sustenta Levin (2013), o psico-funcionalismo parece adotar diretamente uma metodologia da psicologia cognitiva em sua caracterização de estados e processos mentais como entidades definidas pelo seu papel em uma teoria psicológica cognitiva. Todas as versões do funcionalismo, no entanto, caracterizam os estados mentais em termos de seus papéis em *alguma* teoria psicológica ou em outra teoria qualquer. O que é característico sobre o psico-funcionalismo, para Levin, é sua alegação de que estados e processos mentais são apenas aquelas entidades, com apenas aquelas propriedades, postuladas pela melhor explicação *científica* do comportamento humano, mesmo que seja difícil saber qual seriam os critérios para essa tal melhor explicação científica. Isto significa, para eles, em primeiro lugar, que a forma da teoria pode afastar-se das especificações do funcionalismo computacional. Isso também significa que as informações utilizadas na caracterização funcional de estados e processos mentais não precisam ser restritas ao que é considerado conhecimento comum ou o senso comum, mas podem incluir informações disponíveis apenas pela observação laboratorial cuidadosa e experimentação. Por exemplo, uma teoria psico-funcional pode ser capaz de distinguir fenômenos como a depressão da tristeza ou apatia, mesmo embora as causas e os efeitos distintos dessas síndromes sejam difíceis de desembaraçar apenas consultando intuições ou apelando para o senso comum.

Isto pode parecer uma vantagem absoluta, uma vez que as teorias psico-funcionais podem aproveitar-se de todas as ferramentas de pesquisa disponíveis à psicologia científica, e, presumivelmente, fazer todas, e apenas, as distinções que são cientificamente corretas. Esta metodologia, no entanto, deixa o psico-funcionalismo aberto à acusação de que, como a tese da identidade psicofísica, ele seja excessivamente excludentes (Block, 1980), uma vez que criaturas cujos estados internos compartilhem conosco um mesmo arcabouço de padrões

causais, mas que não sejam tão refinados como os nossos, não contariam como partilhando dos nossos estados mentais. Muitos psico-funcionalistas podem não considerar isso como uma consequência infeliz, e argumentam que isto é apropriado pois permite tratar apenas aqueles que são psicologicamente semelhantes como tendo os mesmos estados mentais. Mas há uma preocupação mais séria sobre a tese, a saber, que, se as leis das melhores teorias psicológicas empíricas divergirem até mesmo dos contornos gerais de nossa “*folk psychology*” – isto é, de nossas crenças de senso comum sobre os papéis causais de nossos pensamentos, sensações, percepções e *vontades* – vai ser difícil tomar as teorias psico-funcionais como fornecedoras de um relato de nossos estados mentais, preferindo simplesmente mudar seu objeto (Loar, 1981; Stich, 1983). Alguns teóricos (Horgan & Woodward, 1985), no entanto, argumentam que é provável que futuras teorias psicológicas serão reconhecidamente mais aproximadas da “*folk psychology*”, embora esta questão também tenha sido objeto de debate (Churchland, 1981).

Mas há uma outra importante força no funcionalismo, o funcionalismo analítico, que busca justamente restringir a teoria não exatamente a generalizações suficientemente próximas às que usamos na “*folk psychology*” e que promovam uma sustentação entre os estados mentais, estímulos ambientais e nosso comportamento, mas sim, para além disso, busca uma informação *a priori* sobre essas relações (veja, por exemplo, Smart, 1959; Shoemaker, 1984). Isto é porque, para os funcionalistas analíticos, existem objetivos igualmente importantes que requerem uma caracterização estritamente *a priori* dos estados mentais.

2.3.3. Funcionalismo analítico

Como o behaviorismo lógico, do qual sofreu fortes influências, segundo argumenta Levin (2013), o objetivo do funcionalismo analítico é fornecer traduções “neutras”, ou análises, do nosso vocabulário ordinário de termos referentes a estados mentais. O funcionalismo analítico, no entanto, tem recursos mais abrangentes do que o behaviorismo lógico para tais traduções, uma vez que permite referência às relações causais que um estado mental tem a estímulos, comportamentos e *outros estados mentais*. Uma boa maneira de ver por que os funcionalistas analíticos insistem que caracterizações funcionais fornecem análises significantes é vislumbrada recordando um debate que ocorreu relacionado à “Teoria da Identidade Psicofísica”, sendo esta a tese de que cada tipo de estado mental pode ser identificado com algum tipo de estado cerebral ou atividade neural. Alguns teóricos da identidade, como Smart (1959), argumentaram que faz todo o sentido, e que poderia muito bem

ser verdadeiro, identificar a sensação de *dor* com a *estimulação das fibras C*. Os termos “dor” e “estimulação das fibras C” não têm o mesmo *significado*, mas ainda assim poderiam denotar o mesmo estado; o fato de que uma declaração de identidade não é a priori, eles argumentaram, não significa que isso não seja verdadeiro. E, só porque eu não preciso consultar algum tipo de scanner cerebral ao relatar que estou com dor, não significa que a dor que eu relatar não seja um estado neural que um scanner cerebral pode (em princípio) detectar.

Uma importante e duradoura objeção feita a este argumento foi levantada inicialmente por Max Black (como relatada por Smart, 1959). Black argumentou, seguindo uma concepção fregeana, que a única maneira onde termos com significados diferentes possam denotar o mesmo estado seria ao expressarem propriedades diferentes, ou “modos de apresentação”, desse estado. Mas isso implica que, Black argumentou, se termos como “dor”, “pensamento” e “desejo” não são equivalentes em significado a todas as suas descrições fisicalistas, eles podem indicar estados físicos apenas por expressar as *propriedades irredutivelmente mentais* desses termos. Assim, mesmo que os termos “dor” e “estimulação das fibras C” façam um recorte de um único tipo de estado neural, esse estado deve ter dois tipos de *propriedades*, física e mental, por meio das quais a identificação pode ser feita. Este argumento veio a ser conhecido como “argumento das propriedades distintas” (Levin, 2013).

Segundo Levin (2013), o apelo a caracterizações que preservem o sentido funcional, portanto, é o que, ao fornecer termos equivalentes “neutros” para os nossos termos e conceitos mentais, neutraliza a força anti-materialista do argumento das propriedades distintas. Em verdade, como funcionalistas analíticos podem reconhecer, termos como “dor”, “pensamento” e “desejo” não são equivalentes a quaisquer descrições expressas na linguagem da física, da química, ou da neurofisiologia. Mas se existem descrições *funcionais* que preservam o significado destes termos, então os estados mentais de uma criatura podem ser identificados simplesmente determinando qual dos estados e processos internos daquela criatura desempenham os papéis funcionais relevantes (Lewis, 1966). E, uma vez que a capacidade de desempenhar esses papéis seja meramente uma questão de ter certas relações causais com estímulos, comportamento, ou outros (a vontade incluída?), a posse dessas propriedades é compatível com uma teoria materialista da mente.

A grande questão, de acordo com Levin (2013), é se uma teoria que se limita à informação *a priori* sobre as relações causais entre estímulos, estados mentais e comportamento pode fazer as distinções corretas entre os estados mentais. Outro ponto que merece um melhor esclarecimento é sobre a própria construção das teorias funcionais, visto que até agora a

discussão de como fornecer caracterizações funcionais dos estados mentais individuais tem sido por demais vaga e os exemplos dados por demais simplistas. É possível fazer melhor e, em caso afirmativo, qual a versão do funcionalismo é provável que tenha um maior êxito? Buscaremos abordar essas perguntas na próxima seção, oferecendo um tratamento separado aos estados experienciais, como percepções e sensações corporais, que têm um caráter qualitativo distintivo ou “sentir” e, o que pode estar mais intimamente relacionado ao tema da vontade, – estados intencionais, como pensamentos, crenças e desejos, que pretendem representar o mundo de várias maneiras (embora haja um consenso crescente de que os estados experienciais têm conteúdo de representação e estados intencionais tem caráter qualitativo e, portanto, que esses dois grupos podem não ser mutuamente exclusivas).

Primeiro, no entanto, é importante conhecer mais precisamente como supostamente funcionaria uma definição funcional. Podemos fazer isto concentrando-nos em um método geral para a construção de definições funcionais, que foi inicialmente introduzido por David Lewis (1972), e que veio a se tornar uma prática padrão para os funcionalistas de todas as variedades. Articular este método irá nos ajudar a avaliar os pontos fortes e fracos de diferentes variedades do funcionalismo – enquanto indicamos alguns novos desafios que surgem para todos eles.

2.3.4. Definições funcionais e a sentença de Ramsey

A principal característica deste método, agora canônico de acordo com Levin (2013), é tratar estados e processos mentais como sendo definidos implicitamente pela *sentença de Ramsey* de uma ou outra teoria psicológica – o senso comum, um teoria científica, ou analítica. Como exemplo, considere as generalizações sobre a dor introduzidas anteriormente: dor tende a ser causada por lesões corporais; dor tende a produzir a crença de que algo está errado com o corpo e o desejo de sair desse estado; dor tende a produzir a ansiedade; dor tende a produzir reflexo de afastamento e/ou lamúria. Para construir a *sentença de Ramsey* desta “teoria”, o primeiro passo é unir essas generalizações, em seguida, substituir todos os nomes de diferentes tipos de estados mentais com diferentes variáveis e, em seguida, quantificar essas variáveis, conforme segue:

$\exists x \exists y \exists z \exists w$ (x tende a ser causado por lesão corporal & x tende a produzir estados y , z e w & x tende a produzir afastamento e/ou lamúria).

Tal afirmação é livre de quaisquer termos de estados mentais. Ela inclui apenas *quantificadores* que variam ao longo de estados mentais, termos que denotam estímulos e comportamento, e termos que especificam várias relações causais entre eles. Podem, portanto, serem considerados como fornecendo definições implícitas dos termos referentes a estados mentais da teoria. Um indivíduo terá esses estados mentais apenas no caso em que possui uma família de estados de primeira ordem que interagem nas formas especificadas pela teoria.

Uma maneira útil de pensar a sentença de Ramsey de uma teoria psicológica é considerá-la como definindo um sistema de estados mentais “como um todo”, como os estados que interagem com estímulos de várias maneiras possíveis para produzir comportamento (Lewis, 1972). Isso deixa claro que, nas formulações clássicas de teorias funcionais, estados mentais são destinados a serem caracterizados em termos de suas relações com os estímulos, comportamentos, e *todos os outros estados* que podem ser aceitavelmente invocados pela teoria em questão. Portanto, certas teorias funcionais podem ter mais recursos para individualizar os estados mentais do que o sugerido pelas definições gerais utilizadas como exemplos. As próximas três seções irão discutir o potencial de vários tipos de teorias funcionalistas para dar caracterizações adequadas de estados experienciais e intencionais e também para especificar os *inputs* e *outputs* do sistema.

2.3.5. Caracterização de estados experienciais

A estratégia comum nos tratamentos de maior sucesso de experiências perceptivas e sensações corporais, segundo Shoemaker (1984), é individualizar experiências de vários tipos gerais (experiências de cor, experiências de sons, sensação de temperatura) em parte pelo apelo às suas posições nos “espaços de qualidade” associados com as modalidades sensoriais relevantes – isto é, as matrizes (talvez multidimensionais) determinadas por julgamentos sobre as semelhanças relativas e diferenças entre as experiências em questão. Assim, por exemplo, a experiência de uma cor laranja muito avermelhada poderia ser (parcialmente) caracterizada como o estado produzido pela visualização de uma amostra de cor dentro de algum intervalo de espectro de luz particular, que tende a produzir o julgamento ou opinião que o estado experimentado apenas é mais semelhante à experiência de vermelho do que de laranja. Caracterizações análogas terão de ser dadas para outras experiências de cor. Os julgamentos ou crenças em questão irão ser, eles próprios, (parcialmente) caracterizados em

termos de suas tendências para produzir classificações ou categorizações de comportamentos de certos tipos especificados.

De acordo com Levin (2013), esta estratégia pode parecer fatal para o funcionalismo analítico, que se restringe ao uso de informações *a priori* para distinguir entre os estados mentais, uma vez que não está claro que a informação necessária para distinguir entre as experiências tais como as percepções de cor resultaram da análise conceitual dos termos ou conceitos de nossos estados mentais. No entanto, este problema pode não ser tão terrível quanto parece. Por exemplo, se as sensações e experiências perceptivas são caracterizadas em termos de seus lugares em um “espaço de qualidade” determinado por juízos pré-teóricos pessoais de similaridade e de dissimilaridade (e talvez também em termos de suas tendências para produzir vários efeitos emocionais), então essas caracterizações podem qualificar-se como *a priori*, mesmo que elas tenham de ser provocadas por uma espécie de “questionamento socrático”.

Há limites para esta estratégia, no entanto, que parecem deixar duas opções para os funcionalistas analíticos: luta – ou seja, negar que é coerente supor que existem distinções que os críticos sugerem; ou mudança – ou seja, abraçar uma outra versão do funcionalismo em que as caracterizações dos estados mentais, embora não verdades conceptuais, podem fornecer informações ricas o suficiente para individualizar os estados em questão. Mudar, no entanto, seria desistir dos benefícios (se houver) de uma teoria que oferece traduções de preservação de sentido dos nossos termos do estado mental.

Segundo Levin (2013), houve ceticismo significativo, no entanto, sobre se qualquer teoria funcionalista – seja analítica ou científica – poderia capturar o que parecem ser os caracteres intrínsecos de estados experienciais tais como as percepções de cor, dores e outras sensações corporais; estas questões serão abordadas mais à frente.

2.3.6. Caracterizações de estados intencionais

Por outro lado, estados intencionais, tais como crenças, pensamentos e desejos (às vezes chamados de “atitudes proposicionais”) são frequentemente tomados como sendo facilmente especificados em termos funcionais (mas nem sempre: por exemplo, ver Searle, 1992; e G. Strawson, 1994, que sugerem que os estados intencionais também tem um caráter qualitativo). Como mostra Levin (2013), não é tão difícil ver por onde começar: crenças são (entre outras

coisas) estados produzidos de certa forma pela senso-percepção ou por inferência de outras crenças, que tendem a interagir com certos desejos para produzir o comportamento; desejos são estados com certas relações causais ou contrafactuais com um sistema de metas e necessidades, e que também tendem a interagir com certas crenças para produzir comportamento.

Porém, mais deve ser dito sobre o que faz um estado uma crença ou desejo particular. Por exemplo, a crença – ou o desejo – de que vai nevar amanhã. A maioria das teorias funcionais descrevem tais estados como relações diferentes (ou “atitudes”) em direção ao mesmo estado de coisas ou proposições (como para descrever a crença de que vai nevar amanhã e a crença de que vai chover amanhã, como a mesma atitude para proposições diferentes). Isso permite que as diferenças e semelhanças no conteúdo de estados intencionais sejam interpretadas como diferenças e semelhanças entre as proposições em que esses estados estão relacionados. Mas o que faz um estado mental uma relação para, ou um atitude em relação a, alguma proposição P? E podem essas relações serem capturadas apenas por apelo aos papéis funcionais dos estados em questão?

De acordo com Loar (1981), o desenvolvimento de regras conceituais para a semântica parece dar uma resposta a estas perguntas: o que é para Julian acreditar que P é estar em um estado que tem relações causais e contrafactuais com outras crenças e desejos que espelham certas relações inferenciais, evidenciais e práticas (dirigidas para a ação) entre proposições com essas estruturas formais. Esta proposta levanta uma série de questões importantes. Uma delas é se os estados capazes de entrar em tais inter-relações podem (ou devem?) ser interpretados como compreendendo ou incluindo elementos de uma “linguagem do pensamento” (Fodor, 1975). Outra questão é se idiosincrasias nas inclinações inferenciais ou práticas de diferentes indivíduos produzem diferenças em (ou incomensurabilidades entre) seus estados intencionais.

Como considera Levin (2013), ainda outro desafio para o funcionalismo são as intuições generalizadas que suportam o “externalismo”, a tese de que o que os estados mentais representam não podem ser caracterizados sem o apelo para certas características do ambiente no qual os indivíduos estão incorporados. Assim, se o ambiente de um indivíduo difere de outro, eles podem ser considerados como tendo diferentes estados intencionais, mesmo que eles raciocinem da mesma forma, e tenham exatamente a mesma perspectiva sobre seus ambientes de seus próprios pontos de vista. O cenário de “Terra Gêmea” introduzido por Putnam (1975) é frequentemente invocado para apoiar uma individuação externalista de crenças sobre coisas naturais como a água, uma pedra, árvores ou tigres, por exemplo. A Terra Gêmea, como Putnam apresentou, é um planeta hipotético onde as coisas tem a mesma aparência, cheiro, paladar,

exatamente da mesma maneira que na Terra, mas que têm diferentes estruturas microscópicas; por exemplo, o material que enche os riachos e sai das torneiras, embora se pareça e tenha gosto de água, tem uma estrutura molecular XYZ em vez de H₂O. Parece intuitivo pensar que, assim sendo, nós *significaríamos* algo diferente pelo nosso termo “água” do que os nossos homólogos da Terra Gêmea diriam da deles, e assim as crenças que descrevemos como crenças sobre a água seriam diferentes daquelas que os nossos homólogos da Terra Gêmea. Conclusões semelhantes podem ser elaboradas para todos os casos de crença (e outros estados intencionais) sobre coisas naturais.

O mesmo problema, por outro lado, parece surgir para outros tipos de crenças. Tyler Burge (1979) apresenta casos em que parece intuitivo que uma pessoa, Oscar, e sua funcionalmente equivalente contraparte, tenham diferentes crenças sobre várias síndromes (como a artrite) e objetos (tais como sofás), porque o uso desses termos por sua comunidade linguística diferem. Por exemplo, na comunidade de Oscar, o termo “artrite” é usado como habitualmente usamos, enquanto que “artrite” na comunidade de seu homólogo denota a inflamação das articulações e também várias doenças da coxa. A alegação de Burge é que mesmo que Oscar e seu homólogo se queixem da “artrite” em suas coxas e façam exatamente as mesmas inferências envolvendo “artrite”, elas significam coisas diferentes e eles devem ser considerados como tendo diferentes crenças. Se estes casos são convincentes, então existem diferenças entre os tipos de estados intencionais que só podem ser capturadas por caracterizações destes estados que fazem referência às práticas da comunidade linguística do indivíduo. Estes casos apresentados sugerem que, se as teorias funcionalistas não podem fazer referência ao ambiente de um indivíduo, capturar o conteúdo representacional mesmo que de pelo menos alguns estados intencionais, então isto está fora do âmbito do funcionalismo.

Por outro lado, a individuação externalista de estados intencionais podem deixar de captar algumas semelhanças psicológicas importantes entre nós e os nossos colegas que são relevantes para uma explicação do comportamento. Se meu homólogo da Terra Gêmea e eu, ambos viemos de uma longa caminhada, e ao declarar que estamos com sede dizemos “eu quero um pouco de água” e vamos para a cozinha, parece que o nosso comportamento pode ser explicado citando um desejo e crença comum. De acordo com Levin (2013), alguns teóricos têm sugerido que as teorias funcionais devem se esforçar apenas para capturar o que tem sido chamado de “conteúdo restrito” de crenças e desejos – sejam quais forem as características representacionais que os indivíduos compartilham com seus homólogos da Terra Gêmea. Não há consenso, no entanto, sobre como teorias funcionalistas deveriam tratar esses “restritos”

recursos representacionais (Block, 1986), e alguns filósofos expressaram ceticismo sobre se tais características devem ser interpretadas como representações (Fodor, 1994). Mesmo que fosse desenvolvida uma consideração de um conteúdo representacional restrito que fosse comumente aceitável, no entanto, se as intuições inspiradas no cenários da Terra Gêmea se mantivessem estáveis, então se poderia concluir que o conteúdo representacional completo dos estados intencionais (e estados qualitativos, se eles também têm conteúdo representacional) não pode ser capturado apenas por caracterizações funcionais restritas.

2.3.7. Caracterização dos *inputs* e *outputs* de um sistema

Considerações sobre se certos tipos de crenças devem ser externamente individualizadas levantam a questão sobre a melhor forma de caracterizar os estímulos e comportamentos que servem como *inputs* e *outputs* para um sistema. Elas devem ser interpretadas como eventos envolvendo objetos no ambiente de um sistema, ou melhor, como eventos em sistemas sensoriais e sistemas motores? Teorias do primeiro tipo são frequentemente chamadas “braço longo” das teorias funcionais (Block, 1990), uma vez que caracterizam *inputs* e *outputs* – e, consequentemente, os estados pelos quais ele são produzidos e produzem – pelo contato com o mundo. A adoção de uma teoria “braço longo” impediria que os nossos homólogos da Terra Gêmea compartilhassem nossas crenças e desejos, e pode, assim, honrar intuições que suportariam uma individuação externalista de estados intencionais.

Se caracterizações funcionais dos estados intencionais são destinados a capturar seus “conteúdos restritos”, no entanto, os *inputs* e *outputs* do sistema terão de ser especificados de uma forma que permita que indivíduos em diferentes ambientes compartilhem um mesmo estado intencional. Em vista desta situação *inputs* e *outputs* podem ser melhor caracterizados como uma atividade em receptores sensoriais específicos e neurônios motores. Mas esta opção (conhecida como “braço curto”) também restringe o leque de indivíduos que podem compartilhar nossas crenças e desejos, já que criaturas com diferentes estruturas neurais seriam impedidas de partilhar nossos estados mentais, mesmo se eles compartilhassem todas as nossas disposições comportamentais e inferenciais (Block, 1990).

Talvez haja uma maneira de delimitar estímulos sensoriais que abstraia das especificidades da estrutura neural humana o suficiente para incluir qualquer criatura possível que intuitivamente pareça compartilhar nossos estados mentais, mas que ao mesmo tempo seja suficientemente concreta para excluir entidades que são claramente sistemas não cognitivos

(como a economia da Bolívia; como ironizado por Block, 1980). Se não houver tal formulação, no entanto, então funcionalistas ou terão de dispensar intuições no sentido de que certos sistemas não podem ter crenças e desejos, ou admitirem que suas teorias podem ser mais excludentes do que o inicialmente esperado.

Os problemas aqui espelham as questões relacionadas com a individuação dos estados intencionais discutidos na seção anterior. Seria necessário, porém, um maior trabalho para desenvolver as alternativas de “braço longo” e “braço curto”, e para avaliar os méritos e as deficiências de ambas. Nas seções anteriores fizemos uma apresentação mais geral do funcionalismo e de algumas ramificações deste, quanto a aspectos de influências e a avaliação de seus pontos fortes e fracos. Houve, porém, muitas objeções ao funcionalismo que se aplicam a todas as versões da teoria. Algumas delas já foram introduzidas em discussões anteriores, outras tentaremos abordar aqui em maiores detalhes.

2.3.8. Críticas ao funcionalismo

Uma dificuldade para cada versão da teoria é que sua caracterização funcional é *holística*. Funcionalistas sustentam que os estados mentais devem ser caracterizados em termos de seus papéis em uma teoria psicológica – seja ela *folk-psychology*, uma teoria científica, ou algo entre isso –, mas todas essas teorias incorporam informações sobre um grande número e variedade de estados mentais. Assim, se a dor é definida como relacionada a certas crenças altamente articuladas e desejos, então os animais que não têm estados internos que desempenham os papéis de nossas crenças e desejos não podem compartilhar nossas dores, assim como seres humanos sem a capacidade de sentir dor não podem compartilhar certas (ou talvez nenhuma) de nossas crenças e desejos. Além disso, as diferenças na forma com que as pessoas raciocinam, a forma como suas crenças são fixas, ou a forma com que seus desejos afetam suas crenças – devido tanto a idiossincrasias culturais ou individuais – deveria tornar impossível que pessoas compartilhassem os mesmos estados mentais. Estas são consideradas sérias preocupações para todas as versões do funcionalismo (Stich, 1983; Putnam, 1988).

Outra preocupação para os funcionalistas é o “problema da exclusão causal”: a preocupação sobre qual regra funcional pode ser responsável por aquilo que consideramos ser a eficácia causal dos nossos estados mentais (Malcolm 1968; Kim 1989, 1998). Por exemplo, se a dor é realizada em mim por algum tipo de estado neural, então, na medida em que existe algo parecido a leis puramente físicas que ligam os estados desse tipo com o comportamento

de dor, pode-se dar uma explicação causal completa do meu comportamento, citando a ocorrência desse estados neurais (e as propriedades em virtude das quais eles representam nessas leis). Desta maneira, segundo esta visão, as propriedades da função de nível superior desse estado – no caso exemplificado a dor – são causalmente irrelevantes.

Tem havido um número de diferentes respostas para este problema. Alguns (por exemplo Loewer, 2002, 2007; Burge, 1995; Baker, 1995) sugerem que surge a partir de uma consideração excessivamente restritiva do nexos de causalidade, na qual uma causa deve “gerar” ou “produzir” o efeito, uma visão que colocaria as propriedades macroscópicas de outras ciências como causalmente irrelevantes também. Em vez disso, o nexos de causalidade deve ser considerado como um tipo especial de dependência contrafactual entre estados de certos tipos (Loewer, 2002, 2007; Block, 1997), ou como um tipo especial de regularidade que se mantém entre eles (Melnik, 2003). Se isso estiver correto, então as propriedades funcionais (juntamente com outras propriedades macroscópicas de ciências como Química e Biologia) poderiam contar como causalmente eficazes. No entanto, a plausibilidade dessas considerações sobre a causalidade depende de suas perspectivas para distinguir de boa-fé relações causais daquelas que são claramente epifenômenos, e alguns expressaram ceticismo sobre se isso realmente pode ser feito, entre eles Kim (2007) e Jackson (1996).

Outros filósofos argumentam que a causalidade seria melhor considerada como uma relação entre tipos de eventos que deveria ser invocada para fornecer explicações suficientemente gerais de seus comportamentos (Burge, 1995; Baker, 1995). Embora muitos dos que são movidos pelo problema da exclusão, como por exemplo Jackson (1996) e Kim (2007), sustentam que há uma diferença entre generalizações que são verdadeiramente causais e aquelas que contribuem de alguma outra forma (meramente epistêmica) para a nossa compreensão do mundo, teóricos que defendem esta resposta ao problema acusam que essa objeção, mais uma vez, depende de uma visão restrita do nexos de causalidade que excluiria as demais.

Outra importante questão diz respeito às crenças que temos sobre a ocorrência em nós próprios de estados mentais, como pensamentos, sensações e percepções. Parece que temos imediatamente disponíveis crenças, não-inferenciais, sobre esses estados, e a questão é como isso deve ser explicado se os estados mentais são idênticos com propriedades funcionais. A resposta depende da visão que se tenha dessas crenças introspectivas. Em termos gerais, segundo Levin (2013), há duas visões dominantes do assunto. Um relato popular de introspecção – o modelo “sentido interior” em que a introspecção é considerada como sendo

uma espécie de “varredura interna” do conteúdo de sua mente – foi tomado como sendo hostil ao funcionalismo, alegando que é difícil ver como os objetos de tal verificação podem ser propriedades relacionais de segunda ordem dos próprios estados neurais. Alguns, no entanto, ainda segundo Levin, têm defendido que o funcionalismo pode acomodar as características especiais de crença introspectiva sobre o modelo de “sentido interior”, uma vez que seria apenas um dos muitos domínios em que é plausível pensar que temos conhecimento imediato, não-inferencial, de propriedades causais ou disposicionais.

Há ainda uma outra crítica ao funcionalismo que levanta a questão de saber se qualquer teoria da mente que invoque crenças, desejos e outros estados intencionais jamais poderia ser, mesmo que aproximadamente, uma teoria empírica. Considerando que, mesmo funcionalistas analíticos sustentam que os estados mentais são definidos implicitamente em termos de seus papéis (causais ou probabilísticos) na *produção* de comportamento, esses críticos tomam estados mentais, ou pelo menos estados intencionais, a serem definidos implicitamente em termos de seus papéis na *racionalização*, ou no *fazer sentido*, do comportamento. Esta é uma empreitada diferente, eles afirmam, uma vez que na racionalização, ao contrário da explicação causal, é preciso mostrar como as crenças do indivíduo, seus desejos e comportamentos conformam-se, pelo menos aproximadamente, a certas *normas* ou *ideais a priori* de raciocínio teórico e prático – prescrições sobre como *devemos* raciocinar, ou o que, dadas nossas crenças e desejos, nós *devemos* fazer (Davidson, 1980; Dennett, 1978). Assim, não se pode esperar que a definição de relações normativas ou racionais entre os estados intencionais expressos por estes princípios correspondam a relações empíricas entre os nossos estados internos, estímulos sensoriais e comportamento, uma vez que essa constitui uma espécie de explicação que tem fontes de evidências e padrões de correção que são diferentes dos das teorias empíricas (Davidson, 1980). Ou seja, não se pode extrair fatos de valores. Com isso, embora atribuições dos estados mentais possam, em certo sentido, explicar o comportamento, ao permitir que um observador interprete-os como comportamentos que tem um sentido, não deve ser esperado que denotem entidades que figurem em leis empíricas. Isto não quer dizer, entretanto, que não existam causas ou leis empíricas do comportamento. Mas estes, no entanto, poderão apenas ser expressos no vocabulário das neurociências, ou outras ciências de nível mais baixo, e não como relações entre crenças, desejos e comportamento (Dennett, 1978).

Segundo Levin (2013), funcionalistas têm respondido a estas preocupações de diferentes maneiras. Muitos simplesmente negam a intuição por trás da objeção e afirmam que até mesmo as mais rigorosas análises conceituais dos nossos termos e conceitos intencionais pretende

defini-los em termos de seus papéis causais de boa-fé, e que quaisquer normas que refletem são explicativas em vez de prescritivas. Eles argumentam que, se essas generalizações são idealizações, elas são o tipo de idealizações que ocorrem em qualquer teoria científica: assim como a Lei de Boyle retrata as relações entre a temperatura, pressão e volume de um gás sob certas condições experimentais ideais, nossa teoria da mente consiste em descrições do que os seres humanos normais fazer em condições ideais fisicamente especificáveis, e não prescrições a respeito do que *deveriam* fazer.

Outros funcionalistas concordam que podemos advertir a várias normas de inferência e ação ao atribuir crenças e desejos para os outros, mas negam que haja qualquer incompatibilidade em princípio entre explicações normativas e empíricas. Eles argumentam que, se há relações causais entre crenças, desejos e comportamento que mesmo aproximadamente espelham as normas da racionalidade, em decorrência disto, as atribuições de estados intencionais podem ser empiricamente confirmados (Fodor, 1990). Além disso, muitos dos que sustentam esta opinião sugerem que os princípios da racionalidade aos quais estados intencionais devem satisfazer são mínimos, compreendendo, no máximo, um conjunto de restrições sobre os contornos da nossa teoria da mente, como a de que as pessoas não podem, em geral, manterem crenças contraditórias, ou agirem contra seus sinceramente declarados desejos mais fortes (Loar, 1981).

Como vimos, e embora muitos funcionalistas argumentem que as considerações discutidas acima mostram que não há em princípio uma obstrução a uma teoria funcionalista que tenha força empírica, de acordo com Levin (2013), essas preocupações com a normatividade da atribuição intencional continuam a alimentar um certo ceticismo sobre o funcionalismo e, para além disso, qualquer teoria científica da mente que use noções intencionais. Além dessas preocupações gerais sobre o funcionalismo, há ainda muitas questões específicas que surgem para a perspectiva de dar uma caracterização funcional dos estados qualitativos, ou o problema da *qualia* segundo uma visão funcionalista. Em termos gerais, já que esse assunto extrapolaria o escopo do presente trabalho, esta é uma categoria de estados mentais que parece particularmente resistente a uma caracterização funcional. Como vimos, teorias funcionalistas de todas as variedades buscam caracterizar os estados mentais exclusivamente em termos relacionais, mais especificamente causais. Uma objeção comum e persistente, no entanto, é que não há como tais caracterizações captarem o caráter qualitativo, ou “qualia”, de estados experienciais tais como percepções, emoções e sensações corporais,

uma vez que deixaria de fora algumas das suas propriedades essenciais, a saber, “o como é tê-las” (Nagel, 1974).

Há uma estratégia final para defender uma abordagem funcionalista de estados qualitativos contra todas essas objeções, e que ficou conhecida como eliminativismo (Dennett 1988; Rey, 1997). Segundo esta visão, pode-se negar que haja algo tal qual uma *qualia irreduzível*, e sustentar que a convicção de que se tais coisas existem, ou até mesmo que se poderiam existir, seria devido à uma ilusão. Estratégia similar foi utilizada por Wegner & Wheatley (1999) – *Apparent mental causation: Sources of the experience of will* – e Wegner (2002) – *The illusion of conscious will* – ao sustentar que a vontade consciente, ou melhor, o efeito causal que poderia emanar de um vontade livre, não passa de uma ilusão, ou de um “pseudofenômeno”. Esses trabalhos citados foram fortemente influenciados pelos experimentos de Benjamin Libet, que abordamos mais a frente.

2.4. Construtivismo

A psicologia cognitiva, no geral, quanto a seu posicionamento sobre a questão da origem do conhecimento, está comprometida com uma visão *construtivista* (Castañon, Justi & Araujo, 2014). Tradicionalmente, uma resposta a esta questão era ofertada ou por uma visão de um ambientalismo empirista, ou pelo inatismo racionalista. O construtivismo se apresentou, então, como uma terceira forma de abordar a questão, buscando de certa forma aproximar as duas visões anteriores, reconhecendo não somente a importância dada pelos empiristas ao caráter sensorial da formação do conhecimento, derivado do ambiente e contexto onde o sujeito se insere, mas também a dada pelos racionalistas a estruturas inatas, fundamentais para que o sujeito possa conhecer. O termo surge justamente com Piaget (1967), e busca mostrar o papel ativo do sujeito na construção de suas próprias estruturas cognitivas.

Segundo Castañon (2015), é importante também reconhecer, para uma melhor compreensão do construtivismo contemporâneo, sua forte influência e inspiração kantiana, onde é justamente a *inversão do sentido da relação entre sujeito e objeto* que está na raiz do construtivismo. Onde a filosofia, tradicionalmente, pensava o conhecimento como uma determinação do sujeito que conhece pelo objeto conhecido, Kant apresenta o processo do conhecimento como uma organização ativa do sujeito, através de estruturas *a priori* da razão, do material que nos é fornecido pelos sentidos. Sendo assim, o próprio sujeito constrói suas representações de mundo e não apenas recebe passivamente as impressões causadas pelos

objetos, sendo, portanto, um foco de atividade do universo. De acordo com Castañon (2015), o termo ‘construção’ é empregado em referência à filosofia kantiana em pelo menos dois sentidos: Um primeiro sentido indica que nossas representações dos objetos dos sentidos são construídos de forma automática, pelo trabalho de estruturas inatas de nossa mente, ocorrendo portanto sem interferência da consciência. Um segundo sentido, mais geral, indica um processo autônomo de construção de hipóteses sobre a natureza para um posterior teste empírico de sua validade.

Piaget, portanto, como um herdeiro da tradição kantiana (Piaget, 1987, p.220), recusa tanto o empirismo objetivista, onde as representações mentais seriam impostas à mente pelos objetos, quanto o racionalismo, onde o conhecimento seria inato e seu desenvolvimento se daria mais como uma atualização de estruturas pré-determinadas. Para Piaget (1973), o conhecimento seria o produto de um sujeito ativo que construiria suas representações de mundo em uma interação constante com o objeto do conhecimento. Dada nossa experiência com o mundo, buscamos *assimilar* essa experiência em esquemas e teorias que já possuímos. Porém, quando constatamos que nossas previsões e explicações são repetidamente contraditas, nossos esquemas é que se modificam de forma a *acomodar* estas novas informações. Portanto, somos ativos tanto quando interpretamos a experiência para assimilá-la aos nossos próprios esquemas, e da mesma forma também somos ativos ao mudarmos nossos esquemas de forma a acomodarem-se à realidade.

Segundo Castañon, Justi e Araujo (2014), esta estrutura geral do pensamento piagetiano desde o início provocou uma profunda impressão no movimento cognitivista e se mantém até os dias de hoje inserido na psicologia cognitiva, principalmente através do trabalho dos neopiagetianos, que procuram integrar a abordagem de Piaget à teoria do processamento de informação (Miller, 2012). Como amostra disso, podemos ver, já em 1967, com o primeiro manual de psicologia cognitiva de Ulric Neisser, como o movimento cognitivista começava a chegar a conclusões parecidas, como podemos constatar por essa passagem:

A assertiva central é a de que ver, ouvir e lembrar são todos atos de construção, que podem fazer mais ou menos uso da informação derivada dos estímulos dependendo das circunstâncias. Os processos construtivos são assumidos como tendo dois estágios, dos quais o primeiro é rápido, cru, global e paralelo, enquanto o segundo é deliberado, atencional, detalhado e sequencial (1967, p.10, apud Castañon, Justi e Araújo, 2014).

Ainda segundo Castañon, Justi e Araujo (2014), essas ideias foram fundamentais para o desenvolvimento de outros modelos para a simulação de aspectos construtivistas da cognição humana. Entre eles, e um dos mais tradicionais nos dias de hoje, são os que propõem o modelo

bottom-up e *top-down* do processamento de informação. Neste modelo, todo processamento é executado bi-direcionalmente, ou seja, o *bottom-up* se refere ao processamento diretamente afetado pelo *input* do estímulo, enquanto o *top-down* seria o processamento feito a partir de experiências passadas do sujeito, expectativas que o orientam, tudo aquilo que o sujeito traz à ocorrência do estímulo. Para Eysenck e Keane (1994), a tese dominante na psicologia cognitiva contemporânea, mais uma vez seguindo Neisser (1976), é a de que toda atividade cognitiva envolve ambos os tipos de processamento, que seriam nada mais que um modelo computacional de que nossas hipóteses e expectativas condicionam a seleção das informações que consideramos relevantes num determinado contexto.

2.5. Processamento de Informação

Outro acontecimento importante, e que teve grande influência na psicologia cognitiva, foi, como já citado anteriormente, a cibernética. Um movimento interdisciplinar, envolvendo matemática, engenharia e biologia, juntamente com a psicologia, e que estava, pelo menos inicialmente, fortemente associado às necessidades tecnológicas advindas da Segunda Guerra Mundial. Evoluíram esforços para o desenvolvimento de novas tecnologias, serviços de inteligência, comunicação, treinamento e capacitação. Teve como grande influência também as novas possibilidades abertas pelas descobertas de Turing. Norbert Wiener (1948), um dos principais fundadores da cibernética, tinha como objetivo a elaboração de uma teoria para tratamento de mensagens e controle de sistemas que poderiam se autorregular em função de metas previamente definidas.

Segundo Castañon, Justi e Araujo (2014), duas contribuições foram fundamentais para a psicologia cognitiva. Uma primeira, seria o conceito de *feedback* (retroalimentação), que permitiu o desenvolvimento de sistemas mecânicos que podiam comparar seu comportamento com um resultado planejado, ou metas pré-estabelecidas e, se necessário, modificar seu comportamento para alcançar tais metas. Em segundo lugar, a analogia entre a lógica binária (verdadeiro/falso) com o funcionamento do neurônio e suas conexões (ativado/desativado), que podiam também ser implementados em circuitos elétricos. A partir disso, e da colaboração de um psiquiatra e um matemático, surge então as chamadas redes neurais, e um projeto mais consistente de simulação do cérebro humano (McCullon & Pitts, 1943).

Ainda segundo Castañon, Justi e Araujo (2014), uma outra forte influência para a psicologia cognitiva foi o trabalho do linguista Noam Chomsky. Em 1957, Chomsky apresentou

sua conhecida teoria da linguagem onde propunha um modelo mais eficiente para explicar a linguagem humana natural, em contraposição a teoria da informação e, principalmente, a perspectiva behaviorista da linguagem, que segundo Chomsky não explicava seu caráter criativo e potencialmente infinito de gerar sentenças. Concentrando-se, então, nos aspectos sintáticos da linguagem, Chomsky buscou trazer a precisão formal da matemática para a análise linguística construindo uma visão da linguagem como um processo gerativo de sentenças altamente estruturado, que seria governado por regras gramaticais inconscientes. O âmago da linguagem, de acordo com essa teoria, seria a propriedade da sintaxe, uma capacidade humana inata de combinar e reconstruir símbolos verbais em certas ordens determinadas, o que proporcionaria a criação de uma quantidade potencialmente ilimitada de sentenças gramaticalmente válidas. Segundo esse modelo, a sintaxe é vista como um programa inato e universal de regras para transformação de símbolos, o que independe da semântica, ou seja, seu significado, e da fonologia, ou som.

Além disso, podemos destacar também o surgimento da teoria da informação, ainda sob o impacto do desenvolvimento dos serviços de engenharia e comunicação durante a Segunda Guerra. Uma questão fundamental para a transmissão de informação era como ela poderia ser feita em canais de comunicação com ruído, de forma segura e codificada. Assim, essa teoria vai conceber a informação em termos puramente formais, ou seja, sem a necessidade de se referir ao seu conteúdo ou significação, sendo levado em consideração apenas a fonte dessa informação, o canal de transmissão, o emissor, o receptor e as possibilidades de codificação do sinal. Como essa concepção puramente formal e abstrata, a informação pôde ser generalizada para qualquer sistema de comunicação e foi possível também, além disso, propor uma forma de mensuração da informação em unidades básicas, que vieram a ser conhecidas como *bit*. Com isso, segundo Lachman, Lachman e Butterfield (1979), abria-se a possibilidade de interpretar também os processos mentais em termos de transmissão e processamento de informação.

É comum vermos os cientistas cognitivos descrevendo a computação como “processamento de informação”. É menos comum, porém, um esclarecimento do que eles querem dizer com “informação” ou “processamento”. Sem um tal esclarecimento, a descrição é pouco mais do que um slogan vazio. Claude Shannon introduziu uma noção cientificamente importante de “informação” em seu artigo 1948, “*A Mathematical Theory of Communication*”. A ideia aqui é que a informação mede *redução de incerteza*, em que redução de incerteza se manifesta como uma distribuição de probabilidade alterada sobre possíveis estados. Shannon codificou esta ideia dentro de uma estrutura matemática rigorosa, que

estabelece as bases para a *teoria da informação* (Cover e Thomas, 2006). O conceito de informação de Shannon é fundamental para a engenharia moderna. Ela encontra aplicação frutífera também dentro da ciência cognitiva e neurociência, especialmente a neurociência-cognitiva. Porém, ela suporta uma análise convincente da computação como “processamento de informações”? Considere uma máquina antiquada de fita que grava as mensagens recebidas através de um rádio sem fio. Usando um quadro de Shannon, pode-se medir a quantidade de informação que é carregada por alguma mensagem gravada. Nesse sentido a máquina de fita “processa” informação, no sentido proposto por Shannon, sempre que reproduzir uma mensagem gravada. Ainda assim, a máquina não parece implementar um modelo computacional não-trivial⁴. Certamente, nem o formalismo da máquina de Turing, nem o formalismo da rede neural oferecem grandes *insights* sobre as operações da máquina. Indiscutivelmente, então, um sistema pode processar informações no sentido proposto por Shannon sem executar cálculos em um sentido que seja de maior interesse.

Uma segunda noção importante de informação deriva da influente discussão de Paul Grice (1989) sobre *significado natural*. Significado natural envolveria correlações com suporte contrafactuais confiáveis. Por exemplo, os anéis de uma árvore correlacionam-se com sua idade. Nós coloquialmente descrevemos os anéis das árvores como portadores de informação sobre a idade da árvore, da varíola como portador de informações sobre varicela, e assim por diante. Tais descrições sugerem uma concepção que une informações para correlações com suporte contrafactuais confiáveis. Fred Dretske (1981) desenvolveu essa concepção em uma teoria sistemática, assim como vários filósofos subsequentes. Será que informação no estilo de Dretske seria útil para uma análise plausível da computação como “processamento de informação”? Considere um antiquado *termostato de tira biometálica*. Dois metais são unidos em uma tira e a expansão diferencial dos metais faz com que a tira se dobre, ativando ou desativando uma unidade de aquecimento. O estado dessa tira correlaciona de forma confiável com a temperatura ambiente atual, e o termostato “processa” essas informações de suporte quando ativa ou desativa o aquecedor. No entanto, o termostato não parece implementar qualquer modelo computacional não-trivial. Não se poderia normalmente considerar o

⁴ Um modelo trivial seria aquele que lança pouca ou nenhuma luz sobre as operações da máquina. Uma preocupação recorrente é que se um modelo é trivial, podemos descrever praticamente qualquer sistema físico como executando cálculos. Searle (1990) afirma que uma parede implementaria qualquer programa de computador, uma vez que podemos discernir algum padrão de movimento molecular na parede, que é isomorfo à estrutura formal de um programa. Argumentos de trivialidade desempenham um papel importante na literatura filosófica da área. Anti-computacionalistas usam argumentos de trivialidade contra computacionalistas, enquanto computacionalistas procuram evitar a trivialidade.

termostato como fazendo algum tipo de computação. Indiscutivelmente, então, um sistema pode processar informações no estilo Dretske sem executar cálculos em qualquer sentido de interesse.

Uma terceira noção proeminente de informação é a *informação semântica*, ou seja, o conteúdo representacional. Alguns filósofos sustentam que um sistema físico calcula somente se os estados do sistema têm propriedades representacionais (Dietrich, 1989; Fodor, 1998). Nesse sentido, o processamento de informações é *necessário* para a computação. Como Fodor coloca, “não há computação sem representação” (1975, p.34). No entanto, esta posição também é discutível. Chalmers (2011) afirma que uma máquina de Turing pode executar cálculos embora os símbolos manipulados pela máquina não tenham interpretação semântica. Os cálculos da máquina são puramente sintáticos por natureza, faltando qualquer coisa como propriedades semânticas. Neste ponto de vista, o conteúdo da representação não é necessário para um sistema físico ser considerado como computacional.

Em uma passagem muito citada, o psicólogo David Marr (1982), distingue três níveis em que se pode descrever um “dispositivo de processamento de informação”:

Teoria computacional: “o dispositivo é caracterizado como um mapeamento de um tipo de informação para outro, as propriedades abstratas deste mapeamento são definidas com precisão, e sua pertinência e adequação para a tarefa em mãos são demonstradas” (p. 24).

Representação e algoritmo: “a escolha da representação para o *input* e *output* e o algoritmo a ser utilizado para transformar um no outro” (pp 24-25.).

Implementação de hardware: “os detalhes de como o algoritmo e representação são realizados fisicamente” (p 25).

Esses três níveis de Marr têm atraído intenso escrutínio filosófico. Marr ilustra sua abordagem fornecendo teorias de “nível computacional” de vários processos perceptivos, tais como detecção de bordas. A discussão de Marr sugere uma *concepção funcional da computação*, em que a computação é uma questão de transformar *inputs* em *outputs* apropriados.

De modo geral, como vimos, o processamento de informação é um modelo proposto por teóricos cognitivistas inspirados no funcionamento do computador além dos trabalhos do neo-behaviorista Tolman (1948), que chamou atenção para a importância dos mapas cognitivos, representações mentais da informação circundante relevante para as intenções do indivíduo;

conjugado ainda com as influências da *gestaltheorie*, ao analisar a organização do pensamento, o papel da compreensão, as estratégias cognitivas e a estrutura da consciência. Palmer e Kimchi (1986) apresentam uma visão extensiva sobre as características da abordagem do processamento de informação e, a fim de caracterizar mais claramente essa abordagem dentro da psicologia, propuseram cinco suposições básicas sobre esta:

1. Descrição informacional: os eventos mentais podem ser descritos como eventos informacionais e se dividem em três partes: *input*, operações sobre a informação e *output*. É importante notar que interessa aos psicólogos a noção do conteúdo representacional mais que a noção de informação como uma quantidade mensurável. A visão da informação aqui é como um evento abstrato, uma descrição teórica de eventos mentais. Isso quer dizer que dois sinais fisicamente diversos podem estar transmitindo a mesma informação, com o mesmo significado. Visto por esse nível, o processamento de informação fornece a possibilidade de mapear as informações de entrada e as informações de saída. Muito embora as teorias de mapeamento não forneçam como os mapeamentos *input-output* são realizados, elas fornecem conhecimentos que permitem restringir possíveis mecanismos que não são plausíveis, permitindo o teste da teoria, uma vez que essas teorias lidam diretamente com a realidade empírica e podem especificar quando o *input* e *output* terminam de modo a compará-los com o comportamento humano.

2. Decomposição recursiva: qualquer evento informacional unitário pode ser descrito de forma mais analítica, através de sua decomposição em eventos mais simples. Cada um desses eventos também é um evento informacional, e as relações de ordem temporal entre eles especifica como a informação flui através do sistema de componentes. As teorias informacionais que fazem uso dessa suposição são verdadeiras teorias de processamento de informação (modelos de processos cognitivos), porque elas entram na “caixa preta”. É o que Marr (1982) chama de nível algorítmico da teoria. Esse nível, portanto, especifica as operações mentais e aparecem representadas na literatura da psicologia cognitiva em forma de diagramas de fluxo da informação. A decomposição significa que uma operação na memória, por exemplo, possa ser decomposta em vários níveis e representada por diagramas. A natureza recursiva da decomposição significa a construção de uma hierarquia das operações realizadas numa dimensão vertical (operações de alto nível x operações de baixo nível ou descrições gerais x específicas) e numa dimensão horizontal que simplesmente enumera os componentes que pertencem a cada nível do diagrama de fluxo da informação.

3. Continuidade de fluxo: deve estar disponível no *output* das operações toda a informação necessária para realizar cada operação e estas formam a base do *input* da próxima

informação, representada pelos “boxes” do diagrama. Sendo este um corolário da suposição da decomposição.

4. Dinâmica de fluxo: Nenhuma informação do *output* pode ser produzida por uma operação até que sua informação de *input* esteja disponível para realizá-la e tenha passado tempo suficiente para que o *output* seja processado. A dinâmica do fluxo da informação é importante para o teste empírico da teoria. As restrições são que o processamento não pode começar até que algum *input* esteja disponível e toda operação toma alguma quantidade de tempo. Uma suposição adicional é que cada operação ao longo do fluxo constitui um estágio de processamento discreto, suposição amplamente testada a partir do trabalho de “memory scanning” proposto por Sternberg (1966, 1969) e que gerou inúmeros trabalhos subsequentes (Sternberg, 1975; 2001). Outras concepções da dinâmica de fluxo também foram desenvolvidas. Por exemplo, Normam e Bobrow (1975) propuseram uma dinâmica de fluxo mais flexível que a anterior, sugerindo um “*output* disponível continuamente”, mais tarde foi desenvolvida matematicamente por McClelland (1979) e Miller (1988), e ficou conhecida como “processamento em cascata”. Neste tipo de processamento, cada operação começa produzindo algum *output* assim que surge um *input*, de modo que os estágios estão sempre sendo indexados pela presença de *inputs*. Não há uma decisão final sobre qual modelo de processamento de informação é mais plausível. O importante aqui são as condições amplas de testagem desses modelos.

5. Incorporação física: no sistema físico dinâmico cujo comportamento está sendo descrito como um evento informacional, a informação é realizada pelos estados do sistema, chamadas de representação, e as operações são realizadas pelas mudanças de estado, chamadas de processos.

Um outro conjunto de características do processamento de informação, similar ao de Palmer e Kimchi (1986), porém mais resumido, foi proposto por Eysenck e Keane (1994): (I) As pessoas são consideradas seres autônomos e intencionais que interagem com o mundo externo; (II) a mente através da qual elas interagem com o mundo é um sistema de processamento de símbolos de uso geral; (III) nos símbolos atuam vários processos, que os manipulam e os transformam em outros símbolos que, por fim, terão relação com as coisas do mundo externo; (IV) a meta da pesquisa psicológica é a de especificar os processos simbólicos e representações subjacentes ao desempenho de todas as tarefas cognitivas; (V) os processos cognitivos levam tempo; sendo assim, suposições sobre os tempos de reação podem ser feitas ao se presumir que certos processos ocorrem em sequência e/ou possuem alguma complexidade

especificável; (VI) a mente é um processador de capacidade limitada, que tem limitações tanto estruturais quanto de recursos; (VII) o sistema simbólico depende de um substrato neurológico, mas não está inteiramente limitado por este.

Ainda que o processamento de informação não tenha fornecido uma teoria abrangente da cognição humana, ele tem servido como um paradigma ou estrutura útil, ou seja, um conjunto maleável de ideias que podem ser utilizadas na formulação de teorias cognitivas. Destas características, podemos salientar três aspectos importantes em termos das discussões futuras: primeiro, a continuidade da influência do modelo computacional da mente, surgida com os primeiros modelos computacionais de processamento de informação (SGP: Solucionador Geral de Problemas, Newell, Shaw & Simon, 1958); segundo, a continuidade da experimentação em psicologia cognitiva, reavivada com o emprego do tempo de reação (TR) para a medida de processos mentais; terceiro, a afirmação do substrato neural para o processamento de informação, um elemento importante que vem influenciando renomados psicólogos cognitivos a fazerem neurociência cognitiva há algum tempo, como é o caso de Michael Posner & Cols. (Posner, 1992; Posner & Dehaene, 1994; Posner & Petersen, 1990). A esses elementos citados, podemos acrescentar um quarto elemento, que é crítico aos modelos de processamento de informação, que é quanto ao papel exercido pelo contexto cultural na constituição da mente. A base do processamento de informação, portanto, é a noção de mente como um conjunto de computações que ocorrem num substrato neural. Além disso, dessa relação interdisciplinar se espera a reunião de condições para a construção de modelos e teorias psicológicas mais poderosas e de maior cientificidade.

2.6. Os experimentos de Benjamin Libet e sua repercussão na ciência cognitiva

Benjamin Libet foi um cientista considerado pioneiro no campo de estudo da consciência humana, sendo pesquisador no departamento de fisiologia da Universidade da Califórnia, San Francisco. Em 2003, foi o primeiro ganhador do *Prêmio Nobel Virtual em Psicologia* pela Universidade de Klagenfurt, “por suas realizações pioneiras na investigação experimental de consciência, o início da ação, e livre arbítrio”.

Em meados da década de 1970, Libet estava envolvido em investigações sobre atividade neurais e limiares de sensação. Suas investigações iniciais envolviam determinar a quantidade de ativação em locais específicos no cérebro necessárias para acionar sensações somáticas artificiais, contando com uma rotina de procedimentos psicofísicos. Este trabalho logo cruzou

com uma investigação sobre a consciência humana; seu mais famoso experimento foi feito para demonstrar que os processos eléctricos inconscientes do cérebro da chamada *Bereitschaftspotential* (ou potencial de prontidão), descobertos por Lüder Deecke e Hans Helmut Kornhuber em 1964, precedem as decisões conscientes para executar atos voluntários, ou espontâneos, o que implicaria que processos neuronais inconscientes precedem e potencialmente causam os atos volitivos que são retrospectivamente sentidos como sendo conscientemente motivados. Porém, seus experimentos tem causado controvérsia, não só porque desafiam a crença no livre arbítrio, mas também devido a uma crítica dos seus pressupostos envolvidos, como veremos mais a frente. Ademais isso, foi grande inspirador para um estudo mais aprofundado da neurociência do livre arbítrio.

O experimento de Libet (Libet 1985, Libet et al. 1983) consistia basicamente em se buscar estudar as bases neurológicas acerca da capacidade humana de fazer escolhas, ou *agency*. Para isso, buscava-se determinar a quantidade de ativação em áreas específicas do córtex necessária para que desencadeassem sensações somáticas artificiais (i.e., detectar o limiar mínimo para que um estímulo se tornasse consciente). Libet focava a pesquisa no chamado potencial de prontidão (RP), que é a atividade no córtex motor preparatória do movimento e que seria medida através do uso de um eletroencefalograma (EEG), que media a atividade eléctrica no córtex enquanto um eletromiógrafo registrava o movimento muscular usando eletrodos grudados na pele. O paciente sentava-se de frente para um monitor com o relógio do osciloscópio em que havia um ponteiro giratório em sentido horário. Ao mesmo tempo, era pedido aos participantes a execução de movimentos suaves e simples no ato de flexionar o dedo com o intuito de apertar um botão à frente.

Os sujeitos poderiam apertar o botão quantas vezes desejassem e, além disso, tinham de observar a posição dos pontos no relógio do osciloscópio; ao perceberem a mudança na posição do ponto, eles deveriam pressionar o botão. Concomitantemente, o EEG media a atividade eléctrica no córtex motor (potencial de prontidão), isto é, a intenção de agir (pressionar o botão ou não). Libet e seus pesquisadores perceberam que durante o intervalo de detectar a mudança do ponto no relógio e a intenção de agir, havia um atraso de duzentos milissegundos entre a primeira aparição da *vontade* consciente de apertar e o *ato* de pressionar (Libet, 1985). Há aqui dois estágios – a *intenção* de apertar o botão em virtude da detecção da mudança no relógio e o *ato* motor de apertar o botão. Libet referiu-se ao julgamento do tempo de sua “decisão” (‘desejo’, ‘intenção’) para atuar como “julgamento W” e estava interessado na relação do “julgamento W” com o RP.

Em seu artigo “*Unconscious cerebral initiative and the role of conscious will in voluntary action*”, uma vez que os termos “ação voluntária” e “vontade” podiam assumir diferentes significações, Libet propôs-se a clarificar o sentido que esses termos podiam assumir em seus estudos:

Um ato é considerado como voluntário e como sendo uma função da vontade do sujeito quando (a) ele surge endogenamente, não como resposta direta a um estímulo externo; (b) não existem restrições impostas externamente ou compulsões que controlam a iniciação e execução do ato; e (c) os sujeitos sentem introspectivamente que ele está executando o ato pela sua própria iniciativa e que eles são livres para iniciar ou não o ato de acordo com seus desejos (Libet, 1985, p. 529, tradução nossa).

Durante o experimento, foi solicitado ao sujeito que observasse a posição do ponto marcado no temporizador osciloscópio quando “ele/ela pela primeira vez ficou ciente do desejo ou vontade de agir” (os testes de controle com o equipamento de Libet demonstraram uma confortável margem de erro de apenas 50 milissegundos). Ao se pressionar o botão também era registrada a posição do ponto sobre o oscilador por via eletrônica. Ao comparar o tempo marcado quando o botão fora pressionado e a decisão consciente do sujeito para agir, os pesquisadores foram capazes de calcular o tempo total do experimento, desde a volição inicial do sujeito até a ação resultante. Em média, como dito, registrou-se cerca de duzentos milissegundos decorridos entre a primeira aparição de vontade consciente para pressionar o botão e o ato de pressioná-lo.

Os pesquisadores também analisaram as gravações do EEG para cada experimento em relação ao momento da ação. Notou-se que a atividade do cérebro envolvida na iniciação da ação centrava-se principalmente no córtex motor secundário e ocorria, em média, cerca de quinhentos milissegundos antes do experimento terminar com o pressionamento do botão. Ou seja, os pesquisadores registraram uma preparação da atividade cerebral relacionada à ação resultante por volta de trezentos milissegundos antes de os indivíduos relatarem sua primeira vontade consciente para agir. Em outras palavras, as decisões aparentemente conscientes para agir foram precedidas por um acúmulo inconsciente de atividade elétrica no cérebro – a mudança de sinais do EEG que refletem esse acúmulo seria o chamado *Bereitschaftspotential* ou potencial de prontidão. Alguns novos estudos, feitos em 2008 (Soon, Brass, Heize & Hayne, 2008) demonstraram números ainda mais expressivos onde, como resultado de uma tomada de decisão, poderia ser encontrada uma atividade cerebral preparatória no córtex pré-frontal e parietal até 7 segundos antes que o sujeito estivesse ciente de sua decisão.

Mesmo Libet (1985), apesar dos resultados, buscou de alguma forma ainda salvaguardar algum tipo de liberdade para a vontade, introduzindo uma cláusula que ele denominou de “poder de veto” – a capacidade que o agente tem de, após ter tomado ciência da ação selecionada pelo cérebro, poder usar sua “liberdade” para vetá-la. A cláusula permitiria uma margem mínima de “liberdade”, pois neste interstício, o agente poderia vetar ou prosseguir com a ação. Enquanto a consciência não desempenha nenhum papel na instigação de atos volitivos, Libet sugeriu que ela ainda poderia ter um papel a desempenhar na supressão ou retenção de determinados atos instigados pelo inconsciente. Libet observou que todo mundo já experimentou a retenção de realizar um desejo inconsciente. Uma vez que a experiência subjetiva da vontade consciente para agir precedeu a ação por apenas 200 milissegundos, o que deixaria à consciência apenas 100-150 milissegundos para vetar uma ação (isto é porque os últimos 20 milissegundos antes de um ato são ocupados pela ativação dos neurônios motores na coluna vertebral pelo córtex motor primário, e a margem de erro indicada por testes utilizando o oscilador também deve ser considerada).

Esses experimentos de Libet sugerem a alguns, como por exemplo Wegner (2002), que os processos inconscientes no cérebro são o verdadeiro iniciador de atos volitivos e o livre arbítrio, portanto, não desempenha nenhum papel em sua iniciação. Se os processos cerebrais inconscientes já tomaram medidas para iniciar uma ação antes mesmo que a consciência tenha tomado conhecimento de qualquer desejo para executá-la, desta maneira, segundo esta interpretação, o papel causal da vontade consciente seria totalmente eliminado. Susan Blackmore (1998), na mesma linha que Wegner, oferece a seguinte argumentação: “a experiência consciente leva algum tempo para se construir e esse tempo é muito lento para ser responsável por fazer as coisas acontecerem” (p. 707).

Sugeriu-se, portanto, que a consciência seja apenas um efeito colateral de funções neuronais, um epifenómeno de estados cerebrais, e os experimentos de Libet são proferidos em apoio a esta teoria; a impressão que temos da iniciação consciente de nossos próprios atos são, seguindo esse ponto de vista, um erro de retrospecção. No entanto, alguns filósofos contestaram esta conclusão:

Em suma, as causas [neuronais] e correlatos da experiência consciente não devem ser confundidas com a sua ontologia [...] a única evidência sobre o que as experiências conscientes são provém de fontes em primeira pessoa, que consistentemente sugerem que a consciência seja algo outro que não apenas adicional para a atividade neuronal. (Velmans, 2000, pp 35-37, tradução nossa)

Uma crítica mais geral a partir de uma perspectiva dualista-interacionista foi levantada por Alexander Batthyany (2009) em relação ao pedido de Libet a seus sujeitos que apenas “deixassem que o desejo [para se mover] aparecesse por si mesmo a qualquer momento sem qualquer pré-planejamento ou concentração em quando agir” (Libet et al., 1983, p. 323). De acordo com Batthyany, nem teorias reducionistas ou não reducionistas da *agency* afirmam que os impulsos que aparecem por conta própria são exemplos adequados de (supostos) eventos conscientemente causados porque não se pode esperar passivamente que um desejo ocorra sendo, ao mesmo tempo, aquele que o está trazendo-o conscientemente à tona. Os resultados de Libet, assim, não poderiam ser interpretados como fornecedores de evidência empírica em favor de uma teoria reducionista da *agency*, uma vez que as teorias não reducionistas, incluindo até mesmo o interacionismo dualista, poderiam prever os mesmos resultados experimentais.

Daniel Dennett (2003) argumenta que nenhuma conclusão clara sobre volição pode ser derivada dos experimentos de Libet em motivo de uma ambiguidade na cronometragem dos diferentes eventos envolvidos. De acordo com Dennett, Libet diz quando o potencial de prontidão ocorre objetivamente, utilizando-se de eletrodos, mas depende do sujeito para relatar a posição do ponteiro de um relógio para determinar quando a decisão consciente foi feita. Como aponta Dennett, este é apenas um relato de onde parece ao sujeito que várias coisas se juntam, não o tempo objetivo em que elas realmente ocorrem.

Supondo que Libet soubesse que o potencial de prontidão atingiu um pico à 6810 milissegundo do ensaio experimental, e que o ponteiro do relógio foi direto para baixo (que é o que você relatou que viu) em 7005 milissegundos. Quantos milissegundos ele deveria acrescentar a este número para obter o tempo onde você estava consciente disso? A luz começa a partir do relógio para o seu globo ocular quase que instantaneamente, mas o caminho dos sinais partindo da retina através do núcleo geniculado lateral para o córtex estriado leva de 5 a 10 milissegundos - uma fração insignificante dos 300 milissegundos de compensação, mas quanto tempo isso toma até chegar a você. (Ou você está localizado no córtex estriado?) Os sinais visuais têm de ser processadas antes de chegar aonde quer que eles precisem chegar para que você faça uma decisão consciente da simultaneidade. O método de Libet pressupõe, em suma, que nós podemos localizar a intersecção de duas trajetórias:

- *a tomada de consciência dos sinais que representam a decisão de apertar o botão*
- *a tomada de consciência dos sinais que representam as sucessivas orientações dos ponteiros do relógio*

de modo que esses eventos ocorram lado-a-lado como se estivessem no lugar onde sua simultaneidade pudesse ser notada. (Dennett, 2003, p.231, tradução nossa)

A teoria inicial de Libet (Libet et al., 1979), que consistia no estudo de estímulos e sensações, foi considerada por alguns comentadores, entre eles Patricia Churchland (1981), como oferecendo alguns aspectos curiosos, ou mesmo estranhos, devido à aparente ideia de uma “causalidade reversa”. Libet (1981), em resposta a Patricia Churchland, argumentou que

os dados sugeriram que nós retrospectivamente “antecipávamos” o início de uma sensação para o momento da resposta neuronal primária. Diferentes pensadores interpretaram o trabalho de Libet sobre estímulo e sensação de maneiras diversas. John Eccles (1985) apresentou o trabalho de Libet como se esse sugerisse que uma mente não física desse um passo para trás no tempo. Edoardo Bisiach (1988) descreve a visão de Eccles como tendenciosa, mas acrescenta:

Esta é certamente a conclusão de que os autores (Libet, et al.) parecem estar dispostos a forçar sobre o leitor. [...] Eles disputam uma explicação alternativa, sugerida por Mackay em uma discussão com Libet (1979, p. 219) no sentido de que “a referência subjetiva para trás no tempo pode ser devida a um julgamento ilusório feito pelo sujeito quando relata a cronometragem”, e mais significativo ainda, Libet et al. (1979, p. 220) sugerem que “embora grave essas dificuldades não são insuperáveis” para a teoria da identidade (de mente e matéria) produzida por seus dados. (p. 78, tradução nossa)

Posteriormente, Libet (2004) concluiu que não parecia haver nenhum mecanismo neural que pudesse ser visto como mediador ou que estivesse computando as referências sensoriais subjetivas para trás no tempo. Libet postulou que o primeiro potencial evocado (EP), serviria como um “marcador de tempo”. O EP é um potencial positivo agudo que aparece na região sensorial adequada do cérebro em cerca de 25 milissegundos após um estímulo na pele. Os experimentos de Libet (Libet, et al., 1979) demonstraram que existe um encaminhamento subjetivo automático da experiência consciente para trás no tempo para este marcador de tempo. A sensação na pele não aparece em nossa consciência até cerca de 500 milissegundos após o estímulo da pele, mas nós subjetivamente sentimos que a sensação ocorreu no momento do estímulo. Para Libet, estas referências subjectivas parecem ser puramente uma função mental sem base neural correspondente no cérebro. Com efeito, esta sugestão pode ser mais amplamente generalizada:

A transformação dos padrões neuronais para uma representação subjetiva parece desenvolver-se em uma esfera mental que surgiu a partir desse padrão neuronal. [...] Minha visão da função subjetiva mental é que ela é uma propriedade emergente das funções cerebrais adequadas. O mental consciente não pode existir sem os processos cerebrais que dão origem a ela. No entanto, tendo emergido a partir das atividades do cérebro como uma “propriedade” exclusiva desse sistema físico, o mental pode exibir fenômenos não evidentes no cérebro neural que o produziu. (Libet, 2004, p. 86-87, tradução nossa)

Na parte final de sua carreira, Libet (2004) propôs ainda o que ele chamou de teoria do campo mental consciente (CMF), buscando explicar como o mental surge do cérebro físico. Segundo ele, as duas principais motivações que deram origem a esta proposta foram: (1) o fenômeno da unidade da experiência consciente subjetiva e (2) o fenômeno onde a função mental consciente parece influenciar a atividade das células nervosas.

No que diz respeito à unidade da experiência consciente, era cada vez mais evidente a Libet (2004) que muitas das funções do córtex são localizadas, mesmo a um nível microscópico, a alguma região do cérebro, e ainda assim as experiências conscientes relativas a estes domínios aparecem sempre ao sujeito como integradas e unificadas. Nós não experienciamos uma variedade infinita de eventos individuais, mas sim uma consciência integrada e unitária, por exemplo, sem lacunas em imagens coloridas e espaciais. Para Libet, algum fenômeno ou processo unificador provavelmente medeia a transformação de representações neuronais localizadas e particularizadas em nossa experiência consciente unificada. Este processo parece ser melhor explicado ou entendido em uma esfera mental que parece emergir dos eventos neurais, ou seja, o campo mental consciente.

O CMF seria então, para Libet (2004), o mediador entre as atividades físicas de células nervosas e o surgimento da experiência subjetiva. Assim, o CMF seria a entidade na qual a experiência subjetiva unificada está presente e fornece a capacidade causal para afetar ou alterar algumas funções neuronais. Libet propôs o CMF como uma “propriedade” de um fenômeno emergente do cérebro; ele não existe sem o cérebro, mas sim emerge de um sistema adequado de atividade neural. Esta proposta está relacionada às teorias eletromagnéticas de consciência (McFadden, 2002).

Para testar a proposta habilidade causal do CMF de afetar ou alterar funções neuronais, Libet (2004, 2006) propôs um desenho experimental onde se isolasse cirurgicamente um pedaço do córtex cerebral (em um paciente para o qual um tal procedimento seja terapeuticamente necessário). Se a estimulação elétrica do córtex isolado puder evocar um relato introspectivo pelo sujeito, o CMF deve ser capaz de ativar áreas cerebrais adequadas, a fim de produzir um relato verbal. Este resultado seria uma prova direta de que um campo mental consciente poderia afetar as funções neuronais de uma maneira que seria responsável pela *atividade da vontade consciente*. Uma descrição do teste experimental proposto é a seguinte:

Um pequeno pedaço do córtex sensorial (que auxilia uma modalidade qualquer) é neuralmente isolado, mas se mantém viável por fazer todos os cortes corticais subpialmente. Isso permite que os vasos sanguíneos na pia sejam projetados para o pedaço isolado e forneça o fluxo de sangue dos ramos arteriais que mergulham verticalmente no córtex. A previsão é que a estimulação elétrica do pedaço sensorial produza uma resposta subjetiva que possa ser relatada pelo sujeito. Isto é, a atividade em um pedaço isolado pode contribuir para produzir a sua própria porção do CMF. (Libet, 2006, p. 324, tradução nossa)

Logo abaixo, Libet oferece um maior aprofundamento sobre o CMF:

A CMF não é um fenômeno dualista cartesiano; não é algo separado do cérebro. Em vez disso, propõe-se a ser uma propriedade de um sistema localizável produzido pelas atividades neuronais apropriadas, e não pode existir sem elas. Mais uma vez, não é um “fantasma na

máquina”. Mas, como um sistema produzido por bilhões de células nervosas em ação, isto pode ter propriedades não diretamente previsíveis a partir destas atividades neuronais. É um fenômeno não físico, como a experiência subjetiva que ela representa. O processo pelo qual o CMF surge a partir dos elementos que contribuem para isso não é descritível. Ele deve simplesmente ser considerado como um novo fenômeno natural fundamental “dado”, que é diferente de outros “dados” fundamentais como a gravidade ou o eletromagnetismo. (Libet, 2006, p. 324, tradução nossa)

Porém, grande parte dos seguidores de Libet tomaram principalmente seus primeiros experimentos sobre vontade consciente e atos voluntários como uma prova definitiva de que não há liberdade de escolha. Alguns neurocientistas atualmente prosseguem com este raciocínio, a saber, de que não há consciência plena e irrestrita durante o processo de decisão. Recentemente, Fried et al. (2011) repetiu o experimento de Libet em pacientes com epilepsia intratável usando eletrodos intracranianos diretamente ligados ao tecido cortical na área frontal medial 12. A atividade registrada nestas áreas indicava potencial de prontidão anterior ao movimento voluntário, confirmando o experimento de Libet. Além disso, quando essa mesma área foi estimulada, o relato dos pacientes era de um “sentimento urgente de mover uma parte particular do corpo”, sem a necessidade de executar o movimento. Em outro experimento, desta vez em macacos, descobriu-se que um pequeno subconjunto de neurônios no córtex medial frontal (área motor suplementar) tinha atividade crescente antes do movimento, ou seja, os neurônios se ativavam antes de o macaco preparar o movimento para ação (Shima & Tanji, 2000). O tempo da intenção consciente poderia ser predito a partir da ativação deste subgrupo de neurônios confirmando novamente o achado de Libet.

Haggard (2011) alega que estes dados experimentais provam que a intenção consciente é apenas um corolário subjetivo da ação a ser executada. Neste sentido, a intenção do agente seria um “epifenômeno subjetivo” que teria apenas a impressão de iniciar a ação, porém, o epifenômeno em si, não causaria ação alguma. Haggard baseia seu argumento nesta experiência e afirma que a ação voluntária é controlada e determinada por uma rede de neurônios em nível inconsciente sem que o sujeito tenha acesso; antes de o autor pensar em executar a ação o cérebro já enviou o comando. Em outras palavras, o agente tem acesso apenas ao resultado do que foi *decidido* pelo cérebro e, segundo esta hipótese, o cérebro apenas comunica ao “agente” sua decisão.

Haynes et al. (2007) também relatam que atividades em certas áreas cerebrais predeterminam a intenção consciente e o tempo que levam para iniciar a decisão motora. O fMRI e o EEG seriam ferramentas capazes de dizer, por exemplo, se a intenção do sujeito era mover a mão direita ou esquerda. Os participantes do experimento tinham de executar uma

tarefa de decisão motora em etapas, enquanto a atividade cerebral era medida com fMRI. Semelhante ao experimento de Libet, os participantes se sentavam e fixavam o olhar para uma tela, na qual uma sequência de letras era apresentada e, a seguir, os participantes deveriam apertar um botão tão logo as letras surgissem. Em algum momento, eles tinham de reportar se sentiam a necessidade de pressionar um ou outro botão, segundo as letras que apareciam na tela em sequência num intervalo de quinhentos milissegundos.

O mapa neuronal de atividades mostrava que a ação era codificada nos córtex parietal e frontal em até dez segundos antes de a informação se tornar acessível, ou melhor, disponível conscientemente para o agente. As imagens de fMRI deste experimento indicavam intensa atividade nas áreas corticais supracitadas, sugerindo que aquelas redes neuronais processavam estímulos referente à decisão. Quando o sujeito estava prestes a tomar uma decisão, a ativação do córtex frontal e parietal já indicava que o cérebro havia iniciado os procedimentos, enviando o comando para a área motora. Baseado neste experimento, Haynes et al. (2007) sugerem a possibilidade de prever o pensamento do sujeito, pois dependendo da área ativa, é possível saber que *tipo* de pensamento o sujeito está tendo, embora o conteúdo possa permanecer desconhecido.

De acordo com Sousa (2014), uma crítica que pode ser feita, tanto aos experimentos de Libet sobre atos voluntários, e que poderia ser estendida a alguns dos experimentos referidos acima, se trata da questão do pré-planejamento implícito nestes. Com a ressalva de que ao mesmo tempo que se deve ressaltar a importância do planejamento e da estruturação racional de tarefas não se pode, apenas por isso, deduzir daí nem livre arbítrio nem causação mental. E, mesmo que as capacidades de controle e seleção não disparem a ação, poderiam funcionar como mecanismos subtrativos onde diferentes possibilidades são restringidas em favor de uma determinada. Mesmo assim não haveriam provas suficientes para descartar a relevância do pré-planejamento. Godefroy & Rousseaux (1997) e D'Esposito & Gazzaley (2006), entre outros, mostram que os lobos frontais são responsáveis por funções executivas relacionadas à seleção e ao planejamento de ações e que portanto devem ser tomadas como relevantes para os processos de decisão operando controle descendente. Mesmo Libet (1982) dará atenção a este tópico de alguma forma reconhecendo a importância do pré-planejamento. Aqui, pré-planejamento significa seleção e organização de atividades futuras. E isso não pode ser subestimado no curso de ações não automáticas ou não-reflexas como seria o caso de ações livremente escolhidas.

O simples ato de flexão de dedo ou pulso, chamado por Libet de “atos livremente voluntários” seriam, de acordo com Sousa (2014), mais como “atos de resposta a uma tarefa”. Estes atos são procedidos pelo potencial de prontidão. De qualquer forma, os pacientes estão apenas cumprindo uma série de tarefas ou ordens requeridas pelo experimentador em um contexto laboratorial de pouca relevância ecológica. Assim, pode-se argumentar que os experimentos de Libet não estariam pesquisando livre arbítrio, mas alguma forma muito simples de processo de decisão. Da mesma forma, os sujeitos que realizam a ação estão conscientes do que se deve fazer bem antes do começo do experimento. Mais que orientado pelo experimentador, o paciente é treinado. Então, obviamente, se aquilo que se passa no cérebro durante o processo prévio de treinamento não está sendo também monitorado, não podemos entender de forma mais completa e fidedigna que áreas do cérebro estão envolvidas nos processos de decisão. Estes sujeitos que estão operando a flexão do dedo ou do pulso nos experimentos de Libet certamente compreenderam a tarefa requerida pelo experimentador. Desta forma, podemos dizer que estes sujeitos já têm um certo nível de consciência de como agir e assim algum pré-planejamento. Podemos dizer que, se eles entendem a tarefa que lhes é requisitada, bem como as condições nas quais esta deve ser exercida, estes sujeitos já formaram uma intenção para agir. De qualquer forma, o processo de decisão presente nos experimentos de Libet começa bem antes do experimento propriamente dito, durante o processo de explicação e treinamento.

Não devemos esquecer também que existem diversos casos totalmente diferentes de decisão. Quando nós automaticamente pisamos no freio para evitar a colisão de nosso carro com o garoto que corre atrás de uma bola, como no exemplo de Libet (2004), não estamos exercendo uma ação consciente. Mesmo sabendo que não há tempo para que essa ação seja uma ação consciente não poderíamos negar que aqui se dá algum processo de decisão. Nesse tipo de processo de decisão, obviamente seria mais problemático a afirmação de que estamos agindo segundo uma vontade livre, já que se trata de uma decisão automática. Considerando que supostamente temos diferentes níveis de processamento de informação no cérebro, teríamos como uma implicação natural que podemos ter diferentes níveis de decisão. Estes processos envolvidos em notar o garoto e pisar no freio não são tão semanticamente carregados como o caso de decisões complexas como a de se casar com uma determinada mulher ou votar por este ou aquele partido ou candidato.

CAPÍTULO III

O PROBLEMA DA VONTADE NA PSICOLOGIA COGNITIVA

Tendo apresentado no primeiro capítulo o problema da vontade e livre arbítrio, além de vários questionamentos que surgem daí, como a questão da responsabilidade moral, o determinismo e o compatibilismo/incompatibilismo, de forma mais geral e resumida, dentro do espaço e tempo a que nos cabe essa investigação e; no segundo, uma delimitação do que chamamos de psicologia cognitiva, além de tentarmos deixar mais explícitos seus pressupostos epistemológicos, vemos que em muitos pontos esses pressupostos tocam questões fundamentais para o debate sobre o problema da vontade e livre arbítrio, como a questão da eficácia causal dos processos mentais e sua relação com o comportamento, ou de como a mente funcionaria como uma espécie de máquina de computar, seguindo regras (que podem se adaptar ou acomodar ao contexto, desenvolverem-se de certa forma, através do teste com a realidade), ou ainda como processos neurais inconscientes podem estar na origem de um ato voluntário, da tomada de decisão, etc.

Neste terceiro capítulo apresentamos mais de perto o que a psicologia cognitiva tem a dizer, de forma mais direta, sobre o conceito da vontade, e seu posicionamento sobre a questão da liberdade da vontade, ou livre arbítrio. Começamos pela análise de duas obras que foram fundamentais para a psicologia cognitiva e onde se trata de questões relevantes acerca do problema da vontade e livre arbítrio. A primeira delas, *Plans and the Structure of Behavior* (1960), escrita por George Miller, juntamente com Eugene Galanter e Karl Pribram, é um marco importante para o cognitivismo, e apresenta um reconhecimento sobre como o problema da vontade humana seria um problema central na também nascente psicologia cognitiva. Na segunda obra que gostaríamos de destacar, *Cognition and Reality* (1976), Ulric Neisser oferece algumas considerações sobre a liberdade humana e controle e predição do comportamento. Posteriormente, buscamos investigar como o conceito de vontade e livre arbítrio foi definido em dicionários e enciclopédias de psicologia e ciência cognitiva, para termos uma visão mais atual desses conceitos nos dias de hoje. Fizemos, também, uma incursão mais geral na área de estudo da motivação, devido a abrangência da área e os vários modelos existentes de abordagem dos motivos da ação humana, além de apresentarmos o conceito neuropsicológico de função executiva, muito em voga em estudos recentes da psicologia cognitiva e que é relacionado aos atos voluntários, controle cognitivo global, planejamento e execução do comportamento. E, por

fim, levantamos a questão de se seria possível falar em uma vontade livre partindo-se de uma visão cognitivista.

3.1. *Plans and the structure of behavior*

George A. Miller, psicólogo estadunidense, considerado como um dos fundadores da psicologia cognitiva, se tornou muito conhecido por seu experimento clássico e muito influente na história da psicologia, onde evidenciou que a memória de curto prazo tem uma capacidade limitada a aproximadamente sete itens. Foi igualmente sustentado que, se os itens forem combinados em agrupamentos (*chunks*) distintos, podemos lembrar mais itens. Por exemplo, não conseguimos lembrar uma série de 20 dígitos seguidos, após ouvi-la somente uma vez. Porém, se dividirmos esses números em uma série de cinco agrupamentos de quatro dígitos, cada um deles formando uma data diferente, então seríamos capazes de recordá-los (Miller, 1956).

Outro exemplo clássico de *chunks* é a capacidade de recordar sequências longas de números binários, pois eles podem ser codificados na forma decimal. Por exemplo: a sequência 0010 1000 1001 1100 1101 1010 poderia ser facilmente lembrada como 2 8 9 C D A. Isto, obviamente, só funcionaria para alguém que pudesse converter números binários em números hexadecimais, mostrando como os *chunks* são “unidades significativas”. Um *chunk* pode se referir a dígitos, palavras, posições de xadrez, ou rosto de pessoas. O conceito de *chunks* e a capacidade limitada da memória de curto prazo se tornaram um elemento básico em todas as teorias de memória posteriores. Como isso, ao dar ênfase as limitações da capacidade humana de receber e processar informações, Miller ofereceu mais um dado importante para o estabelecimento da ideia de que a mente é um sistema de processamento de informação (Lachman, Lachman & Butterfield, 1979).

Posteriormente, Miller, junto com Eugene Galanter e Karl Pribram, estabeleceu outro marco importante para o cognitivismo, com o livro *Plans and the Structure of Behavior* (1960). Ao propor uma síntese das influências de Turing, da cibernética, da teoria da informação e de Chomsky, entre outras, os autores propõem que o comportamento é determinado por representações internas, estruturadas hierarquicamente, de acordo com planos bem definidos. Introduziram também o conceito de meta que, reinserido ao cenário psicológico, abre espaço novamente para uma explicação teleológica do comportamento. Nesse sentido, todo o

comportamento humano é *prospectivo* e visa atingir *metas* através de *planos* e *estratégias* de ação consciente (Castañon, Justi & Araujo, 2014).

Como estabeleceram Miller, Galanter e Pribram (1960), um plano pode ser definido de maneira rigorosa como um processo hierárquico de sequências de operações a serem executadas por um organismo, da mesma forma como um programa para um computador. Um conceito importante, proposto pelos autores, para se entender esse processo, é o modelo TOTE (*test-operate-test-exit*). Miller et al. sugeriu que o TOTE deveria substituir o modelo estímulo-resposta como a unidade básica de comportamento. Em uma unidade TOTE uma meta é testada para ver se ela foi alcançada e, se não, uma operação é executada para alcançar a meta. Este ciclo de testar-operar seria repetido até que a meta seja eventualmente alcançada ou abandonada. O conceito TOTE forneceu a base de muitas teorias posteriores de resolução de problemas (como, por exemplo, o GPS) e sistemas de produção. Um exemplo clássico de um TOTE é um plano para martelar um prego. O *Exit Test* (Teste de Saída) é se o prego está rente à superfície. Se estiver, então o martelo é testado para verificar se está em cima (caso contrário ele é levantado), para poder bater no prego. Assim, parte do *output* volta como *input*, de forma a permitir que um sistema funcionando a base desse modelo calcule a margem de erro entre a meta estabelecida e sua posição atual, o que permitiria ao sistema ajustar seu comportamento (*output*) em relação à meta.

Portanto, segundo o modelo proposto por Miller et al. (1960), o planejamento (na forma de unidades TOTE) seria considerado um processo cognitivo fundamental. Onde o comportamento pode ser visto como hierarquicamente organizado (por exemplo, através de *chunks*, unidades TOTE) e a mente como um sistema de processamento de informação que controla o comportamento. Ainda assim, conforme admitido por Miller et al., suas preferências teóricas estão todas do lado dos teóricos cognitivos (p. 09). Porém, apesar disso, admite ainda que há algo na posição cognitiva cujas críticas não receberam uma resposta satisfatória – a crítica de que os processos cognitivos tal qual Tolman e outros postularam não são de fato suficientes para explicar o comportamento.

A posição de Tolman é a de que a descrição da estrutura cognitiva seria indispensável para o entendimento do comportamento, posição esta compartilhada por Miller et al. (1960), e que pode ser melhor compreendida na seguinte passagem de Tolman (1948):

[O cérebro] é muito mais parecido com um centro de controle do mapa do que uma central de telefone à moda antiga. Os estímulos não estão conectados por apenas simples interruptores de um-para-um com as respostas enviadas. Em vez disso, os impulsos de entrada são

normalmente trabalhados e elaborados na sala de controle central em uma tentativa, na forma de um mapa cognitivo do ambiente. E é esta tentativa, indicando rotas e caminhos e relações ambientais, que finalmente determina quais respostas, se houver alguma, que o animal vai finalmente liberar. (Tolman, 1948, p.193; apud Miller et al., 1960, p. 8, tradução nossa)

Apesar de admitir ser simpático a esse tipo de teorização, onde Miller et al. (1960) diz ser óbvio que há muito mais acontecendo entre estímulo e resposta do que pode ser explicado por uma simples declaração sobre forças associativas. Todavia, há uma importante crítica a essa posição cognitiva que ainda não foi adequadamente respondida. Mesmo se você admitir, como dizem os críticos, esse “algo fantasmagórico interno”, você não vai ter explicado satisfatoriamente o comportamento. Como apontado por Miller et al., a forma penetrante como Guthrie (1935) coloca essa questão nos é muito proveitosa:

Signos, na teoria de Tolman, ocasionam a realização do rato, ou cognição, ou julgamento, ou hipóteses, ou abstração, mas eles não ocasionam sua ação. Em sua preocupação com o que se passa na mente do rato, Tolman tem negligenciado a predição do que o rato vai fazer. Como, até agora, a teoria está preocupada em como o rato está enterrado em seus pensamentos; se ele consegue a caixa de alimentos no final [do labirinto] isso é uma preocupação do rato, não a preocupação da teoria. (Guthrie, 1935, p. 172; apud Miller et al., 1960, p. 9, tradução nossa)

Para Miller et al. (1960), parece que os teóricos cognitivos não entenderam a força dessa crítica. Se um rato faminto sabe onde achar comida – se ele tem um mapa cognitivo dizendo onde a caixa de comida está localizada – ele irá até lá e a comerá. Parece que não há nada mais aí para se explicar. No entanto, “o *gap* entre conhecimento e ação parece menor que o *gap* entre estímulo e resposta – mas ele continua ali, continua indefinidamente grande” (Miller et al., 1960, p. 09). Os teóricos cognitivos utilizaram desse argumento como o melhor curso para mostrar que a teoria dos reflexos era inadequada para explicar o comportamento, porém, segundo o próprio Miller et al., eles parecem igualmente despreparados quando o mesmo argumento é direcionado contra eles próprios – as coisas se mostraram mais complicadas do que eles se atreveram a imaginar. Assim, se Guthrie estiver correto, cada vez mais teoria cognitiva seria necessária, mais do que os teóricos cognitivos estão providos. Isso quer dizer que, longe de respeitar a Navalha de Occam, o teórico cognitivo necessita de mais bagagem teórica para carregar, talvez mais do que ele possa carregar. E algo é preciso para preencher a lacuna entre conhecimento e ação.

Segundo Miller et al. (1960), muitos psicólogos, estando eles próprios inclusos, foram perturbados por esse vácuo teórico entre cognição e ação. Podemos nos perguntar, nesse ponto, o que esse vácuo teórico apontado por Miller et al. significa para o objetivo do nosso presente trabalho, deixaremos o próprio Miller et al. apresentar uma resposta:

Sem dúvida, é perfeitamente óbvio para o leitor que temos aqui uma versão moderna de um

antigo enigma. Em uma data anterior poderíamos ter introduzido o tópico diretamente ao anunciar que pretendemos discutir a vontade. Mas hoje a vontade parece ter desaparecido da teoria psicológica, assimilada anonimamente no tópico mais amplo da motivação. (p. 11, tradução nossa)

Como podemos ler na passagem acima, o problema da vontade, apesar de ter desaparecido da teoria psicológica, ao menos em um nível terminológico, é um problema que ainda se impõe. A psicologia cognitiva, ao tentar oferecer uma saída para o problema do *gap* explicativo entre estímulo-resposta, introduzindo aí a cognição como fator explicativo para o comportamento, não consegue apropriadamente preencher esse *gap*, visto ainda existir o mesmo problema entre cognição e ação. Ao mostrar como se dá a cognição, ou os processos, leis, regras envolvidas em seu funcionamento, ainda assim não se explicaria o comportamento. Obviamente pode-se entender de forma mais acurada o que estava envolvido naquela forma de se comportar. Mas, conhecer não equivale a agir. Há ainda a necessidade, seguindo esse raciocínio, de que algo aja, mesmo que em função desse conhecimento, ou da informação processada, há a necessidade de um sujeito que faça uso desse conhecimento. “O problema é descrever como as ações são controladas por uma representação interna do organismo de seu universo” (Miller et al., 1960, p.12)

Podemos conjecturar ainda que há a necessidade de uma força que mova o organismo, ou seja, que promova a ação em função desse conhecimento. Nesse sentido entendemos o problema da vontade como uma formulação teórica que busca explicar a causa do comportamento. Todavia, como fora observado anteriormente, daí surgem todos os problemas envolvidos nesse tipo de formulação, como a questão da relação mente/corpo, ou como o mental pode ter influência causal no físico, se essa força que move a ação é determinada ou não e, se não, como compatibiliza-la com uma visão determinista de mundo. Mesmo ignorando, ou tentando ignorar esse antigo enigma, mesmo que se tente evitar todo esse debate já por demais sobrecarregado, que atravessa questões de cunho filosófico das mais abrangentes, ainda que tentando dissolver o problema da vontade em micro teorias explicativas num tópico sobre motivação, pela natureza do que a psicologia se propõem estudar, esse problema irá sempre se impor.

Como Miller et al. (1960) reconheceu, a psicologia cognitiva de sua época ainda buscava preencher essa lacuna explicativa entre cognição e comportamento. Nos parece que ainda hoje essa constatação de Miller et al. continua pertinente. Para tanto, devemos ainda investigar dentro do tópico da motivação, o que podemos encontrar relativo ao problema pesquisado. O leitor pode encontrar tal investigação mais à frente. Por agora, vamos abordar a visão que Ulric

Neisser oferece relativo à liberdade humana, controle e predição do comportamento e que está diretamente relacionada com o problema investigado.

3.2. *Cognition and Reality*

Ulric Neisser foi outro autor de grande importância para a nascente psicologia cognitiva. Podemos dizer que a nova disciplina se consolidou com o lançamento de seu primeiro manual, *Cognitive Psychology* (1967). Logo no prefácio, Neisser esclarece o que entende por cognição, sendo esta todos os processos envolvidos na transformação e armazenamento do *input* sensorial: sensação, percepção, retenção, etc. Define, então, a psicologia cognitiva, parafraseando Freud, como o estudo da “informação do estímulo e suas vicissitudes” (p. 4). Prosseguindo, explica o chiste:

Como é usado aqui, o termo ‘cognição’ se refere a todos os processos pelos quais o input sensorial é transformado, reduzido, elaborado, estocado, recuperado e usado. Ele diz respeito a estes processos mesmo quando eles operam na ausência de estimulação relevante, como em imagens ou alucinações. Termos como sensação, percepção, imaginação, retenção, recordação, resolução de problemas e pensamento, entre muitos outros, referem-se a estágios hipotéticos ou aspectos da cognição (1967, p. 4, tradução nossa).

De tal modo, o escopo da psicologia cognitiva seria o de estabelecer as *regras* de transformação da informação que chega ao sujeito como *input* e sai como *output*. Tais regras podem ter a forma de procedimentos de manipulação de símbolos, processos de satisfação de restrições, estratégias para realização de metas e estruturas cognitivas. Essas regras, estratégias e estruturas não são formas fixas, não permanecem inalteradas desde o nascimento, ou seja, podemos investiga-las tanto em seu estado maduro quanto através de seu processo de desenvolvimento. Esse é um ponto importante, e está relacionado à visão de Neisser sobre um possível aspecto que possibilitaria falar em liberdade no comportamento humano, e que foi apresentado em sua obra posterior, *Cognition and Reality* (1976). Além disso, deixa claro também que, com o lançamento de seu manual, busca integrar as contribuições da teoria da informação, dos estudos de Miller e das simulações computacionais de Newell e Simon, entre outras (Castañon, Justi & Araujo, 2014).

Ainda em seu manual, *Cognitive Psychology* (1967), assume a proposta filosófica de Putnam devido a utilidade da metáfora computacional para a psicologia, que, embora feita com ressalvas pôr a considerar inadequada “de muitos modos” (p. 6), permitiria falar em dois níveis diversos de análise: sendo hardware/software como análogos ao físico/psicológico. Deste modo, o objeto primário de investigação do psicólogo não seria o registro físico dos dados, mas

sim o funcionamento dos “programas”, as regras de processamento de informação. O que distinguiria a psicologia da neurofisiologia: pois caberia ao psicólogo principalmente entender dos dados “sua utilização, não sua encarnação” (p. 6).

Nove anos depois de publicar o livro que lançou o campo, Neisser publica *Cognition and Reality* (1976), no qual exprime sua insatisfação com o que considerava um estreitamento da posição cognitiva com a tendência de acentuar situações artificiais de laboratório em lugar do mundo real. De certa forma desiludido, considerava que o movimento da psicologia cognitiva tal como existente em meados da década de setenta não mantinha muito de sua função principal que seria proporcionar uma melhor compreensão da maneira como os seres humanos enfrentam a realidade. Via também de forma mais crítica a aplicação da teoria do processamento de informação aos processos psicológicos, da mesma forma que a metáfora computacional. A crença de que os processos cognitivos se desenvolvem de acordo com leis, que podem ser descobertas e representadas, segundo Neisser (1967, 1976), seria mais adequada principalmente aos processos mais simples e automáticos da cognição, e não em relação aos processos considerados superiores (como criatividade e atribuição de significado).

Um ponto que mais nos interessa aqui é a discussão apresentada por Neisser, no nono capítulo de *Cognition and Reality*, sobre algumas consequências da cognição. Segundo ele, é frequentemente argumentado que o progresso na psicologia poderia representar um perigo para a liberdade pessoal. Pois, se a psicologia for uma ciência, sua meta seria então entender os princípios gerais da natureza humana. Neisser então se pergunta se, armados com esses princípios, não estaremos aptos a controlar nosso objeto de estudo? Segundo ele, muitos tomaram ou tomam essa possibilidade seriamente. Skinner, por exemplo, repetidamente sugeriu que métodos de controle comportamental já estariam disponíveis e que podemos utiliza-los para aperfeiçoar a condição humana se não formos acanhados o suficiente para coloca-los em prática (por exemplo, em Skinner, *Beyond Freedom and Dignity*, 1971).

Na opinião de Neisser, porém, tanto as esperanças quanto os medos relativos a noção de controle comportamental são mal entendidos: “os fatos da cognição implicam que a manipulação psicológica do comportamento está fadada ao fracasso – pois não podem levar a resultados sistematicamente previsíveis sob condições culturais comuns” (1976, p. 177). Porém, isto não é o mesmo que falar que o comportamento não pode ser predito ou manipulado já que todos utilizamos de uma grande dose de predição de comportamento todos os dias e mesmo de manipulação. A sociedade, de forma geral, dificilmente poderia existir se assim não o fosse. Em nossa relação com o outro, naturalmente fazemos predições de seu comportamento,

além de casos não tão raros de conhecimento geral onde pessoas manipulam outras para o bem próprio, etc.

No entanto, de acordo com Neisser (1976), o fato de que a manipulação pode ter sucesso eventualmente em si não prova que se o manipulador tivesse um maior conhecimento de psicologia teria maior sucesso em sua manipulação; o alcance do controle comportamental pode ter limites intrínsecos. Antes, porém, de decidir se o comportamento pode ser controlado por meio psicológicos, Neisser propõem que consideremos como ele é controlado normalmente, isto é, do que esse controle depende? De forma simplificada, podemos reduzir o controle do comportamento como ele normalmente ocorre em duas opções: ou o indivíduo controla o que ele mesmo vê ou faz (ele é livre); ou então seu ambiente o determina (ele é controlado).

Neisser (1976), então, nos lembra dos estudos da atividade de controle perceptual, onde duas soluções para esse problema são populares entre os psicólogos cognitivos. Uma primeira delas, estaria relacionada a distinção feita entre percepção e atenção, sendo percepção determinada pelos estímulos que nela incidem, enquanto os mecanismos de atenção seletiva permanecem sobre o controle do próprio indivíduo. Conforme Neisser, essa seletividade é inerente a muitos dos processos de coleta de informação e não pode ser relegada a nenhum dispositivo em separado. Uma segunda solução estaria relacionada aos trabalhos de J. S. Bruner (1957, 1973), como apontado por Neisser, que atribui controle ao perceptor, onde ele diz que esse vai cada vez mais longe além da informação dada enquanto adquirir habilidades perceptuais mais sofisticadas. Nesse ponto de vista, o principal impulso do desenvolvimento cognitivo é tornar o adulto mais livre que a criança: ele passa a ser menos “limitado aos estímulos” (*stimulus-bound*) e mais “internamente direcionado” (*inner-directed*).

Nos estudos de movimentos oculares, por exemplo, Mackworth e Brunner sugerem que “a busca visual deve se desenvolver a partir de um estado em que o olhar é controlado pela natureza do estímulo e de suas características intrínsecas a um em que ele é, nas palavras de Yarbus, um instrumento do pensamento” (1970, p.166). Na mesma linha, Eleanor J. Gibson diz que a atenção “muda do estado de ser capturada para ser exploratória”, e continua, citando William James, “o caráter reflexivo e passivo da atenção [da criança], que ... faz a criança parecer menos pertencente a si mesma que a cada objeto que advém a capturar sua atenção, é a primeira coisa que o professor tem que superar” (1969, p. 456).

Fazendo um pequeno parênteses na apresentação da visão de Neisser, gostaríamos de fazer algumas observações conforme apresentadas por Christopher Mole (2013), segundo ele,

o envolvimento da atenção na ação voluntária é muito mais difícil de se estudar de uma forma controlada do que na percepção. Sendo assim, teorias da atenção advindas da psicologia experimental tem, como resultado, centrado quase exclusivamente em instancias perceptuais da atenção. O que contrasta com as teorias que foram desenvolvidas num período inicial da psicologia, no qual os aspectos de atenção envolvendo ação eram muito mais proeminente. William James, por exemplo, sugere nos *Principles of Psychology* que “vontade não é nada mais que atenção” (James, 1890, p. 424) e, ao mesmo tempo, propõem que “atenção, crença, afirmação, e volição motora, são [...] quatro nomes para um processo idêntico, incidental para o conflito de ideias por si só, a sobrevivência de um a despeito da oposição dos outros” (1880, p.582).

Há alguns aspectos em que as questões que envolvem o relacionamento de atenção para a voluntariedade da ação são paralelas as questões que envolvem o relacionamento de atenção para a consciência da percepção. Parece natural pensar que a atenção é necessária para comportamentos finamente ajustados, tanto quanto parece afigurar necessária para a percepção de detalhes finos. Mas não parece natural pensar que temos de estar prestando atenção à execução de cada ato que realizamos voluntariamente. Um caso poderia ser hipotetizado, como exemplo, para alegar que a atenção figura na produção de alguns, mas não todos, comportamentos voluntários. Pode-se imaginar essa visão sendo desafiada por um teórico que afirma que a atenção é necessária para que uma ação seja voluntária, argumentando que há uma ilusão (análoga à ilusão da luz do refrigerador) que nos dá a impressão errada de que nossos atos desatentos são voluntários também. Segundo Mole (2013), no caso da atenção e consciência há um quebra-cabeça epistemológico sobre o tipo de prova que possa resolver esta questão.

Retomando Neisser (1976), a noção de que adultos são mais livres e menos controlados por seu ambiente do que as crianças, sendo melhor habilitados a irem além da informação dada, aponta para algo importante sobre maturidade e liberdade. Crianças, geralmente, são vistas como mais espontâneas, imprevisíveis e livres das amarras das convenções. Como explicar essa aparente contradição, como elas podem ser “limitadas pelos estímulos”? Além disso, como apontado por Neisser, alguns estudos sugerem que os movimentos dos olhos de adultos seguem um padrão mais convencional que o das crianças e esse aparente paradoxo parece se estender a todas as áreas da percepção e comportamento (p.179).

Para entendê-lo, Neisser (1976) apresenta como exemplo o jogo de xadrez. Uma das características de um bom jogador de xadrez é sua capacidade de colher informações relevantes

do tabuleiro. Um mestre nesse jogo, por exemplo, pode reproduzir a posição de todas as pedras do tabuleiro depois de observá-lo por apenas poucos segundos, já um amador não faria isso tão facilmente. Essa habilidade do mestre se deve a sua capacidade de perceber certos aspectos da posição das peças que escapariam a um jogador menos experiente. A experiência no jogo o dotou com um grande número de esquemas para posições de peças recorrentes; e, de acordo com alguns estudos citados por Neisser, o vocabulário de tais esquemas em grandes mestres de xadrez podem ser maior que o vocabulário de palavras da maioria das pessoas comuns.

Essa informação que o mestre de xadrez absorve do tabuleiro determina não só como ele vai mover suas peças mas também como ele move seus olhos ao observar o tabuleiro. Ou seja, ele literalmente vê o tabuleiro de forma diferente – mais adequada e compreensivamente – do que um novato ou alguém que não sabe jogar. Segundo Neisser (1976), essa diferença de percepções não é questão de perceber certo ou errado, mas antes de perceber mais ou menos informação. As informações que especificam o movimento mais adequado estão disponíveis na luz emitida para qualquer um, tanto para um bebê quanto para um grande mestre enxadrista. Porém, apenas o mestre está preparado para colher tais informações. Como lembra Neisser, a luz carrega a informação sobre o *layout* do ambiente apenas em virtude das leis da óptica, já as especificações sobre possíveis movimentos, entretanto, são estabelecidas pelas regras do xadrez, que são essencialmente uma convenção social.

Os sujeitos que percebem não vão além da informação dada, mas a cultura vai além das contingências elementares da natureza para fazer com que informação adicional esteja disponível. As regras do xadrez não controlam a percepção do mestre, elas tornam isso possível dando-lhes algo para perceber. (Neisser, 1976, p.181, tradução nossa)

Tal interpretação sugere que o jogador de xadrez nem é totalmente livre para olhar e mover onde ele escolhe, nem totalmente à mercê do ambiente. O controle dos movimentos oculares e todo comportamento adaptativo só é compreensível como uma interação. Os esquemas possuídos pelo jogador dirigem o movimento de seus olhos que captam informações que por sua vez modificam os esquemas e vão direcionar novos movimentos. Para Neisser (1976), qualquer movimento particular é “causado” por toda a história do ciclo em que ele se encontra; a causa mais imediata é o esquema em si: um estado momentâneo do sistema nervoso do observador. Isso é tão verdadeiro para o olhar reflexo de uma criança em direção a um som repentino como para o olhar pensativo do mestre de xadrez em direção a rainha de seu oponente. Mas, desde que o estado de cada esquema seja parcialmente determinado pela informação do estímulo aceito anteriormente, ambos os movimentos oculares são, pelo menos parcialmente, regidos por (isto é, adequados para) suas respectivas situações.

Segundo Neisser (1976), o mesmo princípio pode ser aplicado ao desenvolvimento de toda percepção especializada. Os ciclos perceptivos variam de acordo com o tipo e alcance da informação que os guia. Assim, perceptores inexperientes ou que não possuem suficiente sofisticação estarão sintonizados a características relativamente mais superficiais de seu ambiente; perceptores mais sofisticados ou habilidosos para características mais sutis. No geral, o adulto é mais orientado para o futuro e mais efetivamente direcionado a metas do que uma criança, mas não é mais independente do mundo que o cerca que esta última. Ambos dependem do mundo que o cercam, porém um pode tirar mais vantagens, ou melhor, esta mais preparado para agir a partir das informações disponíveis.

No mais, escolher para onde olhar não é o mesmo que escolher o que fazer. Assim, de acordo com Neisser (1976), quando selecionamos uma *ação* em vez de outra, o esquema embutido geralmente inclui alguma antecipação de nossa própria situação futura; como um mapa cognitivo, ele contém o ego. Esperamos ficar satisfeitos com nós mesmos, ou sermos recompensado, ou pelo menos ficarmos mais contentes, se escolhermos A; podemos esperar um bem menor se escolhermos B. Escolhas perceptivas, porém, parecem menos importantes do que estas e geralmente são feitas com pouca referência a esse tipo de resultado. Apesar desta diferença, no entanto, para Neisser as escolhas de percepção e de comportamento têm o mesmo status existencial – “nenhuma escolha é livre das informações em que se baseia” (p.182). A informação é selecionada, obviamente, por aquele que a seleciona – o seletor. Nenhuma escolha é determinada diretamente pelo ambiente. Ainda assim, o ambiente fornece a informação que o seletor vai usar.

Desta maneira, os argumentos acima apresentados por Neisser, implicam que predição e controle do comportamento não são primariamente uma matéria psicológica. Por exemplo, o que deveríamos conhecer para predizer como um mestre enxadrista moverá suas peças, ou seus olhos? Como seu movimento é baseado na informação que ele tira do tabuleiro, desta maneira, para prever seus movimentos teríamos de ter acesso as mesmas informações que ele tem. Em outras palavras, para prever os movimentos de um mestre de xadrez, nós mesmos teríamos de ser um mestre em xadrez. Um psicólogo, por maior que seja seu conhecimento sobre psicologia, ao jogar contra um mestre de xadrez irá sempre perder, precisamente porque o mestre poderá prever e controlar nosso comportamento, enquanto nós não poderíamos fazer o mesmo. Para vencer, deveríamos aprimorar nosso conhecimento sobre xadrez, e não sobre psicologia.

Neisser (1976) chama a atenção também para o fato de que o mestre de xadrez não controla o comportamento do adversário utilizando-se de algum artifício psicológico⁵; ele apenas faz um movimento de forma a provocar uma mudança no ambiente (as posições do tabuleiro) e as possibilidades que ele oferece. Para Neisser, é dessa forma que quase sempre o comportamento é controlado: “mudar o mundo é uma forma muito poderosa de mudar o comportamento; mudar o indivíduo, deixando o mundo de lado, é uma proposição duvidosa” (p. 183). Desta maneira, nenhuma mudança pode promover controle, ou resultados previsíveis, a menos que o setor relevante do mundo onde esse controle ou predição irá se dar seja muito bem conhecido.

Apesar de habilidade no jogo de xadrez ser uma perícia bem específica, Neisser (1976) diz que o ponto em questão aqui pode ser aplicado de forma geral. Ou seja, toda percepção e ação se dá em contínua dependência do ambiente, e não podem ser compreendidas sem um entendimento do ambiente em si. Um psicólogo, por exemplo, não pode fazer predições sobre o comportamento de ninguém, ou ter controle sobre ninguém, que tenha mais conhecimento sobre a situação em que este se encontra do que ele tem, ou que consiga ter acesso a mais informações do que ele foi capaz de reconhecer. Como o ambiente humano é imensamente complexo, uma enorme variedade de disciplinas desde as ciências políticas até engenharia de tráfego surgiram com o esforço para compreendê-lo. E, como nos lembra Neisser, poucas dessas disciplinas podem reivindicar qualquer grande sucesso até a data. Neisser clama então por “uma pouca de humildade para a disciplina que se chama de “ciência do comportamento” (p. 184), uma vez que predição e controle do comportamento no mundo real requerem um conhecimento detalhado do mundo em um nível que usualmente não temos, e que em qualquer caso isso sempre cairá também fora do âmbito da especialidade psicológica.

Essa análise sugere não só que controle absoluto estaria fora de questão sem um conhecimento total do ambiente, como também indica condições sobre as quais algum nível de controle é possível. Por exemplo, em ambientes limitados fortemente controlados, tais como instituições penais, hospitais psiquiátricos, campos de concentração, etc., muitas vezes oferecem condições onde se consegue controlar seus detentos razoavelmente bem. – e, mesmo em tais situações, pessoas podem encontrar formas de manter sua individualidade. Também já

⁵ Um jogador de xadrez às vezes pode jogar “psicologicamente”; ou seja, ele pode tomar em consideração o temperamento de seu oponente ou os hábitos que esse tem ao jogar. Porém, esta é uma estratégia muito limitada e de importância apenas secundária ou acessória; não irá ajudar um iniciante a vencer um especialista, e não pode se contar com ela em posições fracas.

ouvimos falar nesses ambientes de métodos de modificação de comportamento ou lavagem cerebral, por exemplo, que podem apresentar relativo sucesso. Esse sucesso no entanto, segundo Neisser, não é devido ao sucesso de alguma teoria psicológica em particular, mas apenas mostra como pessoas podem agir adaptativamente em situações que oferecem poucas alternativas. Mais importante aqui é que há pouca evidencia de que manipulações empreendidas nesses ambientes tiveram consequências previsíveis depois que o sujeito foi liberado ou solto. Encarceramento pode produzir um efeito poderoso e prolongado, obviamente, mas o comportamento que vai resultar disso não pode ser previsto.

Outro exemplo apresentado por Neisser (1976) seria o controle dos pais sobre os filhos. Inicialmente, quando a criança está sempre próxima aos pais, e em um ambiente conhecido por esses, o controle normalmente é mais poderoso; porém, quando a criança começa a tomar contato com novos ambientes, principalmente em ambientes não muito bem conhecido pelos pais, esse controle começa a diminuir. Assim, sugere Neisser, quanto mais uma pessoa aprende se torna menos suscetível de controle. Pessoas com maior conhecimento são mais difíceis de serem manipuladas, pela mesma razão com que jogadores de xadrez mais habilidosos são mais difíceis de serem derrotados. O que Neisser busca exprimir com toda essa argumentação se torna explícito na seguinte passagem:

A verdade nos torna livres. A verdadeira educação não é primordialmente uma técnica para a manipulação de alunos, como alguns sugeriram, mas exatamente o oposto. Isto não é porque a escolaridade faz as pessoas rebeldes, mas porque lhes permite ver mais possibilidades de ação. (p. 185, tradução nossa)

Desta maneira, Neisser parece querer nos mostrar, da mesma forma que inicialmente afirmou que com o desenvolvimento cognitivo nos tornamos “mais livres”, e que por isso o adulto pode ser visto como mais livre que uma criança, já que ele passa a ser menos “limitado aos estímulos” e mais “internamente direcionado”, também o conhecimento nos torna menos suscetíveis de sermos manipulados ou controlados. Uma vez que possuímos uma sofisticação maior em nossas habilidade cognitivas, podemos utilizar de um maior nível de informação, “ler” o ambiente de forma mais apurada e profunda, o que, não somente faz com que seja mais difícil prever nosso comportamento, manipulá-lo e controlá-lo, como também nos proporciona mais possibilidades de ação.

Neisser (1976) acrescenta ainda que seu argumento levou em conta apenas percepção e ação, não mencionando imaginação, abstração ou fala, etc. Uma vez que estas realizações são baseadas no destacamento de processos cognitivos da situação imediata do indivíduo, portanto, eles são necessariamente ainda mais difíceis de prever ou controlar. Um manipulador pode estar

satisfeito em controlar o que uma pessoa faz, mas ignora o que ela sente ou pensa. A longo termo porém sabemos que imagens e ideais podem ter consequências inesperadas e poderosas na ação.

Da mesma forma que conhecimento pode ter esse efeito libertador, desde que o comportamento é baseado em informação, ele também pode ser afetado por “desinformação”. A mentira muitas vezes é usada para manipular e controlar. Onde a única defesa contra ela seria buscar por fontes confiáveis de informação e desenvolver uma visão mais crítica da informação absorvida. Sobre a manipulação da informação, Neisser acrescenta:

A conexão entre a liberdade de escolha e o acesso à informação válida é fundamental; um não pode existir de forma significativa sem o outro. Isso representa um perigo particularmente grave à liberdade na sociedade contemporânea, em que vários tipos de intermediários e a mídia controlam o nosso acesso a fatos importantes. A suposta ameaça à liberdade representada por sinistros psicólogos manipuladores é uma ilusão, mas o perigo de ser sistematicamente mal informado é muito real. (1976, p. 186, tradução nossa)

Por fim, a comunicação humana oferece oportunidades sem paralelo para compreendermos um ao outro e a nós mesmos – ou mesmo temas como psicologia cognitiva – porém, esse conhecimento nunca é completo e muitas vezes estamos simplesmente enganados. Por outro lado, o ciclo de percepção tende a ser autocorretivo, e há sempre mais informação disponível que ainda não foi utilizada. “O resultado de qualquer único encontro entre cognição e realidade é sempre imprevisível, entretanto, a longo prazo, esses encontros devem mover-nos mais próximos da verdade” (Neisser, 1976, p.194).

3.3. Vontade e livre arbítrio segundo dicionários e enciclopédias de psicologia cognitiva

Segundo Houdé (2004), em *Dictionary of cognitive Science*, o conceito de vontade (will) pode ser definido como:

O conceito de vontade geralmente se refere a um fenômeno psicológico que muitas vezes é difícil de distinguir de desejo e intenção; que é concebido como um evento mental (uma volição) dentro de um agente, e é um determinante de suas ações. Neste sentido, ela pode tanto ser vista de um ângulo dualista, como uma causa mental que é separada dos efeitos físicos que a produz, ou a partir de um ângulo materialista, como sendo idêntica a um evento cerebral que provoca uma série de eventos físicos. Mas se a vontade é um evento mental interior (físico ou não), segue-se que a ação começa quando esse evento ocorre, de tal forma que, para fazer A, um agente só precisa ter a vontade de fazer A? Tentar fazer A não é fazer A. Para contornar esta aporia, alguns filósofos, como Elizabeth Anscombe, argumentam que os termos voluntário e intencional, ao qualificar uma ação, não se referem a qualquer evento mental em particular, mas a um conjunto complexo de atitudes psicológicas (em particular, desejos e crenças) e comportamentos. No entanto, esta concepção apresenta um problema: no raciocínio prático e em ações intencionais, de fato parecem ser eventos psicológicos distintos, identificáveis como

intenções ou volições, que causam as ações. (p. 440, tradução nossa)

Como podemos ver, independente do ponto de vista que se toma da vontade, ou dualista ou materialista, ela é posta como iniciadora da ação, ou comportamento. Sendo, portanto, causa deste. Porém, não se entra em detalhes de se ela pode ser considerada como causa primeira da ação, ou causa suficiente da ação, ou se causada por fatores externos, pertencente a uma cadeia causal determinística. Nesse aspecto, entraríamos na questão de se a vontade pode ser exercida de forma livre ou não, ou seja, o problema do livre arbítrio, que levantaria questões como o problema do compatibilismo e incompatibilismo apresentado no primeiro capítulo. Outra questão apresentada nessa definição de vontade, onde esta é tratada como “um conjunto complexo de atitudes psicológicas e comportamentos”, podemos relacionar com uma visão comum utilizada na psicologia cognitiva, como apresentada por Boden (2006) ou Reeve (2006) em seu manual sobre motivação, que trata o problema de se determinar o que causa o comportamento com uma estratégia explicativa baseada em “micro teorias”, buscando dar conta de responder a questão por diversas frentes, e não mais por uma “macro teoria”, como seria a teoria da vontade como apresentada pelos filósofos apresentados, e também por alguns psicólogos como Wundt e William James, na época da nascente psicologia científica. Porém, como vemos, essa estratégia é contra-intuitiva, no sentido que em nossa experiência psicológica imediata conseguimos identificar, como um evento psicológico distinto, intenções e volições, e estas parecem ter força causal sobre nossas ações.

Em outros dicionários e enciclopédias de ciência cognitiva, ou de psicologia cognitiva, não foram encontradas entradas para o termo *vontade* (Eyseck, 1990; Nadel, 2005; Wilson & Keil, 1999). Porém, em Nadel (2005) encontramos uma entrada para vontade livre (*free will*), escrita por David Hodgson, que introduz o termo como se segue:

A expressão “vontade livre” tem sido entendida como a capacidade dos seres humanos de se envolverem em conduta voluntária de acordo com as suas próprias escolhas sem restrições. As questões de como tal capacidade poderia relacionar-se à causalidade como entendida pelas ciências físicas, e se os seres humanos têm essa capacidade, ainda não foram decisivamente respondidas. (p. 2822, tradução nossa)

Prosseguindo, ao oferecer uma análise do conceito em questão, Hodgson (2005) acrescenta que, com o desenvolvimento da ciência, particularmente durante os últimos quatro séculos, houve uma forte tendência em se apoiar a aplicabilidade universal das leis da natureza, o que, aparentemente, deixa pouco espaço para qualquer capacidade de conduta voluntária que em si não esteja inteiramente sujeita e determinada pelas “leis naturais”. Com a ressalva de que a compreensão científica da consciência ainda é muito limitada, e que há razões para pensar

que a consciência possa desempenhar um papel no comportamento humano que não seja totalmente governado por ou explicável apenas em termos das operações de leis da natureza por meio de sistemas físicos. Assim, reconhece Hodgson, ainda há questões sobre vontade livre que estão longe de serem resolvidas.

Um ponto que gostaria de chamar atenção aqui é sobre a aplicabilidade das leis naturais. Leis da natureza nesse sentido, parecem ser restritas a um sentido de leis naturais físicas, que de certa maneira representam uma visão determinista do mundo. Embora essa visão não seja uma inferência imediatamente dada quando falamos em leis da natureza, já que, como vimos, podem existir leis naturais que não sejam determinísticas, e a visão determinista em si é um pressuposto metafísico. Ademais, não possuímos um compêndio fechado de todas as leis naturais existentes. A descoberta de uma lei, que rege um aspecto particular dos fenômenos da natureza (considerando os fenômenos psicológicos como tais) e que tenha como característica a não submissão a determinantes externos, ou físicos, por exemplo, é plenamente possível, ainda que, obviamente, seja algo que deva ser comprovado cientificamente. O que queremos dizer é que o argumento de que vontade livre não é possível pois ele não seria consistente com as “leis naturais”, parece ser muito fraco e defensável apenas por se buscar defender um paradigma materialista/determinista e afastar qualquer indício de algo que coloque essa visão em xeque.

Continuando, conforme o termo vontade livre foi apresentado por Hodgson (2005), a questão de se ou não os seres humanos têm essa capacidade e, em caso afirmativo, como ela funciona e qual é o seu significado, depende em parte do que o próprio conceito é tomado como envolvendo. A maioria concorda que o conceito envolve uma capacidade de agir de forma consciente, racional e sem restrições. A palavra “vontade” refere-se à capacidade de volição, no sentido de iniciação e execução da conduta voluntária; e conduta voluntária é entendida como exigindo consciência e controle da conduta em questão. A palavra “livre” carrega a noção de que o comportamento é o resultado de uma escolha sem restrições, por questões fora do controle da pessoa que age. Controle de conduta e escolha pressupõem alguma racionalidade mínima.

Associado com o conceito de vontade livre está o conceito de responsabilidade: os seres humanos são amplamente considerados como sendo responsáveis por suas ações voluntárias, precisamente porque essas ações são consideradas sob seu controle. Por conseguinte, é amplamente considerado adequado elogiar ou culpar, ou premiar e castigar, uma pessoa por sua conduta voluntária, numa medida que é de certa forma proporcional ao mérito ou demérito da

conduta. Isto reflete-se no direito penal, que inclui regras bastante elaborados para determinar se as pessoas podem ser consideradas como responsáveis por sua conduta, o que seria considerado, em caso positivo, objetivamente uma violação da lei.

Segundo Hodgson (2005), um problema central sobre o conceito de vontade livre é como ele se relaciona com as teorias da causalidade. Um ponto de vista comum de causalidade é que para um evento A ser causado por eventos anteriores B, C, D, etc., esses eventos anteriores devem equivaler a condições suficientes para a ocorrência de evento A. Para vontade livre fazer sentido, parece que a conduta voluntária de uma pessoa teria de ser causada pelo exercício da capacidade da pessoa de se envolver em tal conduta; mas o que então causa o exercício da capacidade? Deste ponto de vista da causalidade, deve ter havido condições prévias suficientes para isso, e outras condições suficientes prévias para essas condições suficientes anteriores, e assim por diante. Se todos os eventos são causados por condições suficientes anteriores, parece que qualquer evento que ocorra, incluindo qualquer ação humana, deve ser determinado, de fato pré-determinado, pelos seus antecedentes causais. Outra maneira de colocar a mesma ideia é dizer que qualquer evento que ocorre é determinado unicamente por condições prévias e as leis da natureza.

Uma controvérsia sobre vontade livre é se o conceito deve ser entendido num sentido básico, segundo o qual a liberdade não está excluída, mesmo se todas as opções e exercícios de controle estão se predeterminado por condições suficientes anteriores fora do controle da pessoa que age; ou se deve ser entendido num sentido forte, como atribuindo aos seres humanos um poder de escolha, cujo exercício não é em si inteiramente predeterminado por condições prévias fora do controle da pessoa que está agindo. Assim, uma outra controvérsia importante é se ou não os seres humanos realmente têm vontade livre, especialmente se este for entendido num sentido forte.

Depois de fazer essa breve apresentação, Hodgson (2005) faz uma rápida incursão histórica sobre o conceito de vontade livre, iniciando por Descartes e passando por Newton, Hume e Kant. Não entraremos nesses detalhes pois já buscamos apresentar a visão desses filósofos no primeiro capítulo. Porém, gostaríamos de salientar alguns pontos interessantes apresentados por Hodgson. Ao falar das leis do movimento de Isaac Newton, indica que elas foram propostas como aplicáveis, sem exceção, a toda e qualquer matéria física, e elas afirmam que, dada a massa, posição e movimento de qualquer pedaço de matéria em um tempo e as forças que agem sobre ela ao longo de um período posterior a tal tempo, suas posições e movimentos durante esse período foram determinadas unicamente por esta lei.

Hodgson acrescenta então que o próprio Newton não explicitou o que isso poderia significar para a matéria física do cérebro humano ou para o dualismo de Descartes, mas no século XVIII, o francês matemático Pierre Laplace apontou que, no âmbito do regime de Newton, dada as condições iniciais de qualquer evento, mais leis da natureza, estas determinariam o futuro, e que poderia, assim, ser exatamente calculado por um ser com dados e informações suficientes; e Laplace não fez nenhuma exceção para o cérebro humano, e não deixou espaço para qualquer eficácia independente da vontade humana.

Durante o século XIX, a evolução científica deu ainda mais suporte à visão de que toda a matéria se comporta em total conformidade com as leis impessoais. Dois eram de particular importância: James Clerk Maxwell mostrou como as forças associadas com a radiação eletromagnética, inclusive a luz visível, pode ser calculada por referência a regras matemáticas, análoga à lei da gravitação de Newton; e a teoria da evolução de Charles Darwin que forneceu uma explicação de como a enorme complexidade e variedade de vida poderia ter surgido a partir de começos simples pela operação das leis impessoais da natureza, e também de como os animais e seres humanos poderiam estar apenas em aparência fazendo escolhas livres, mesmo que, na realidade, estejam funcionando em conformidade com leis impessoais.

No século XX, as teorias da relatividade de Albert Einstein apareceram para apoiar e fazer avançar a ideia de que existem leis universais da natureza que regem todos os sistemas: a sua teoria especial, entre outras coisas, previa uma modificação na mecânica de Newton, de modo a assegurar a coerência integral com a teoria de Maxwell; e sua teoria geral ofereceu uma explicação mais profunda da força gravitacional. Hodgson aponta também como a obra de Sigmund Freud dirigiu a atenção para o fato de que grande parte da motivação humana está inconsciente ou quase inconsciente; e, desde então, tem sido impossível manter uma visão da ação humana como uma questão de decisões totalmente conscientes com base em motivos totalmente conscientes. A invenção da mecânica quântica (MQ) na década de 1920 desafiou o determinismo da física clássica de Newton, Maxwell e Einstein; mas a única indeterminação postulada pela MQ foi a aleatoriedade, de modo que a MQ não suporta diretamente o forte senso de vontade livre. Em outros aspectos, a MQ tende a confirmar a aplicação universal das leis da natureza: ela mostrou como os eventos em nível atômico poderiam ser explicados por leis que também foram responsáveis pela exatidão substancial da física clássica em escalas maiores; e também forneceu a base para a explicação de Linus Pauling da ligação química em termos de física quântica, mostrando como as propriedades químicas da matéria poderiam ser explicadas pela física básica.

Na segunda metade do século, a descoberta da estrutura do DNA de Crick e Watson mostrou como a própria vida pode ser explicada pelas leis da física e da química. E os avanços das ciências cognitivas permitiram que mais e mais aspectos do funcionamento do cérebro humano pudessem ser entendidos em termos da operação das leis da natureza em sistemas físicos. No entanto, salienta novamente Hodgson, ainda há pouca compreensão da consciência em geral, e seu papel na conduta voluntária em particular; para que os problemas relativos ao conceito de vontade livre pudessem ser considerados como resolvidos.

Sobre as contribuições oferecidas pelos avanços das ciências cognitivas para o problema em questão, Hodgson destaca o que agora é conhecido sobre o funcionamento do cérebro como um sistema físico. Ao nível dos neurônios individuais do cérebro, uma grande quantidade é agora conhecida sobre o que faz com que eles disparem ou não disparem; como os sinais elétricos são transmitidos dentro dos neurônios e, em seguida, passam através das sinapses com outros neurônios por meio de transmissores químicos; e como tudo isto é afetado pela química do cérebro. Muito também se sabe sobre os padrões de conexão entre os neurônios do cérebro, e quais regiões do cérebro estão envolvidas em determinadas funções cognitivas e comportamentos.

O cérebro é considerado como um sistema de processamento de informação, que leva a informação a partir de estímulos sensoriais, processando-a de várias maneiras, e dá vários tipos de saídas – em geral, resultando em ações físicas, como andar ou falar. Hodgson lembra-nos ainda que os processos cerebrais são por vezes considerados análogos aos de uma máquina de computação – como vimos várias vezes no segundo capítulo – e reforça que esta ideia tem a sua base na tese de Church-Turing, “no sentido de que qualquer coisa que pode ser computada ou calculada pode ser computado, com o tempo devido, por uma máquina de propósito geral” (Hodgson, 2005, p. 5). Houve muito trabalho teórico e prático na ideia de que o desempenho humano cognitivo (sensação, percepção, raciocínio e execução das decisões, etc.) pode ser reproduzido ou simulado por computadores.

Acrescenta ainda que, esforços estão sendo feitos para identificar os processos cerebrais envolvidos na ação voluntária, com algum sucesso, embora sem muita compreensão do papel da consciência nesses processos. Três outras áreas de investigação são de particular relevância para se considerar a vontade livre, e que sugere que nossos processos conscientes podem ser considerados a ponta de um iceberg cognitivo, tendo pouco ou nenhum papel na determinação do que dizemos e fazemos: (1) as investigações sobre as habilidades cognitivas prodigiosas de processos inconscientes – por exemplo, investigações sobre o processamento de informação

pré-consciente necessária para a percepção e para o uso da linguagem; no fenômeno chamado “*sleep-on-it*”, onde a resposta a um problema de repente ocorre a uma pessoa algum tempo depois de deixar de pensar conscientemente sobre isso; e para o cálculo e outras capacidades de alguns *savants*, sugerindo que todos nós podemos ter capacidades semelhantes inacessíveis à consciência; (2) investigação da natureza e extensão da motivação inconsciente, não apenas confirmando que grande parte de nossa motivação é inconsciente, mas também sugerindo que somos adeptos a racionalizar nossa conduta para sustentar sua fabricação (inconscientemente), e, em seguida, acreditamos nessa racionalização como histórias plausíveis para justificar por que fizemos o que fizemos (Gazzaniga, 1988; Wegner, 2002); e (3) as investigações que sugerem que a consciência vem demasiado tarde para o controle em tempo real de ações (os experimentos de Libet et al, 1979, 1983).

No entanto, estes resultados não excluem controle consciente na formação da conduta voluntária: em um concerto por um pianista, a consciência vem demasiado tarde para dirigir os dedos para as teclas certas, mas não necessariamente muito tarde para moldar o desempenho em resposta a sons ouvidos e o sentimento de emoções. Além disso, como o próprio Libet apontou, os resultados não excluem um veto consciente exercível até o momento de iniciar uma ação. E eles não dizem nada sobre decisões consideradas.

Sobre a relevância da questão da vontade livre nas ciências cognitivas, Hodgson destaca o fato de que o cérebro humano parece estar adaptado, com algum custo em termos de complexidade e uso de energia, para apoiar os processos conscientes, sugerindo que os processos conscientes devem contribuir significativamente para a sobrevivência e reprodução. Seria plausível que tal contribuição seja feita como meio de determinar que a conduta voluntária seja realizada – ou seja, de que a vontade livre seja exercida. As ciências cognitivas ainda têm de explicar como tal contribuição poderia ser feita, em termos da operação das leis da natureza em sistemas físicos. A consciência, ainda segundo Hodgson, envolve dois truques que parecem que não podem ser realizados por computadores e que não têm nenhuma função conhecida no processamento da informação como atualmente entendida: o truque *qualia* e o truque *chunking*. O primeiro é o truque de associar tipos de processos físicos no cérebro com os tipos de experiência qualitativa tais como ver azul ou sentir dor; e o último é o truque de juntar casos particulares desses tipos gerais de experiência em experiências particulares inteiras de uma pessoa. Este segundo truque é associado com o que é chamado o “problema de ligação” da consciência, sobre como a informação é transportada em partes distintas do cérebro (como “vendo uma cor vermelha” e “vendo um círculo”) e trazido juntos em uma experiência

consciente unitária (tais como “vendo um círculo vermelho”). As ciências cognitivas têm tido algum sucesso na identificação de processos cerebrais envolvidos em ambos esses truques, mas praticamente nenhum em explicar seu papel funcional. A compreensão destes truques neste momento está muito aquém de permitir a criação de um sistema artificial que possa realizar tais feitos, ou de afirmar o que distingue um tal sistema de um que tenha o poder de processamento de informação semelhante mas que não possa realizá-los.

Nesse ponto, Hodgson apresenta seu ponto de vista pessoal sobre o assunto:

Como alguém que prefere a visão libertária da vontade, eu acredito que os dois truques podem, eventualmente, ser entendidos como tendo uma função em combinação com um terceiro truque que faz com que eles sejam possíveis, o que eu chamo o truque de seleção, envolvendo a capacidade de fazer seleções entre alternativas levantadas pelo processamento de informações inconscientes. Neste ponto de vista, o processamento da informação inconsciente dá origem a alternativas disponíveis para seleção, fornecendo razões com base em como uma seleção entre essas alternativas pode ser feita, e determina a intensidade com que essas razões são sentidas; e também dá origem a experiências inteiras que têm qualidades únicas com as quais leis gerais da natureza não podem se envolver, mas ao qual a pessoa que tem as experiências pode responder de forma racional e fazer a seleção. Esta é a visão de uma vontade livre que eu acredito que poderia ser tratada utilmente pelas ciências cognitivas, particularmente como eles continuam a investigar os problemas não resolvidos associados com a ação voluntária, raciocínio plausível, o truque qualia, e o truque chunking. (2005, p.6, tradução nossa)

Como vimos nessa citação de Hodgson, onde ele expõe sua visão pessoal e é bem direto ao se posicionar quanto a possibilidade de uma vontade livre. Hodgson é um juiz na corte australiana, mas também tem vários trabalhos de cunho filosófico publicados. Defende uma visão libertária de uma vontade livre, assim como uma divisão entre tipos de leis naturais, dentre um desses tipos propostos de funcionamento da lei natural, que segundo ele faria sentido defender uma vontade livre de forma que não seria contraditório com as leis naturais. Apresentaremos, portanto, de forma breve, um artigo de Hodgson, onde ele resume todo seu pensamento esclarecendo melhor seu posicionamento.

No artigo intitulado “*Why I (Still) Believe in Free Will and Responsibility*”, Hodgson (2007) apresenta sua argumentação em dois pontos principais a favor da existência de uma vontade livre relacionados: ao raciocínio plausível e às experiências conscientes; a como a ciência pode acomodar uma vontade livre. Ao discorrer sobre o raciocínio plausível, Hodgson afirma que os seres humanos são hábeis a decidir razoavelmente sobre o que fazer e o que acreditar. Segundo ele, apesar de todo mundo cometer erros e algumas pessoas acreditarem em coisas tolas, geralmente razões são encontradas para isso. Dessa forma, qualquer investigação intelectual deve assumir que as pessoas envolvidas têm a capacidade de tomar decisões razoáveis sobre o que acreditam.

Prossegue sua argumentação expondo a importância da capacidade de se envolver em raciocínio plausível, no qual premissas ou dados não impliquem em conclusões, mas sim apoiem um julgamento razoável. A maior parte do raciocínio humano não é algorítmico, ou seja, não é um processo regido por leis da lógica, matemática e/ou probabilidade, ou qualquer outra regra que possa ser incorporada dentro de um programa de computador. Hodgson (2007) exemplifica tal assertiva ao falar sobre o método científico, o qual envolve formulação e teste das hipóteses, porém depende muito de um raciocínio plausível, seja na formulação das hipóteses a serem testadas, seja na elaboração dos experimentos para testá-las, ou na seleção de quais hipóteses incontestáveis podem ser provisoriamente aceitas (por limitação do experimento). Assim, se o raciocínio plausível não pudesse apoiar decisões razoáveis sobre o que acreditar, ciência e filosofia não seriam possíveis.

Hodgson (2007) afirma que há uma lacuna entre razões e decisões/ações, neste ponto cita Hume e suas considerações sobre a ação estar de acordo com a preponderância de desejos. Ao correlacionar os desejos e o raciocínio plausível como fatores de influência na tomada de uma decisão, Hodgson afirma que, como esses dois são forças incomensuráveis, não é possível prever uma resultante que definiria a tomada de decisão final.

Correlaciona também o raciocínio plausível com as experiências conscientes pessoais (incluindo experiências visuais e auditivas, pensamentos e sentimentos), sendo que as experiências contribuiriam para a formação do raciocínio plausível. Reconhece que o inconsciente realiza processos algorítmicos admiráveis, e que esses são importantes diante de uma situação em que se necessita de uma tomada de decisão e de ação, por sumariarem informações que irão auxiliar na decisão a ser feita, porém reafirma que apesar de boas ideias virem como resultado de processos inconscientes, não são adotadas sem o aval consciente. Além disso, Hodgson ressalta, ao citar outros autores como Daniel Kahneman e Amos Tversky⁶, que o raciocínio plausível é afetado por vieses inconscientes, porém apenas uma apreciação consciente cuidadosa desse raciocínio minimiza esses vieses.

Ao citar a tese de Alan Turing sobre a computação e o funcionamento de computadores, a qual baseia que qualquer conclusão que pode ser alcançada pela operação de regras gerais em circunstâncias existentes pode ser alcançada sem consciência, Hodgson (2007) questiona sobre

⁶ Kahneman, D., Slovic, P., and Tversky, A. (eds) (1982). *Judgement under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

ausência de regra plausível que determinasse a experiência consciente, se o processo cerebral fosse precisamente determinado por regras.

Ainda sobre as experiências conscientes, Hodgson reconhece que estas são processos físicos cerebrais e que, portanto, as informações nelas contidas podem ser codificadas em processos físicos, porém, como essas informações são experimentadas conscientemente, são caracteristicamente combinadas em conjunto ou *gestalts* unificadas. Prossegue arguindo que apesar dessas *gestalts* não poderem, como um todo, se envolver com leis ou regras de qualquer tipo, elas podem, plausivelmente, fazer uma contribuição positiva para tomada de decisão.

O segundo pilar de argumentação de Hodgson consiste em estabelecer e apoiar as assertivas que indicam que a ciência tende mais a aceitar uma vontade livre do que o contrário. Discorre, inicialmente, sobre a mecânica quântica, onde relata um famoso artigo⁷ publicado em 1935 por Einstein e dois colaboradores, no qual argumentavam que a mecânica quântica deve ser incompleta, com base em uma suposição, apoiado pela teoria da relatividade, que uma medição feita em uma partícula não pode afetar outra partícula que estava distante dela. Porém essa suposição foi refutada por um teorema formulado por John Bell e experiências realizadas no início de 1980 por Alain Aspect⁸, que demonstraram que onde partículas de matéria interagem e se correlacionam por essa interação, uma medição feita em uma delas pode afetar a outra, mesmo que estejam afastadas uma da outra no espaço e não possam se comunicar à velocidade da luz, ou menos.

Apesar de alguns teóricos tentarem formular uma versão determinista da teoria quântica, um teorema recentemente idealizado por John Conway e Simon Kochen⁹ mostra que os resultados de certas medidas realizadas sobre as partículas correlacionadas não podem ser fixados antecipadamente, a menos que de alguma forma a natureza impeça pesquisadores de fazerem medidas particulares que revelem contradições. Hodgson considera essa conspiração da natureza altamente improvável, e acredita que prejudicaria o método científico, negando o acesso dos pesquisadores a amostras aleatórias. Dessa forma, qualquer sugestão de que o método científico suporta o determinismo rigoroso seria refutada por si mesma.

⁷ Einstein et al. (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47, 777 – 780.

⁸ Aspect, A; Dalibard, J.; & Roger, G. (1982). Experimental test of Bell's inequalities using time-varying analyzers. *Physical Review Letters*, 49, 1804-7.

⁹ Conway, J. e Kochen, S. (2006) The free will theorem. *Foundations of Physics*, 36, 1441.

Hodgson (2007) finaliza a argumentação afirmando que a neurociência atual não sugere que as experiências conscientes unificadas podem ter efeitos nos processos cerebrais, além de tender a sugerir que indeterminismo associado à física quântica ocorre em muito pequena escala para acomodar tais efeitos. Porém, Hodgson sugere que é especialmente porque a neurociência falha em explicar ou acomodar experiências conscientes, e a tendência compreensível de neurocientistas para minimizar o significado do que eles não entendem, como a consciência, sendo isso indicativo de limitações da neurociência atual ao invés de ineficiência das experiências conscientes na tomada de decisão.

3.4. Motivação

Como vimos, a área de motivação dentro da psicologia cognitiva parece estar relacionada com o problema da vontade tal como investigado aqui, ou seja, como uma área de interesse onde se busca examinar as causas do comportamento e, principalmente, se esse é determinado ou livre. O próprio Miller et al. (Miller, Galanter & Pribram, 1960), no *Plans and the structure of Behavior*, afirma que “a vontade parece ter desaparecido da teoria psicológica, assimilada anonimamente no tópico mais amplo da motivação” (p.11). Reeve (2006), em seu manual *Understanding Motivation and Emotion*, da mesma forma liga diretamente a investigação filosófica sobre vontade e os estudos modernos da psicologia cognitiva sobre motivação (p.26). Porém, assim como Miller et al. ao afirmar que ela foi “assimilada anonimamente”, diz que a forma de se abordar essas questões no estudo da motivação seria por micro-teorias que buscam dar conta de responder diversos aspectos relacionados a fenômenos específicos do comportamento motivado, suas possíveis causas, determinações, intensidade, o que faz com que alguns comportamentos perdurem e outros não, etc.

Por ser um campo muito vasto, e que abarca uma gama muito ampla de fenômenos, nos mais diversos contextos, como educacional, organizacional, do esporte, terapêutico, entre outros; além de existirem diversas teorias próprias, às vezes muito limitadas a alguns desses contextos notamos uma grande dificuldade em abranger todo o campo. Além de existirem diversos autores que desenvolvem abordagens diferentes, muitas vezes com um vocabulário conceitual próprio. Não só algumas dessas abordagens competem entre si, como algumas partem de visões conflitantes sobre a natureza humana, e mesmo sobre o que seria motivação. Por questões de tempo para o desenvolvimento do presente trabalho, infelizmente só nos será

possível apresentar uma visão mais ampla, e muita das vezes simplificada dos autores, buscando apenas investigar pontos de interesse para os nossos objetivos propostos.

Inicialmente, portanto, buscamos então compreender o que os teóricos que estudam o campo querem dizer com o termo motivação. Notamos que é um termo amplamente utilizado e que é empregado também em diferentes contextos, relacionado, por exemplo, a psicoterapia, a psicomетria com os testes motivacionais ou a teoria da aprendizagem, podendo até mesmo apresentar diferentes significados não só quando aplicado em diferentes contextos mas as vezes dentro de um mesmo contexto. Apresentamos, portanto, a seguir, algumas definições de motivação:

Um motivo é uma necessidade ou desejo acoplado com a intenção de atingir um objetivo apropriado (Krench & Crutchfield, 1959, p. 272, tradução nossa).

Uma busca dos determinantes (todos os determinantes) da atividade humana e animal. (Young, 1961, p. 24, tradução nossa).

O energizador do comportamento. (Lewis, 1963, p. 560, tradução nossa).

Pode-se falar em uma teoria da motivação e significar uma concepção coerente dos determinantes contemporâneos da direção, do vigor e da persistência da ação. (Atkinson, 1964, p. 274, tradução nossa).

A psicologia tende a limitar a palavra motivação... aos fatores envolvidos em processos de energia, e a incluir outros fatores na determinação do comportamento. (Cofer, 1972, p. 2, tradução nossa).

Para cada ação que uma pessoa ou animal executa, nós perguntamos: 'Por que ele ou ela fez aquilo'. Quando fazemos esta pergunta, estamos perguntando sobre a motivação daquela pessoa ou animal... Questões sobre motivação, então, são questões sobre as causas de uma ação específica. (Mook, 1987, p. 3, tradução nossa).

Em abordagem operacional, (motivação) é o conjunto de relações entre as operações de estimulação ou privação e as modificações observadas no comportamento que se processa após as citadas operações. (Penna, 2001, p. 19).

A motivação tem sido entendida ora como um fator psicológico, ou conjunto de fatores, ora como um processo. Existe um consenso generalizado entre os autores quanto à dinâmica desses fatores psicológicos ou do processo, em qualquer atividade humana. Eles levam a uma escolha, instigam, fazem iniciar um comportamento direcionado a um objetivo. (Bzuneck, 2004, p. 9).

Como podemos ver pelos trechos supracitados, o conceito de motivação é abordado de formas bem diversas e, algumas vezes, até conflitantes entre si. A abrangência também em algumas formas com que é tratado parece fazê-lo se confundir com os objetivos da própria psicologia como um todo. Como, por exemplo, ‘por que as pessoas se comportam desta ou daquela maneira?’. Motivação também pode ser entendida como aquilo que move o comportamento, sendo algumas vezes vista como uma força (mas não determinada de qual espécie), ou como um “energizador” do comportamento.

No mais, psicólogos da motivação de orientação cognitiva estão mais interessados na sequência cognição-ação e, como vimos, Miller et al. (1960) reconhecem que existe uma lacuna explicativa nesse ponto que remetem ao estudo da vontade como vista na filosofia. Alguns dos constructos cognitivos implicados na motivação e na sequência cognição-ação são os que se seguem: os planos, as metas, a dissonância, os esquemas, as expectativas, as avaliações, as atribuições (Reeve, 2006).

De acordo com Reeve (2006), acontecimentos do meio incidem nos sentidos humanos construindo vivências sensoriais que são processadas pelo sistema nervoso central. A informação é recebida, transformada, organizada, elaborada e tida em conta enquanto recupera-se da memória a informação que facilita o fluxo do processamento de informação. Segundo o significado que damos a este processamento da informação nós criamos expectativas, fazemos planos, etc. Nossas expectativas, planos e metas são agentes ativos que dão lugar aos fenômenos motivacionais porque dirigem nossa atenção e nosso comportamento para uma sequência de ação designada. Depois que agimos, produzem-se as consequências dessa ação, que por sua vez são avaliadas. Esta atividade cognitiva provocada pelas consequências da ação une-se ao fluxo do processamento da informação no que constituiria a regulação cognitiva da conduta. Ao fazerem uso de seu conhecimento, as pessoas constroem planos e metas para aumentar a probabilidade de que terão consequências positivas e diminuir a probabilidade de consequência negativas. Uma vez que as pessoas têm uma certa expectativa da consequência de um determinado acontecimento, sua atuação se coadunará a essas expectativas, ou seja, seu comportamento irá se ajustar a suas cognições. Portanto, para a psicologia cognitiva, a cognição leva ao comportamento. E, desde um foco motivacional, a cognição é um “trampolim à ação”.

Um teórico importante, pioneiro na abordagem cognitiva da motivação, foi Edward Tolman (1948). Em sua época as principais teorias da aprendizagem se baseavam na abordagem estímulo-resposta. Porém, Tolman propunha que a cognição seria uma variável interveniente entre o estímulo e a resposta. Buscou ainda mostrar como se adquirem as expectativas e como uma vez adquiridas incidem na motivação e no comportamento, que seria dirigido a metas. Baseando-se nessa tese, Tolman concluiu que o comportamento era principalmente *intencionado*. Também fez conjecturas de que o comportamento é uma manifestação da cognição ou do conhecimento, refletindo os processos cognitivos enquanto a pessoa está utilizando hipóteses, expectativas e estratégias de forma constante com *intenção* de atingir metas e evitar obstáculos.

Tolman costumava utilizar dois termos intercambiáveis com propósito e cognição: exigência e expectativa. A exigência e o propósito eram intercambiáveis no sentido em que o propósito de conseguir uma meta ambiental equivalia ao desejo de satisfazer a exigência corporal. Quando surgem exigências corporais, os estímulos que são atendidos de forma seletiva se convertem em metas ambientais. Os termos cognição e expectativa são intercambiáveis no sentido de que a cognição era a base de uma expectativa segundo a qual uma meta ambiental particular podia satisfazer uma exigência corporal. As expectativas seriam então unidades de conhecimento armazenadas na memória que seriam ativadas por acontecimentos biologicamente significativos, e dar-se-iam pela repetição de sequências ambientais.

Outro teórico pioneiro na abordagem cognitiva da motivação, Kurt Lewin (1935), tem uma abordagem da motivação com semelhanças a de Tolman, uma vez que vê o indivíduo como um organismo em busca de metas. Para melhor explicar a motivação do comportamento, propôs sua teoria de campo, que se baseia em duas suposições fundamentais: (i) o comportamento humano é derivado da totalidade de fatos existentes e; (ii) esses fatos tem um caráter de um campo dinâmico, no qual cada parte do campo depende de uma inter-relação com as demais outras partes. Esse campo dinâmico é o espaço de vida que contém a pessoa e seu ambiente psicológico, sendo o ambiente psicológico o ambiente tal como é percebido ou interpretado pela pessoa. Mais do que isso, o ambiente está relacionado com as necessidades atuais do indivíduo. A partir disso, propõem uma equação para explicar o comportamento humano: $C=f(P,M)$. Onde o comportamento (C) é a função (f) ou resultado da interação entre a pessoa (P) e o meio ambiente (M) que a rodeia.

Segundo Lewin (1935), as necessidades tanto fisiológicas como não fisiológicas criam intenções, as intenções por sua vez produzem tensões, e as tensões contribuem para se formar metas que provocavam no organismo a conduta, ou seja, a ação consumatória dirigida para uma meta. Um constructo cognitivo importante era a “valência” que se refere ao grau de valor positivo ou negativo dos objetos do meio da pessoa. Os objetos atraentes ou os objetos que satisfazem uma necessidade têm um grau de valência positivo enquanto os objetos que se interpõem à satisfação de necessidades têm uma valência negativa. Assim, a pessoa busca no meio os objetos-meta capazes de satisfazer a intenção e reduzir a tensão. Os objetos, pessoas ou situações de valência positiva atraem o indivíduo e os de valência negativa o repelem. A atração é a força ou vetor dirigido para o objeto, pessoa ou situação; a repulsa é a força ou vetor que o leva a se afastar do objeto, pessoa ou situação, tentando escapar. Um vetor tende sempre

a produzir locomoção em uma certa direção. Quando dois ou mais vetores atuam sobre uma pessoa ao mesmo tempo, a locomoção é uma espécie de resultante de forças. Lewin utilizou uma combinação de análise topológica (para mapear o espaço vital) e vetorial (para indicar a força dos motivos no comportamento), desenvolvendo uma série de experimentos sobre a motivação, a satisfação e a frustração e os efeitos da liderança autocrática e democrática em grupos de trabalho. Suas ideias são importantes porque grande parte da terminologia introduzida por ele serviu de base para as gerações seguintes de teorias cognitivas da motivação.

De acordo com Reeve (2006), outro grande influenciador sobre as teorias cognitivas motivacionais foram as formulações teóricas de Miller et al. (1960) em *Plans and the Structure of Behavior*. De acordo com Miller et al., as pessoas têm representações cognitivas ideais de suas condutas, dos objetos do meio e dos acontecimentos. Quando as pessoas se encontram ante um desajuste ou uma incongruência entre seu estado ideal e seu estado atual as pessoas adotam um plano de ação de forma que possa promover um ajuste, ou uma maior congruência entre seus estados ideal e atual. Essa incongruência, portanto, é de natureza motivacional e o plano representa um esforço por resolver os problemas para assim reestabelecer a congruência cognitiva – ou homeostase. O mecanismo cognitivo mediante o qual os planos “energizam” e dirigem a conduta ficou conhecido como a unidade TOTE (*test-operate-test-exit*), já mencionada anteriormente. Na unidade TOTE a fase de teste corresponderia à comparação entre o estado atual e o estado ideal. Se o estado atual e o estado ideal não se correspondem, então se dá a incongruência e o indivíduo põe em marcha uma sequência de ação. Ao atuar sobre o meio, a pessoa se esforça por atingir esse estado ideal. Depois dessa operação, a pessoa volta a comparar seu estado atual com seu estado ideal. Se o segundo teste dá resultado de incongruência então a pessoa ajusta sua conduta e continua atuando sobre o meio. Se o segundo teste dá resultado de congruência entre o atual e o ideal então a pessoa sai do plano.

Para Reeve (2006), uma forma mais atual de ver esse processo é conhecido como motivação corretiva, onde os planos não são vistos como algo fixo e inalterável mas sim como sujeitos a frequentes modificações. As pessoas são vistas como sujeitos ativos, tomadores de decisões e que exercem seu direito de eleger que caminho seguir. Deste ponto de vista, qualquer incongruência entre os estados atual e ideal não provoca uma sequência de condutas quase automáticas, mas sim leva a uma “motivação corretiva” mais geral. A motivação corretiva ativa um processo de tomada de decisões no qual o indivíduo tem em conta múltiplas formas de reduzir a incongruência entre os estados ideal e atual.

Quanto às metas, o estado ideal é uma previsão de como as coisas podem estar melhores, enquanto que com os planos o estado ideal costuma ser uma condição que já está presente no meio. As metas, porém, nem sempre melhoram o rendimento. Para que se dê uma melhoria no rendimento, as metas devem ser específicas e desafiantes. Metas específicas e desafiantes melhoram o rendimento porque centram a atenção da pessoa, mobilizam o esforço, aumentam a persistência e motivam a pessoa a desenvolver novas estratégias. Ademais, outra variável importante é o *feedback*. A fixação de metas só funciona se existe um *feedback* progressivo em relação à meta. Mediante o *feedback* a pessoa pode avaliar se seu rendimento foi inferior, equivalente ou superior ao estabelecido para satisfazer a meta.

Uma das limitações da união entre a fixação de metas e a ação dirigida para uma meta é a decisão da pessoa sobre se vai aceitar ou recusar a meta. Antes que uma pessoa aceite ou recuse uma meta tem que avaliar seu nível de dificuldade. Quando se aceita uma meta fixada externamente a pessoa internaliza a meta e a converte assim em uma meta fixada internamente. Por outro lado, as metas que são recusadas mas que se impõem externamente não melhoram o rendimento. A aceitação de uma meta, que implica a decisão de se auto impor uma meta, provoca um compromisso de esforço enquanto a rejeição de uma meta provoca indiferença ou, no máximo, pouco compromisso.

Outra teoria importante que tem grande influência na abordagem cognitiva da motivação é a teoria da dissonância cognitiva, inicialmente desenvolvida por Leon Festinger (1957). Segundo esta teoria as pessoas têm uma tendência a manter a máxima consistência entre suas crenças e isto tem um efeito sobre sua conduta. Cada vez que uma crença se contradiz com outra provoca uma inconsistência que por sua vez produz uma sensação de mal-estar. Frente a esse mal-estar, a pessoa se esforça para eliminar ou ao menos reduzir a base desta inconsistência. De acordo com esta teoria os seres humanos dispõem de uma série de recursos para mudar estes estados tão incômodos. Ante um estado de dissonância, a teoria da dissonância cognitiva prediz que o indivíduo tomará um desses quatro caminhos para resolvê-la: 1) mudar suas crenças originais; 2) mudar sua conduta; 3) mudar algum aspecto de seu meio; ou 4) acrescentar novos elementos cognitivos. A ideia mais importante aqui é que a dissonância cognitiva produz um motivo para que a consistência cognitiva seja restabelecida e que este motivo se manifesta de acordo com uma dessas quatro formas apresentadas.

Visto isso, como argumenta Barberá (2002), muitos modelos cognitivos desenvolvidos para explicar a motivação assumiram ou reproduziram um desenho característico que ficou conhecido como concepção *homeostática*. Um exemplo mais claro desse modelo pode ser

observado numa série de teorias denominadas, a grosso modo, de “consistência cognitiva”. Embora essas teorias possuam formulações específicas e âmbito de aplicação diferenciados todas convergem na ideia de apresentar a motivação dentro de uma visão onde consistência/inconsistência seriam características centrais. Entre elas, segundo Barberá, podemos apontar, não só a teoria da dissonância de Festinger (1957), mas igualmente o modelo de reatância psicológica (Brehm, 1966), o modelo de auto-percepção (Bem, 1972) ou a teoria da equidade (Adamns, 1965); todas estas compartilham a estrutura prototípica do modelo homeostático.

O esboço geral da teoria de redução do impulso, que na concepção de Hull aplica-se a necessidade de satisfazer o desejo por comida em uma situação de fome, é usado para explicar o funcionamento da psique ante pensamentos dissonantes, percepções de desigualdade ou qualquer outra inconsistência entre crenças e comportamento. O conteúdo motivacional que se pretende conhecer, entretanto, se modifica (de motivos básicos para motivos secundários), assim como o desenho geral da pesquisa (de animais de laboratório para observação e registro de reações humanas). Porém, de acordo com Barberá (2002), a estrutura explicativa homeostática permanece intacta no modelo de redução do impulso e nas teorias de consistência cognitiva. Ambos partilham uma interpretação mecanicista da motivação humana baseada na alternância dinâmica entre equilíbrio inicial (consonância) e desequilíbrio posterior (situação dissonante). O sujeito se motiva, por conseguinte, diante uma situação de estímulo concreta e reage mediante comportamentos predeterminados, a fim de reconstruir o estado anterior, ou estado ideal.

Entre as explicações cognitivas da motivação humana, uma das que tiveram maior impacto, ainda segundo Barberá (2002), foram as que destacaram o papel que as “expectativas” e “valências”, como conceitos antecipatórios da ação, exercem sobre o nível de esforço (tendência motivacional) atribuído a determinados comportamentos. Uma série de modelos desenvolvidos por Atkinson (1957) ou Vroom (1964), por exemplo, e que foram aplicados a áreas diversas, podem ser agrupados em torno do nome comum de teorias de expectativa/valência (E/V). De acordo com Barberá, essas teorias compartilham com as teorias de consonância/dissonância uma visão racional do ser humano e consideram que o comportamento é intencional, na medida em que segue uma finalidade funcional. Da mesma forma, compartilham uma abordagem reagente da motivação, já que as tendências motivacionais são interpretadas como reações específicas a mudanças provocadas pelos estímulos. No caso de expectativas e valências, os estímulos são internos, cognitivos e

antecipatórios. A intencionalidade, ainda que assumida como característica inerente da conduta ou comportamento, não representa um constructo explicativo da motivação humana. Mantendo-se, assim, um critério mecanicista ao atribuir às expectativas e valências plena responsabilidade da tendência motivacional dominante (Barberá, 2002).

Alguns exemplos de reelaboração da abordagem de E/V no estudo da motivação, segundo Barberá (2002), podem ser encontrados nas propostas de Bandura (1977) ou de Heckhausen (1977) para diferenciar tipos específicos de expectativas. De forma bem simplificada, a classificação de Bandura distingue entre expectativas de eficácia e resultado. A primeira, referindo-se à percepção de auto-capacidade de executar um comportamento, e se situa conceitualmente entre o sujeito e a ação. As expectativas de resultado, no entanto, fazem alusão a convicção de que uma determinada ação produzirá um certo resultado. Interveem como atividades mediadoras entre a ação e a resultado esperado.

Heckhausen (1977), por sua vez, classifica três tipos de expectativas: i) de situação-resultado (S-R); ii) ação-resultado (A-R) e os resultados-consequências (R-C). A expectativa de A-R praticamente coincide com a noção de expectativa de resultado de Bandura (1977). A expectativa de S-R é definida como a crença de que uma situação em si mesma, independentemente do comportamento do sujeito, conduzirá um determinado resultado. Por exemplo, a crença de que certas situações de pobreza estão emparelhadas com um fracasso escolar. Finalmente, a expectativa de R-C é a crença de que o resultado esperado agirá como um instrumento de mediação para atingir certas consequências. Se se termina a faculdade (resultado), se poderá ascender a uma posição profissional (consequência).

Para Barberá (2002), tanto a noção de expectativa de eficácia de Bandura como a de expectativas de R-C de Heckhausen, geraram impacto significativo sobre a interpretação psicológica da motivação. Bandura desenvolveu o conceito de auto-eficácia (Bandura, 1982), enfocando o papel motivador que tem a percepção de si mesmo como um agente capaz de levar a cabo certas ações. Sobre a noção de auto-eficácia e a percepção de auto-capacidade, retomaremos ao falar de motivação intrínseca. Por sua vez, a distinção entre resultado e consequências, como expõem Heckhausen (1977), possibilita a interpretação dos resultados como passos instrumentais encaminhados em direção a metas superiores que agregam valor significativo para nossas ações. Assim, embora não possamos ter uma influência direta sobre as consequências mas apenas sobre os resultados, a crença na relação entre os resultados e as consequências mais tarde adquire valor motivacional. A relevância das ações futuras sobre o

comportamento presente será enfatizada por Raynor (1981) por meio de sua concepção de “orientação futura” e também através do conceito de “distância temporal da meta”.

Conforme relata Barberá (2002), além de ampliar o conceito tradicional de expectativa ou incorporar tipos diferenciados de valorações antecipadas (comparação social, autoestima, incentivos extrínsecos, custo/benefício, ou valor cultural, por exemplo), há diversas outras propostas que incorporaram parâmetros adicionais às expectativas e valências para explicar as tendências motivacionais do comportamento. Um bom exemplo desse alargamento do modelo E/V situa-se na “teoria dinâmica da ação” de Atkinson e Birch, (1978). Neste, medidas de ação dominante são representadas como o resultado final de um embate de forças instigadoras e consumatórias, por um lado, contra as forças de resistência e inibitórias, por outro. A ideia de incluir na explicação motivacional tanto as tendências de ação como as de evitação já estavam presentes em Atkinson (1957) sobre a conduta de realização. A novidade dessa proposta, de acordo com Barberá, reside no fato de que, ao incorporar parâmetros mediacionais (forças instigadoras e consumatórias/forças inibidoras e de resistência) entre expectativas/valências e as tendências motivacionais resultantes, surge a possibilidade teórica de produzir-se uma mudança no comportamento e na estrutura motivacional subjacente, mesmo quando as expectativas do sujeito e sua avaliação sobre a situação de vida permaneçam constantes.

Em suma, a partir dos anos sessenta proliferam uma série de modelos que tentam explicar a estreita relação entre os fatores motivacional e o comportamento, ou entre o conhecimento, motivação e ação. Apesar dos inegáveis progressos que representam algumas dessas propostas cognitivas a respeito de uma consideração mais ativa e auto reguladora da motivação (Bandura, 1982; Raynor, 1981), a série de teorias discutidas até aqui partilham uma abordagem comum, e que podem ser categorizadas como *reativas*.

Em contraste com uma abordagem reativa da motivação, podemos encontrar ainda, conforme vemos em Barberá (2002), um enfoque mais ativo do ser humano, onde este é considerado como agente causal de suas próprias ações. A motivação, segundo esse modelo, é interpretada como uma atividade que, muitas vezes, se manifesta espontaneamente, sem ser reduzida a uma mera reação a uma situação específica de estímulo. Além disso, o comportamento motivado é considerado como proposicional, posto que, em grande parte, se desenvolve impulsionado por planos, metas e objetivos.

A partir de uma perspectiva psicofisiológica, o conceito de motivação de Hebb (1955), definido como a tendência de cada organismo para produzir atividade organizada, e o

conhecimento do substrato neural da excitação, em torno da formação reticular e de núcleos talâmicos, constituem dois importantes pontos de apoio, de acordo com Hebb, para sustentar a concepção do ser humano como um agente causal. A subsequente crítica da representação de ativação como uma construção unitária (Vila & Fernandez, 1990) permitirá a incorporação de uma abordagem sistêmica, que leva em consideração tanto as estruturas neurais, central e periférica, como o papel modulador dos componentes neuroendócrinos e, especialmente, a série de interações que se estabelecem entre os referidos componentes. A identificação do sistema cerebral da motivação, a partir dos experimentos de Olds e Milner (1954) sobre a área cerebral responsável pela sensação de recompensa, ou *brain stimulation reward* (BSR), foi utilizada para se inferir a existência de um circuito neural cuja função é regular o nível de ativação e dar o impulso necessário para agir em uma determinada direção.

Ao contrário das abordagens reativas em motivação humana, teorias como as anteriores caracterizam-se basicamente por se interessarem principalmente em propriedades como espontaneidade e propositividade. Curiosidade, o desejo exploratório e o sentido de autodeterminação do comportamento são, em si, capazes de ativação psicológica e o objetivo de tal atividade não é para restabelecer uma situação anterior ou restaurar o equilíbrio quebrado, mas para guiar o comportamento a direção novas, desconhecidas e desafiadoras. Segundo Barberá (2002), a importância motivacional que a intencionalidade exerce sobre o comportamento a partir de um ponto de vista propositivo tem se consolidado nos últimos anos, em torno de uma teoria geral da ação claramente separada do esquema tradicional em torno do comportamento reativo. Os antecedentes mais imediatos de tais posições teóricas, ainda segundo Barberá, referem-se a psicologia filosófica do início do século. O desenvolvimento da motivação intrínseca e a investigação psicológica desenvolvida para explicar sua origem se interessam, sobretudo, pela análise da espontaneidade como traço motivacional característico, que está presente em várias atividades humanas. Por sua vez, o caráter propositivo, alinhado com a noção de intencionalidade, vai ocupar o foco das explicações motivacionais dos modelos de ação (Heckhausen, 1977; Kuhl, 1985).

A definição de motivação intrínseca sugere que grande parte das atividades humanas são realizadas pelo prazer pressuposto nessa atividade ou pelo interesse que sua execução acarreta. Uma questão-chave dentro dessa abordagem motivacional para a pesquisa psicológica tem girado em torno de se conhecer quais são os fatores subjacentes que permitem explicar a motivação intrínseca. Em última análise, o que faz com que uma atividade seja intrinsecamente motivadora? As respostas obtidas tem sido diversas e em alguns aspectos complementares.

Alguns pesquisadores têm se interessado em analisar as propriedades de alguns objetos específicos que os tornam inerentemente motivadores. Por exemplo, estudos de Berlyne (1960), aplicados num contexto educacional, incidem nas características de novidade, complexidade e imprevisibilidade, além da curiosidade, possuídas por determinadas atividades, enquanto determinantes de interesse motivacional. Um grau intermediário de cada uma dessas propriedades desperta o interesse dos sujeitos e promove a curiosidade e o desejo exploratório acerca delas. Para uma atividade ser moderadamente complexa, imprevisível, ou ter caráter de novidade, depende, em parte, da comparação entre informação derivada a partir de diferentes fontes.

De certa forma, pode-se dizer que a visão de esquema reagente é mantido nesse tipo de explicação, na medida em que as propriedades (novidade, complexidade, imprevisibilidade) de certos estímulos causam curiosidade e levam à exploração e manipulação de objetos para sua melhor compreensão; do mesmo modo como no esquema clássico hulliano a falta de alimentos gera fome e leva a busca de alimento para atender as necessidades. Neste sentido, a curiosidade pode ser representada como uma necessidade psicológica, atuando de maneira semelhante as necessidades básicas de fome ou sede, como mencionado anteriormente.

Porém, a pesquisa psicológica interessada na motivação intrínseca também tem indagado sobre suas origens dentro do sujeito, através da auto-percepção do sujeito como uma pessoa competente, eficaz e determinada a agir. As noções de auto-competência (White, 1959), de auto-eficácia (Bandura, 1982) e de ação personalizada (Nuttin, 1985) reforçam o protagonismo motivacional da subjetividade e o papel do indivíduo como agente causador da sua própria atividade comportamental. Cada um desses conceitos tentam definir, com maior precisão, a natureza da motivação intrínseca.

Contudo, quaisquer desses modelos explicativos da motivação, ao invés de responder a certas características específicas do estímulo, parecem estar mais interessados em oferecer oportunidades para que as pessoas testem suas habilidades e sua “determinação” de forma mais ativa e intencionalmente. As características de dinamismo e ativação interna prejudicam a noção de motivação intrínseca de atividades prazerosas resultantes, por exemplo, da experiência sensorial passiva. Como aponta Reeve (2006), podemos gostar de ir ao cinema ou ouvir música, mas não podemos definir essas atividades como comportamentos desencadeados por motivação intrínseca. Outra coisa bem distinta é que tais atividades sensoriais favoreçam em nós o interesse, por exemplo, em aprender teoria musical, aprender a tocar algum instrumento ou estudar cinematografia. Do ponto de vista que aqui se analisa, relativo a diferenciação entre as

teorias reativas e as teorias de ativação, é esta última consideração da motivação intrínseca a que apresenta o indivíduo ativo e espontaneamente motivado, que procura ser avaliado como competente e autodeterminado em relação ao ambiente.

Segundo Barberá (2002), junto com o desenvolvimento da motivação intrínseca, durante os anos oitenta, se formalizaram uma série de modelos psicológicos que incorporam a intenção e a vontade como aspectos relevantes do comportamento propositivo humano, atribuindo a esses aspectos um papel específico na explicação motivacional. Os dois mais representativos são a teoria de controle da ação de Kuhl (1985, 1986) e as fases da ação desenvolvidas por Heckhausen (1977). Ambas as explicações estão ligadas diretamente com a tradição alemã da escola de Wuerzburg, e a noção de intencionalidade como a força motriz de uma tendência determinante que incentiva a realização do comportamento. Para Barberá, as duas maiores contribuições desses modelos que serviram de apoio a uma abordagem ativa, e não meramente reativa, da motivação são: i) O estabelecimento de uma importante distinção conceitual entre tendência motivacional e alcance de metas, e ii) a ideia de que o comportamento dirigido a uma meta é hierarquicamente organizado, embora o desenvolvimento de sua estrutura e da fundamentação empírica da hierarquia ainda estejam por ser decifrados.

No que diz respeito ao ponto i), ambos os modelos apresentam uma análise diferencial entre os processos envolvidos na tomada de decisões e os que atuam agindo sobre a realização das metas propostas. Para alcançar uma meta voluntariamente é necessário, segundo esses modelos, ter a intenção de querer obtê-la. Porém, a intenção, enquanto conceito motivacional central que representa um nível de compromisso com a ação, não garante por si só o sucesso do resultado almejado. Na própria experiência da vida diária vemos a distância existente entre os objetivos e as realizações.

Em particular, Kuhl (1985, 1986) destaca como insuficiente o pressuposto implícito que realizam as teorias de E/V, ao identificar o comportamento motivado com meta. Para ele, ao contrário, seu modelo parte da distinção entre intenção e ação. Sua proposta busca sustentação em duas ideias básicas sobre: a) impulsos, desejos, expectativas, valorações e demais tendências motivacionais como determinantes do grau de compromisso com a ação, cujo nível mais elevado se situa na intenção de alcançar uma meta, e b) entre a intenção e o comportamento mediante uma série de processos volitivos complexos que tendem a impor-se frente diversas tendências difíceis, tanto externas como internas, para obter o objetivo proposto. São esses processos e estratégias, mais do que os parâmetros impulsivos, afetivos e/ou cognitivos,

determinantes do nível de compromisso com a ação, e que o modelo de Kuhl está mais interessado em estudar.

O modelo de Heckhausen (1977) representa, de forma geral, a mesma ideia de Kuhl referente a uma diferenciação entre os processos motivacionais e volitivos, incluindo em sua representação quatro fases e dois momentos-chave de reflexão (a travessia do Rubicão) como partes do processo de tomada de decisão, onde a fase de motivação seria anterior ao momento de decisão, denominada “fase de eleição”, e que seria separada das duas fases seguintes, essas ditas volitivas, pelo primeiro momento-chave de reflexão, onde se daria a decisão. A partir do momento que houve uma tomada de decisão consciente, o estado denominado de pré-ação, já seria uma fase volitiva, assim como a fase de ação, que seria seguida por outro momento-chave de reflexão, e uma volta a uma fase de motivação, a quarta e última fase de pós-ação, denominada “fase de valoração”.

A quarta fase do modelo incorpora, além disso, um novo componente motivacional a sequência, que corresponde a avaliação feita pelo sujeito da ação alcançada, ideia esta que tinha sido anteriormente delineada pela noção de expectativa de resultado e consequência. A incorporação da avaliação para o esclarecimento da atividade motivada dos comportamentos voluntários permite, por um lado, de acordo com essa teoria, estabelecer na representando um ciclo de retroalimentação, ou *feedback*, já que apenas um avaliação positiva dos resultados obtidos permite desativar definitivamente a intenção. Por outro lado, a avaliação da ação anterior possibilita implicitamente a avaliação das consequências, compreendendo assim o critério de que o futuro pode influenciar o presente, através das metas antecipadas e da avaliação das mesmas.

Um balanço sobre o estado da arte em psicologia da motivação revela características como a predominância de pesquisas aplicadas, com ênfase especial nos domínios da educação e do trabalho. A psicologia educacional afirma a importância dos processos motivacionais na formação e interações específicas que ocorrem entre a motivação e o desempenho acadêmico. Em ambientes organizacionais, os debates giram em torno do papel da motivação, intrínseca e extrínseca, na satisfação e desempenho no trabalho, com o interesse principal girando em torno de como essa relação pode ser traduzida em benefícios econômicos.

Uma revisão recente das áreas de pesquisa preferidas dentro do estudo da motivação feita por universidades espanholas (Barberá e Mateos, 2000) apresentou como resultado, por um lado, uma enorme dispersão temática e, por outro, um claro predomínio da investigação

aplicada sobre a pesquisa básica. A pesquisa psicológica atual está longe das velhas pretensões universalistas de conceitos motivacionais únicos, tais como “vontade”, “impulso”, “instinto”, “incentivo” ou “pulsão”. Reconhecendo que, no entanto, a maioria das atividades humanas são tão complexas, com suas motivações diversas em uma pluralidade de fatores nelas envolvidas, que torna muito difícil reunir em um único paradigma explicativo toda essa complexidade (Barberá, 2000).

No entanto, uma característica comum compartilhada pelos diversos modelos de motivação é a escolha do que se poderia denominar de um “quadro cognitivo”. Apesar da diversidade temática e mesmo metodológica, ou ainda do enfoque conceitual adotado, quase todos os autores assumem, explícita ou implicitamente, uma interpretação psicológica que prioriza os componentes racionais da motivação humana e a conduta voluntária destinada a atingir metas (Barberá, 2000; Barberá & Mateos, 2000). Uma tendência que parece estar surgindo em modelos motivacionais recentes, de acordo com Barberá (2002) é a influência da abordagem sistêmica, onde o foco não são nem os componentes motivacionais internos, nem os fatores ambientais, mas as relações interativas, de forma contínua, que ocorrem entre as pessoas e o universo subjetivo que é social e culturalmente construído. No entanto, também são apontadas questões ainda sem solução e que são direcionadas como os principais desafios para o futuro, como algumas questões referentes às interações estreitas entre motivação e emoções, bem como o desenvolvimento de modelos que tentam representar as maneiras pelas quais se dá a influência do entrelaçamento das emoções e motivações do conhecimento, além da atividade subjetiva e sua influência nos resultados comportamentais.

Por fim, como já dito anteriormente, o estudo da motivação, para além de se buscar entender as causas do comportamento, mas principalmente o que faz com que um sujeito se comporte de tal maneira, e o que faz com que ele mantenha aquele comportamento, também busca aplicar esse conhecimento de forma prática, em situações específicas. Busca-se motivar o sujeito a agir de determinada maneira. Por exemplo um aluno que não vai bem nos estudos pois não está motivado para tal. Um atleta que precisa se motivar mais em seus treinos para alcançar um melhor rendimento. Uma pessoa que é dependente de uma substância química e precisa de motivação para se livrar da dependência. Ao buscar aplicar os princípios motivacionais para resolver problemas práticos, busca-se motivar a si mesmo ou ao outro, por exemplo através da psicoterapia, para um determinado fim de interesse. Porém, nos perguntamos, o que guia esse interesse? O próprio estudo da motivação em si permite ao sujeito “manipular” o que lhe motiva ao tomar consciência do que motiva seu comportamento. Se o

sujeito pode tomar consciência, e escolher fortalecer, ou reforçar, ou mesmo ir contra algo que o motiva, racionalmente criando, escolhendo novos caminhos ou comportamentos desejados e se motivando portanto em função dessa escolha, há algo aí que da margem a uma escolha do sujeito, ou mesmo do psicoterapeuta que junto com o sujeito irá reforçar motivações para promover ou manter comportamentos desejados.

Essa seletividade, que remete ao que Neisser (1976) falou da seletividade da atenção, parece ser intrínseca ao sujeito, e até mesmo anterior ao que lhe motiva, pois pode de certa forma buscar tomar controle de sua própria motivação. Se esse domínio sobre o que lhe motiva vai ser forte o suficiente já seria outra questão, mas em potência, ao sujeito parece ser permitido buscar uma autonomia, se motivar e ter uma certa liberdade de agir segundo motivos que lhe são caros. A própria possibilidade de aplicação dos princípios motivacionais parece sugerir essa hipótese. Por um lado parece que somos determinados por estes motivos, externos e internos; porém, por outro lado, parece que podemos, ao tomarmos consciência dos motivos que nos afetam, termos um certo controle e usarmos esse conhecimento segundo nossa vontade. Essa seletividade e controle, remete a um outro conceito que trataremos a seguir.

3.5. Função Executiva

Tomando como referência Mourão Junior e Melo (2011), abordaremos aqui o conceito de função executiva, que também pode ser conhecido como controle executivo ou sistema supervisor atencional e, em alguns autores ainda pode ser encontrado no plural, devido a existirem diversas funções executivas, da mesma forma que existem diversas modalidades de memória, de linguagem ou atenção. Assim como essas últimas funções cognitivas são usadas comumente no singular, preferimos utilizar o termo função executiva da mesma forma no singular, a fim de evitar confusões e ambiguidades conceituais.

Função executiva é um conceito neuropsicológico que se refere a um sistema gerenciador e que tem como função organizar uma sequência de ações a fim de atingir um objetivo definido. Sendo um termo abrangente para a gestão, ou seja, regulação e controle, dos processos cognitivos de forma geral, incluindo memória de trabalho, raciocínio, resolução de problemas, bem como planejamento e execução. A função executiva do cérebro vem sendo definida, principalmente no âmbito das neurociências, mas também tem ganhado crescente destaque na psicologia cognitiva, como um conjunto de habilidades que de forma integrada possibilitam ao indivíduo o direcionamento do comportamento a objetivos, ou seja, a realização

de ações voluntárias. Tais ações ditas voluntárias são auto-organizadas mediante a avaliação de sua adequação e eficiência em relação ao objetivo pretendido de modo a escolher, entre as estratégias disponíveis, as mais eficientes, buscando resolver problemas imediatos, e/ou de médio e longo prazo (Capovilla, Assef, & Cozza, 2007; Mourão Junior & Melo, 2011).

Em psicologia cognitiva, a função executiva muitas vezes é ligada a questão da distinção estabelecida experimentalmente por Richard Schiffrin e Walter Schneider (1977) entre os dois modos básicos de processamento de informações: sendo o primeiro os *processos automáticos*, caracterizados por pouca ou nenhuma carga de atenção, sem controle, não dependentes da consciência, por operações paralelas e de execução rápida; e o segundo, os *processos controlados*, que são mais lentos, seriais, estrategicamente determinados e, portanto, incorrem em uma carga de atenção na memória. Os processos controlados, pelo menos aqueles envolvidos em tarefas cognitivas complexas, são frequentemente apontados como dependentes de um sistema cognitivo de nível superior, como o sistema executivo central de Alan Baddeley (1986) na memória de trabalho ou o sistema de supervisão atencional de Don Norman e Tim Shallice (1988). Esses sistemas propostos pelos autores citados, que funcionam como organizadores e controladores da memória de trabalho e atenção, são apontados como possuindo localização cerebral equivalente aquelas apontadas pela neurociência relativa as funções executivas, e são da mesma forma afetadas por acidentes no córtex pré-frontal. Outras formas de controle, ou comportamentos ditos voluntários, são provavelmente mais altamente distribuídos, como apontado, por exemplo, pelo fisiologista Alain Berthoz (2000) em seus estudos sobre o movimento.

Seguindo John Flavell (1979), a função executiva têm sido estudadas por sua função metacognitiva em várias atividades, incluindo o processamento da memória (*metamemória*), a aquisição da linguagem (*metalinguagem*), comunicação (*metacomunicação*), e na construção de teorias da mente. Análises realizadas dentro desse quadro ou estrutura têm lidado essencialmente com o *metaconhecimento* (conhecimento sobre as funções cognitivas) usado pelos processos de controle. Em psicologia infantil, por exemplo, modelos neo-piagetianos enfatizam especificamente o papel dos processos de controle executivo, ou funções executivas, na consideração para as etapas e mecanismos do desenvolvimento cognitivo. Robbie Case (1985) descreve uma hierarquia de etapas entre a infância e a idade adulta (sensório-motor, inter-relacional, dimensional e vectorial), cada uma dessas divididas em subestágios (unifocal, bifocal e elaborado), no decurso dos quais as estruturas de controle executivo se tornam mais e mais complexas. A transição entre subestágios é uma função da capacidade da memória de

trabalho maior, ela própria uma função da eficiência operatória (esquema de automação); a transição entre as fases é assegurada pela integração de estruturas de controle hierárquico cada vez mais potentes, tanto cognitivas e sociais.

Para Houdé (2004), quer a função executiva refira-se a ações, percepções, ou representações na memória, uma das questões-chave na pesquisa de hoje, tanto em psicologia cognitiva quanto na neurociência, diz respeito às respectivas funções de ativação e inibição na nossa arquitetura funcional de vários níveis, onde o controle é exercido em todos os níveis, desde a comunicação entre os neurônios, até as funções executivas no córtex pré-frontal.

Alguns neurofisiologistas, entre eles Wolf Singer (1990), entretanto, rejeitam a ideia de um controlador executivo central, ou seja, de uma função executiva central, especificamente localizada, e que tem controle de tudo; defendendo, ao invés disso, um processo de sincronização temporal (*temporal synchronization process*). Neste ponto de vista, os grupos de neurônios ou módulos cognitivos envolvidos em uma determinada tarefa tornam-se temporariamente ligados através da sincronização de sua atividade elétrica (que oscila a uma frequência de cerca de 40 Hz). A rede momentânea que emerge (chamada de “*neural network*”) capta todos os recursos disponíveis do sujeito, entre eles, obviamente, sua atenção e consciência, para a realização de uma determinada tarefa. Para cada nova tarefa, uma nova rede é construída, e assim por diante. Aqui, a unidade da mente é o resultado da associação transitória de módulos especializados, e não de um supervisor de alto nível. Esta visão joga definitivamente com a ideia de um “teatro cartesiano” (para usar a expressão do filósofo Daniel Dennett, 1991) no núcleo do sistema cognitivo. Porém, esse é um debate também ainda em aberto. Estes dois processos (seleção por um supervisor central e sincronização temporal) podem estar envolvidos de formas relacionadas ou não, de acordo com a estrutura do material e as exigências de uma tarefa desempenhada (Houdé, 2004).

De qualquer forma, sempre que se faz necessário formular planos de ação ou quando uma sequência de respostas apropriadas deve ser selecionada e esquematizada, ou seja, quando uma decisão é feita, a função executiva é acionada. Deste modo, a função executiva está relacionada diretamente aos fenômenos de flexibilidade cognitiva e de tomada de decisões. Atualmente é sabido que os módulos corticais atuantes nas funções executivas se localizam nos lobos frontais direito e esquerdo (Alvarez & Emory, 2006; Mourão Junior & Melo, 2011).

A fisiologia dos lobos frontais, de acordo com Mourão Junior e Melo (2011), por muito tempo foi pouco conhecida, mesmo quando outras áreas do cérebro continham descrições bem

mais elaboradas das funções que desempenhavam. Assim, muito tempo se passou até que os neurocientistas pudessem compreender a importância representada pelos lobos frontais para a cognição. Ainda que o cérebro seja constituído por diversos componentes que desempenham diferentes funções, contudo, os lobos frontais não demonstram essa especificidade de apresentar uma função única que seja prontamente classificável. Desta maneira, por um longo tempo, foram considerados “lobos silenciosos” (Goldberg, 2002). Com o agora famoso acidente, ocorrido em 1848, e relatado por António Damásio (1996) em seu livro *O Erro de Descartes*, onde o jovem Phineas Gage teve o crânio atravessado por uma barra de ferro em uma explosão, da qual sobreviveu, e espantosamente ainda se manteve lúcido nas horas que se seguiram, os lobos silenciosos ganharam vozes. Embora a barra de ferro tenha atravessado-os e destruído a parte anterior do cérebro, o jovem Gage ainda podia ver, ouvir, sentir, falar e movimentar seus músculos; entretanto, de alguma forma, houve uma modificação drástica em sua personalidade. Ele passou a se mostrar incapaz de se adaptar às normas, sua moralidade fora destruída de alguma forma, e ele nunca mais foi o mesmo (Damásio, 1996). Porém, o caso de Gage trouxe um melhor entendimento da fisiologia do lobo frontal ao sugerir que:

... existiam sistemas no cérebro humano mais dedicados do que quaisquer outros às dimensões pessoais e sociais do raciocínio. A observância de convenções sociais e regras éticas previamente adquiridas poderia ser perdida como resultado de uma lesão cerebral, mesmo quando nem o intelecto de base nem a linguagem se mostrassem comprometidos [...] algo no cérebro estava envolvido especialmente em propriedades humanas únicas, e entre elas se encontrava a capacidade de antecipar o futuro e de elaborar planos, de acordo com essa antecipação, no contexto de um ambiente social complexo; o sentido de responsabilidade perante si próprio e perante os outros; a capacidade de orquestrar deliberadamente sua própria sobrevivência sob o comando do livre-arbítrio. (Damásio, 1996, pp.30-31).

Assim, pode-se dizer, segundo Mourão Junior e Melo (2011), que a função exercida pelos lobos frontais parece ser mais metacognitiva do que propriamente cognitiva, já que ela não se refere a nenhuma habilidade mental específica, contudo abrange todas elas. Por este motivo, ainda segundo Mourão Junior e Melo, a função dos lobos frontais é chamada de função executiva e, mais especificamente, o córtex pré-frontal, que é considerado como uma região filogeneticamente mais moderna do cérebro humano. Esta compreende as regiões do lobo frontal anteriores ao córtex motor primário e desempenham papel essencial na formação de metas e objetivos, no planejamento de estratégias de ação essenciais para se alcançar estes objetivos, além de selecionar também as habilidades cognitivas requeridas para a implementação dos planos, assim como a coordenação dos mesmos para aplicá-los na ordem correta. Além disso, o córtex pré-frontal também é apontado como responsável pela avaliação do sucesso ou fracasso das ações dirigidas a metas estabelecidas (Goldberg, 2002).

Fuster (2003) postulou que a função executiva opera por meio de redes neurais interativas e sobrepostas, distribuídas nos córtices de associação, principalmente o córtex pré-frontal. Essas redes, que são denominadas cognitos, alimentam o ciclo percepção-ação, constituindo as unidades básicas do processamento executivo. As conexões do córtex pré-frontal lateral com o hipocampo e com o córtex parietal posterior são especialmente importantes para os aspectos cognitivos de todas as formas de comportamento. E, por intermédio dessas conexões, torna-se possível a mais importante função do córtex pré-frontal, de acordo com Fuster, que é a integração temporal de ações para o cumprimento de metas. O processo neural de integrar – onde integrar significa incluir, excluir e organizar elementos em um conjunto, formando um todo coerente – as informações ao longo da linha do tempo, por intermédio da ordenação dos cognitos, é a base para a programação temporal das ações.

A organização no tempo de novas e complexas sequências de comportamento é possível graças a integração de estímulos externos (sensoriais) e estímulos internos (memórias armazenadas). Dito de outra forma, a integração temporal nada mais é do que o processamento dos estímulos que chegam ao cérebro no tempo presente e das memórias armazenadas, que remetem a um tempo passado. Em função de sua especialização na estruturação temporal de novas e complexas séries de comportamentos direcionados a metas, além da participação na escolha entre as alternativas e na tomada de decisão, o córtex pré-frontal foi considerado o centro executivo do cérebro (Fuster, 2002).

De acordo com Mourão Junior e Melo (2011), a fim de desempenhar esse papel integrativo, o córtex pré-frontal precisa ter acesso a todos os componentes de informação sensorial, motora, e mnemônica que dão forma à estrutura do comportamento. Deste modo, pode-se dizer então que a execução do comportamento temporalmente estruturado ocorreria devido a ativação de um comando central e da ativação adequada dos componentes de suas redes neurais.

Ainda segundo Mourão Junior e Melo (2011), para fins didáticos, uma subdivisão da função executiva de integração temporal do córtex pré-frontal pode ser feita levando-se em conta três funções cognitivas: ajuste preparatório, controle inibitório e memória de trabalho. De acordo com esses autores, uma análise psicológica e fisiológica do funcionamento das três regiões anatômicas pré-frontais (lateral, medial e orbital) produz a seguinte correspondência topográfica das funções cognitivas: as três regiões pré-frontais estão envolvidas em um ou outro aspecto da atenção; a região medial e o giro cingulado anterior estão envolvidos na movimentação e na motivação; a região lateral se relaciona com o ajuste preparatório e com a

memória de trabalho, e a região orbital (e, em alguma medida, também a medial) se relaciona com o controle inibitório de impulsos e interferências.

O ajuste preparatório é uma função prospectiva (dirigida para o futuro), com sede no córtex pré-frontal lateral, e que prepara o organismo para ações que dependem das informações recebidas. Esta função de ajuste do córtex pré-frontal lateral pode ser justificada devido a evidências eletrofisiológicas. Entre um *input* sensorial e a resposta motora pertinente, potenciais lentos podem ser gravados na superfície do lobo frontal, os quais são relacionados com a duração do tempo de reação e a exatidão da resposta. Pode-se identificar, a partir desse procedimento, dois tipos de potenciais, embora ambos pareçam ser parte de uma série contínua. O primeiro é o contingente de variação negativa (CVN), chamado de “onda de expectativa”, que é relacionado à necessidade de negociar a contingência transtemporal entre o estímulo e a resposta. O segundo é o potencial de prontidão (PP), já mencionado anteriormente na discussão dos experimentos de Benjamin Libet, e está relacionado à preparação de uma ação motora. Pacientes com lesão pré-frontal apresentam um déficit significativo na habilidade de planejar uma ação futura que parece refletir a falha da função do ajuste preparatório (Fuster, 2003).

Outra função do córtex pré-frontal, segundo Mourão Junior e Melo (2011), que possui caráter integrativo temporal é o controle inibitório. Este consiste na capacidade de inibir respostas inadequadas ou respostas a estímulos distratores que possam interromper o curso efetivo de uma ação ou uma resposta adequada em andamento. Dificuldades referentes ao controle inibitório são, frequentemente, associadas à impulsividade. De acordo com Fuster (2002), experiências com lesões em animais e também observações clínicas indicam que o mecanismo neural para esta função inibitória reside nas porções medial e orbital do córtex pré-frontal. A finalidade fisiológica desta função é a supressão de entradas internas (por intermédio da memória) e externas (por intermédio dos sentidos) que possam interferir na estruturação do comportamento, do discurso, ou da cognição e que estejam a ponto de serem empreendidas, ou que já estejam em curso. O controle inibitório do córtex pré-frontal, operando através de seus efeitos seletivos, como por exemplo focalização da figura e inibição do fundo, é essencial para a integridade do sistema atencional.

Um dos indicadores do desenvolvimento psicossocial da criança, como sugere Mourão Junior e Melo (2011), é justamente o estabelecimento progressivo do controle inibitório sobre os impulsos internos, sobre o sensorio e sobre a mobilidade. Conforme a criança cresce, os dois componentes principais da atenção, inclusivo e exclusivo, vão amadurecendo de forma gradual e a criança vai se tornando cada vez mais capaz de focalizar e concentrar a atenção nas tarefas

realizadas, tornando-se, desta maneira, menos suscetível a distrações, menos impulsiva, e com maior capacidade de exercer seu autocontrole.

Por fim, um outro mecanismo de caráter integrativo temporal é a memória de trabalho. O conceito teórico mais comum de memória de trabalho, conforme apresentado por Mourão Junior e Melo (2011), descreve a mesma como um sistema de capacidade limitada, que mantém e armazena informações temporariamente, a fim de sustentar os processos de pensamento humano, fornecendo desta maneira uma interface entre percepção, memória de longo prazo e ação.

Como podemos notar, essas funções estão em estreita relação com aspectos importantes e que são enumerados como fundamentais para uma possível decisão livre, ou para um comportamento que seja considerado como livre. Como, por exemplo, autocontrole, foco atencional, planejamento futuro na tomada de decisões, integração dos processos cognitivos e outros. Não é por menos que alguns estudos tem apontado essa área do cérebro como uma área estreitamente relacionada e que fundamenta as ações voluntárias, ou como área que dá início a tais ações (Fried, Mukamel & Kreiman, 2011; Goldberg, 2002; Haggard, 2008; Soon, Brass, Heinze & Haynes, 2008). Porém, grande parte desses estudos se limitam apenas a localizar regiões do cérebro que parecem estar envolvidas com essas atividades relacionadas a ação voluntária, tomada de decisão, etc.

No que concerne a essa investigação, essas questões não são nosso foco, a não ser pelo fato de, para que essa identificação seja feita, ou seja, para que se busque localizar no cérebro estruturas relacionadas por exemplo a ação livre, obviamente, mesmo que implicitamente, ou de forma não intencional, deve-se admitir que ações livres são possíveis. O fato de haver uma estrutura que esteja ligada a tomada de decisões, ou a comportamento voluntário em si não implica necessariamente que uma vontade livre seja ilusória ou que o que causa a ação voluntária seja o cérebro. Argumentações desse tipo nos parecem o mesmo que falar que o que causou a morte de alguém que levou um tiro foi o dedo que pressionou o gatilho, ou melhor, a área do cérebro que ao ser ativada provocou o movimento do dedo e não o sujeito que atirou. Outra forma de ver essa questão é pensar que o que esses autores fazem seria apenas mostrar que, para ações comumente consideradas como ações voluntárias, existem áreas específicas do cérebro cuja atividade estão estreitamente correlacionadas, tanto que, no caso de dano desses áreas, a ação dita voluntária seria prejudicada. O que, bem da verdade, não toca o âmago da questão que nos interessa, já que não se pode negar ou afirmar a liberdade de escolha, ou que possuímos ou não livre arbítrio a partir desse tipo de constatação.

3.6. É possível se falar em vontade livre a partir dos pressupostos da psicologia cognitiva?

Buscarei neste tópico oferecer uma síntese, a luz da bibliografia considerada, do que se considera ‘vontade’ na ciência cognitiva, comparando-a com conceitos próximos. Mais ainda, oferecerei minha avaliação de se é possível, diante do que vimos dos pressupostos e realizações particulares da psicologia cognitiva, falar em uma vontade livre, de acordo com o que é entendido como o problema da vontade na filosofia.

Obviamente, para se responder a essa pergunta, inicialmente devemos delimitar o que se quer dizer por *vontade* e o que se quer dizer por *liberdade*, e ainda, o que é necessário para que a vontade seja livre. Eis aí, no nosso ponto de vista, o maior problema dessa questão. Todo o debate envolto no tema como visto no primeiro capítulo parece principalmente girar em torno desse problema. Diversos autores deram sua definição de vontade e igualmente do que seria necessário para que a vontade fosse considerada livre. Como vimos, o debate relativo ao determinismo/indeterminismo, compatibilismo/incompatibilismo, gira em torno de se definir o que seria liberdade e se essa definição é compatível ou não com nosso entendimento de mundo.

Por outro lado, ainda que não exista uma definição de vontade que seja consensual, pôde-se falar em vontade como uma característica ou atributo da psique humana desde os filósofos pré-socráticos até os dias atuais e ser minimamente compreendido. Da mesma forma, o mesmo pode ser dito sobre o conceito de liberdade e livre arbítrio. Como vimos, alguns autores apontaram a relação do conceito de vontade tal como tratado na filosofia e uma área da psicologia cognitiva que trata da motivação. O próprio Miller et al. (1960) afirma que a vontade foi “assimilada anonimamente no tópico mais amplo da motivação”. Ao abordar a motivação, esta em geral é vista como a tentativa de se estudar as causas do comportamento, ou a busca dos determinantes da atividade humana, o que nos motiva a agir como agimos. Vista também como uma espécie de força, ou conjunto de forças, que “energiza”, ou direcionam o comportamento. Alguns estudos (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Gailliot et al., 2007; Vohs et al., 2008) sugerem que o que chamamos de força de vontade, no senso comum, ou seja, uma energia que utilizamos ao fazer e manter escolhas, tem uma forte ligação com uma capacidade de auto-controle, sendo algo que pode ser constatado fisicamente, e que temos realmente uma fonte de recursos de energia que é despendida com atividades dessa natureza. O uso dessa “energia” provoca um esgotamento do ego (*ego depletion*), e tem influência em ações futuras que exigem auto-controle e tomada de decisões.

Voltando ao ponto da relação entre motivação e vontade, não é problemático afirmar que o conceito dominante de motivação no cognitivismo como aquilo que causa o comportamento é bem próximo do conceito de vontade como adotado em uma parte da tradição filosófica, como vimos, por exemplo, na vontade aristotélica como motor da ação; em São Tomás de Aquino onde é a vontade que move o agente para um fim (porém em Aquino fazemos isto em virtude de nosso intelecto à luz de seu desejo fundamental para o Bem); em Hume, onde liberdade é o poder de agir ou não agir de acordo com a determinação da vontade ; ou mesmo em Kant, onde “a *vontade* é uma espécie de causalidade dos seres vivos” (2001b, p. 93); entre outros.

Porém, ao avaliarmos a motivação no tópico relativo, notamos uma certa confusão em torno do conceito de motivação nas ciências cognitivas, não havendo consenso no campo. Essa mesma confusão, ou falta de acordo ou consenso, pôde ser observada em torno do conceito de vontade. Notamos também que, em alguns estudiosos da motivação (por exemplo Kuhl, 1985, 1986; e Heckhausen, 1977), é feito uma diferenciação explícita entre processos motivacionais e volitivos. O que contradiz de certa forma a afirmação de Reeve (2006) de que a motivação seria uma continuação, ou se coloca no lugar de uma explicação do comportamento pela vontade. Como apontado por Barberá (2002), parece existir uma tendência atual no campo da motivação em se considerar aspectos como intencionalidade e processos volitivos como determinantes da motivação. Percebemos, portanto, ainda que não se fale explicitamente em vontade, um uso corriqueiro de termos como ‘volitivo’ ou ‘comportamentos voluntários’, não só no campo da motivação como também em outros estudos da ciência cognitiva em geral, entre eles a neurociência, com o conceito de função executiva.

Um aspecto importante do problema da vontade diz respeito a como algo mental pode ter influência sobre o físico, ou seja, está relacionado ao problema da relação mente/corpo. Quanto a esse aspecto, a posição nas ciências cognitivas parece ser mais uniforme, levando em conta a adoção do funcionalismo. É permitido no funcionalismo, de modo geral, identificar estados e processos mentais por meio de seus papéis causais e, da mesma forma, sabemos que os papéis funcionais são possuídos por estados e processos neurais. É claro que o termo “funcionalismo” tem sido usado também de forma vaga e de diferentes maneiras, como mostramos em nossa discussão no referido tópico. Apesar disso, podemos dizer que, segundo uma visão mais corrente desta abordagem, o que faz algo como um pensamento, desejo, dor (ou mesmo uma volição) não depende tanto de sua constituição interna, mas sim de sua função, o papel que desempenha, no sistema cognitivo do qual é parte. Desta maneira, teorias

funcionalistas assumem a identidade de um estado mental a ser determinado por suas relações causais com estímulos sensoriais, outros estados mentais e comportamento. Podemos falar portanto, seguindo Levin (2013), que o funcionalismo seria compatível mesmo com um tipo de dualismo que leva estados mentais a causar, e serem causados por, estados físicos.

Quanto ao problema da liberdade ou determinação da vontade, nos encontramos em maior controvérsia. A própria definição mais aceita da psicologia cognitiva como abordagem para a investigação da cognição baseada na ideia de processamento de informação parece fazer crer em uma visão determinista, e até mecanicista do funcionamento mental humano. Se funcionamos de acordo com regras, portanto nosso comportamento seria determinado por essas regras. Como vimos, essa visão parece ter uma certa sustentação quando relacionamos o funcionamento de nossa mente com o de um computador. Porém, essa analogia parece ser exagerada, como reconhecido pelo próprio Neisser (1976). Não existem evidências científicas que sustentem, por exemplo, que o cérebro funciona exatamente como um computador. Assim, como vimos, essa crença parece advir de uma supervalorização da tese de Church-Turing, visto que esta não implica que o cérebro (ou a mente, ou a consciência) possa ser simulado por uma máquina de Turing, nem mesmo em conjunto com a crença de que o cérebro (ou mente, etc.) seja cientificamente explicável, ou exiba um padrão sistemático de respostas ao meio ambiente, ou ainda, seja “governado por regras”.

Um outro fator que poderia ser apontado como direcionador de nossas escolhas, seriam os motivos apontados pelo estudo da motivação. Mesmo aí ainda se consideram fatores intencionais e admite-se um certo controle, uma seletividade, que permite que se possa manipular de forma favorável esses motivos para se atingir uma meta, ou objetivos. Consideramos o fator tomada de consciência como um fator essencial para que se possa ter controle de potenciais motivos influenciadores do comportamento. Para Neisser (1976), com o desenvolvimento e refinamento cognitivo nos tornamos “mais livres” – por isso o adulto pode ser visto como mais independente que uma criança, visto ser menos “limitado aos estímulos” e mais “internamente direcionado”. O conhecimento também nos tornaria menos suscetíveis a sermos manipulados ou controlados, uma vez que ao possuírmos uma maior sofisticação em nossas habilidades cognitivas, poderíamos utilizar de um nível mais apurado de informação, o que faria com que não somente tivéssemos mais controle de nosso comportamento como também nos proporcionaria maiores possibilidades de ação.

Assim, como podemos nos orientar por metas, valores, fins, a consciência intencional se torna foco de atividade capaz de produzir comportamento (Penna, 1986). Passa-se a

considerar não só causas como determinantes do comportamento, mas também razões, reintroduzindo uma visão explicativa teleológica, que pôde ser elucidada pela aplicação do modelo da unidade TOTE proposto por Miller et al. (1960). O conceito de função executiva também pode dar sustentação a essa visão, visto existir um sistema identificado com uma região cerebral que tem como função justamente gerenciar os processos cognitivos, organizar sequências de ação mobilizando um conjunto de habilidades, que de forma integrada, vão possibilitar ao indivíduo direcionar comportamentos a objetivos, ou seja, a realização de ações ditas voluntárias.

Essa priorização de aspectos racionais que direcionam o comportamento, porém, pode ser um possível ponto de crítica sobre uma visão cognitivista aqui levantada. Na parte histórica de como o conceito de vontade foi tratado na filosofia, mostramos como a tradição intelectualista apresentava uma tendência a considerar que a vontade fosse dirigida por “razões”, ou preferências, que podem ser objeto de deliberação. Essa priorização de aspectos racionais pode estar sujeita as mesmas críticas direcionadas à essa tradição intelectualista da vontade. Como vimos, na psicologia cognitiva há realmente um maior foco nas pesquisas relativas à esses aspectos. Ainda assim, questões referentes a interação entre cognição e emoções, bem como o desenvolvimento de modelos que tentam representar as maneiras pelas quais se dão essa interação sejam apontados como desafios futuros, e novas tentandências de pesquisa, na psicologia cognitiva.

O compromisso do cognitivismo com o determinismo não é tão claro, isto porque, de certa maneira, aceitam-se tanto *causas* quanto *razões* como determinantes do comportamento (Castañon, Justi & Araujo, 2014). Podemos fazer uma leitura desse posicionamento quanto ao determinismo de um ponto de vista kantiano, por exemplo. Para Kant, como vimos no primeiro capítulo, determinismo não se opõe à liberdade da vontade. Aplicado ao comportamento humano, parece ser sustentável um duplo nível explicativo, tanto por causas, quanto por razões. Quando se examina um comportamento já realizado, e se busca explicar esse comportamento como fenômeno dado, este pode ser visto de uma forma determinista, visto o fenômeno ser necessariamente organizado por nossa razão de acordo com a lei da causalidade. Por outro lado, em um nível intelectual, podemos pensar em uma liberdade que nos guia, ou melhor, que nos permite escolher razões para nos guiarmos. E, assim que concretizamos essas razões como comportamentos dados como fenômenos, estes estariam submetidos, por sua vez, a categoria da causalidade.

De um ponto de vista kantiano, o princípio da causalidade incide sob o objeto tomado como fenômeno, aí sim a vontade pode ser pensada conforme a lei natural, sendo determinada e não-livre. Ao mesmo tempo, a vontade pode ser pensada enquanto coisa em si mesma, logo, não submetida à lei natural, sendo, portanto, livre. Poderíamos, assim, considerar a causalidade por duas perspectivas: “como *inteligível* quanto à sua *ação*, considerada a de uma coisa em si, e como *sensível* pelos seus *efeitos*, enquanto fenômeno no mundo sensível” (Kant, 2001a, p.466). Uma vez que estamos limitados a uma estrutura (nossa própria estrutura cognitiva) que nos permite conhecer o mundo e a nós mesmos, num sentido kantiano, mesmo que limitados a essa estrutura, não somos determinados por ela, uma vez que somos nós que a possuímos e, por esse mesmo motivo, estamos para além dela. Há algo em nós que escapa ser limitado a própria razão, escapa a nossa própria estrutura de conhecer. Se, o fruto desse algo tem consequências dentro desse limite dos fenômenos que podem ser conhecidos, obviamente, esses só serão entendidos dentro desses limites. Como se uma parcela da causa permanecesse sempre indeterminada, pois não se limita ao que pode ser conhecido, já que não se enquadra dentro da própria estrutura da razão. Estrutura essa, como vemos em Kant (2001a), submetida a categorias, tais quais a própria causalidade. Portanto, quando dizemos 'uma parcela da causa', não podemos a priori determinar como causa, pois essa é uma categoria que só pode ser aplicada aos fenômenos, e essa espontaneidade pura a qual tentamos nos remeter pertence a um reino que não se limita às categorias do conhecimento. Nesse sentido podemos falar em liberdade (como livre de causas), não como algo que conhecemos, mas como algo que nos dá a possibilidade mesma de conhecer, e que por si só, nesse sentido, é anterior (logicamente) ao conhecimento e, portanto, não é passível de ser conhecida, categorizada, determinada.

Assim sendo, essa dupla interpretação do comportamento seria possível porque teríamos no sujeito do mundo dos sentidos um *caráter empírico*, onde suas ações como fenômenos estão conectadas de forma causal com outros fenômenos, segundo leis constantes da natureza; e, ao mesmo tempo, um *caráter inteligível*, que embora seja causa de suas ações como fenômenos, não se subordina ele mesmo a quaisquer condições do mundo dos sentidos, não sendo fenômeno, mas uma coisa em si mesma. Desta maneira, Kant afirma que a liberdade e a natureza, cada qual com seu significado pleno, se encontrariam, ao mesmo tempo e sem conflito, nas mesmas ações, conforme referência à sua causa sensível ou inteligível.

Entretanto, essa é apenas uma das possíveis formas de se tentar compreender o problema da determinação do comportamento, que cai na questão do compatibilismo / incompatibilismo, e que talvez extrapole em muito o âmbito de uma investigação empírica da psicologia cognitiva.

Buscamos mostrar aqui como pode ser rica uma tentativa de maior aproximação, por um lado do debate filosófico em torno do conceito de vontade e questões relacionadas, e por outro os avanços e conquistas da psicologia cognitiva. Obviamente, não esgotamos aqui todos os possíveis entrelaçamentos e discussões que poderiam ser levantadas. Como vimos, o problema da vontade é um assunto por demais extenso e complexo mas ao mesmo tempo central da psicologia, pois define o aspecto mais importante da própria natureza humana e sua posição na natureza.

Quanto à questão levantada aqui, de se é possível se falar em vontade livre a partir dos pressupostos da psicologia cognitiva, posso finalmente afirmar que considero, por todo o exposto, que sim. Agora, quanto a questão da existência de uma resposta definida da abordagem ao problema, acredito que a resposta é não.

Acredito que uma eventual tomada de posição nesse problema por parte das ciências cognitivas seja por demais problemática, sendo muito preocupante que alguns cientistas, sem base científica sólida, assumam posicionamentos negando a possibilidade de uma vontade livre. Fazem isso geralmente se baseando em argumentos não tão sólidos quanto querem que o público leigo acredite que sejam. Para negar que temos uma vontade livre, precisam negar a própria consciência que temos de nossa liberdade, afirmando com isso que essa se trata apenas de uma ilusão. Que a própria consciência é um pseudofenômeno, ou algo que não tem real importância quando atuamos no mundo. Quanto a isso, a posição de William James nos parece inspiradora:

*O primeiro ato de uma vontade dotada de liberdade deve ser sustentar sua crença na própria liberdade. Eu, em conformidade com isto, acredito livremente em minha liberdade. Faço-o com a melhor da consciência científica. . .*¹⁰

¹⁰ “The very first act of a will endowed with freedom should be to sustain the belief in the freedom itself. I accordingly believe freely in my freedom. I do so with the best of scientific consciences. . .” (1899/1946, p.192)

CONCLUSÃO

O presente trabalho fez um levantamento de formas como o conceito de vontade é tratado na psicologia cognitiva. Apesar de algumas dificuldades, por este conceito ter, de certa forma, se dissolvido em estudos de motivação e da função executiva. Concluímos que o conceito de vontade pode ser definido como um fenômeno psicológico concebido pelo sujeito onde esse fenômeno se dá como algo que move ou causa suas ações. Ou ainda como aquilo que, identificado com o próprio eu do sujeito, *seleciona* entre opções de comportamento disponibilizadas pela cognição. Uma volição, que direciona intencionalmente à uma meta, iniciando e executando a conduta. Pode ser considerada livre pelo sujeito quando percebida como partindo de si mesmo, sem constrangimento externo, ou quanto resulta em um escolha sem restrições, por questões fora do controle da pessoa que age. Nesse sentido envolve responsabilização do sujeito por suas ações.

Retomando nossos passos, buscamos estabelecer, no primeiro capítulo, os sentidos dos termos vontade e liberdade da vontade, ou livre arbítrio, e a questão do compatibilismo/incompatibilismo, central na discussão sobre o tema. Como vimos, o *problema da vontade* está intimamente atravessado pelos problemas ligados aos conceitos de causalidade, acaso, determinação, indeterminação, assim como a responsabilidade moral da ação. Esse tipo de debate é de fundamental importância visto que o problema da liberdade da vontade está estreitamente relacionado com importantes questões no campo da ética e da filosofia do direito. Além disso, vimos diversos modelos propostos para se abordar a vontade humana e como cada um deles se enquadra a visões de mundo possível, referentes a compatibilidade ou não ao determinismo, e argumentos para defender e refutar essas posições. Esse é ainda um debate vivo, em aberto e que vem apresentando renovada efervescência, com grande número de publicações nas últimas décadas.

No segundo capítulo, apresentamos uma definição de psicologia cognitiva e de como ela se enquadra como uma das disciplinas do projeto mais abrangente de ciência da cognição, buscando explicitar ainda alguns de seus pressupostos epistemológicos a fim de rastrear possíveis filiações filosóficas ligadas ao problema da vontade. Procuramos também oferecer um resumo de como as outras disciplinas desse projeto de ciência da cognição, como por exemplo a inteligência artificial, filosofia da mente, linguística, e as neurociências, influenciaram a visão de homem da psicologia cognitiva. Entre essas, fizemos uma exposição

das contribuições da ciência da computação à psicologia cognitiva, notadamente do trabalho de Alan Turing, que foi fundamental por exemplo para expressar ideias sobre os processos mentais mais claramente do que conceitos verbais poderiam expressar, utilizando-se, para tanto, da linguagem da computação ou teoria da informação para descrevê-los. Além disso, apresentamos os experimentos do neurocientista Benjamin Libet e seus colaboradores, que tiveram forte repercussão nas ciências cognitivas de modo geral, mas principalmente no campo de estudo da consciência, dos atos volitivos e livre arbítrio.

No terceiro e último capítulo, inicialmente destacamos, em *Plans and the Structure of Behavior* (1960), de Miller, Galanter e Pribram, como os autores apotam a existência de uma lacuna explicativa entre cognição e comportamento e remetem a discussão sobre como preencher essa lacuna ao tópico da vontade como anteriormente pertencente ao debate na filosofia, falam também que “hoje a vontade parece ter desaparecido da teoria psicológica, assimilada anonimamente no tópico mais amplo da motivação” (Miller et al., 1960, p. 11). Passamos então a análise da concepção apresentada por Ulric Neisser, no *Cognition and Reality* (1967), de como a cognição pode ser um ponto chave de desenvolvimento de liberdade, proporcionando maiores possibilidade de ação no mundo. Analisamos ainda em dicionários e enciclopédias de psicologia e ciências cognitivas como atualmente é definido o conceito de vontade e livre arbítrio, apontando como em apenas dois deles foram encontradas entradas para esses conceitos.

Prosseguimos, por fim, ao apresentar de forma mais ampla o campo de estudo da motivação humana que, como vimos, tem se mostrado disperso e multifacetado, com uma tendência a explicar o comportamento humano e suas causas através de micro-teorias, algumas apresentando uma visão mais reativa do ser humano quanto ao ambiente, outras com uma visão mais ativa e intencionalmente direcionada, apresentando conduta voluntária destinada a atingir metas cognitivamente selecionadas. Abordamos também o conceito neuropsicológico da função executiva, que se refere a um sistema gerenciador com localização razoavelmente bem definida nos lobos frontais direito e esquerdo e que tem como função organizar sequência de ações a fim de atingir um objetivo definido, e que tem sido identificado com a realização de ações voluntárias.

Por fim, devemos destacar alguns limites de nosso estudo referentes principalmente a complexidade e amplitude do tema. De certa forma, em função disso, tivemos de fazer uma escolha por apenas apresentar uma visão mais geral e ampla, e muitas vezes pouco aprofundada de alguns pontos que talvez pudessem ter sido melhor explorados. Dentro do possível,

consideramos termos vislumbrado o quão longe de qualquer abordagem satisfatória ainda está a questão fundamental da vontade. Esta é, no meu entender, uma questão central para a psicologia, que deveria se colocar na vanguarda da investigação da vontade por se tratar de um tema que define a questão da natureza humana. Esperamos ter mostrado um pouco dessa importância, e ajudado a fomentar futuros estudos e pesquisas dentro dessa temática.

REFERÊNCIAS

- Abbagnano, N. (2007). *Dicionário de filosofia*. 5ed. São Paulo: Martins Fontes.
- Adams, J.S. (1965). Inequity in social exchanges. In: L. Berkowitz (ed.) *Advances in experimental social psychology*. New York: Academic Press.
- Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, 16 (1), 17-42.
- Aquino, T. (2001). *Suma de Teología*. 4.ed. Madri: Biblioteca de Autores Cristianos.
- Aristóteles. (2006). *De Anima. Apresentação*, tradução e notas: Maria Cecília Gomes dos Reis. São Paulo: Editora 34.
- Atkinson, J. W. (1957). Motivational determinants of risk taking behavior. *Psychological Review*, 64, 359-372.
- Atkinson, J. W. (1964). *An introduction to motivation*. New York: Van Nostrand.
- Atkinson, J. W. & Birch, D. (1978). *Introduction to motivation*. New York: Van Nostrand.
- Baars, B. J. (1986). *The cognitive revolution in psychology*. New York: Guilford.
- Baddeley, A. (1986). *Working memory*. Oxford, England; New York: Oxford University Press.
- Baker, L. R. (1995). *Metaphysics and mental Causation*, in Heil and Mele 1995, 75–96.
- Bandura, A. (1977). *Social learning theory*. New Jersey: Prentice-Hall.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122-147.
- Barberá, E. (2000). Marco conceptual e investigación de la motivación humana. *Revista Española de Motivación y Emoción*, 1, 23-36.
- Barberá, E. (2002). Modelos Explicativos en Psicología de la Motivación. Ester. *Revista Electrónica de Motivación y Emoción (REME)*, 5(10).
- Barberá, E. & Mateos, P. (2000). Investigación sobre psicología de la motivación en las universidades españolas, *Revista Española de Motivación y Emoción*, 5-6(3).
- Barker-Plummer, D. (2016) Turing Machines. The Stanford Encyclopedia of Philosophy (Spring 2016 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2016/entries/turing-machine/>>. Acessado em 10/02/2016.
- Batthyany, A. (2009). *Mental Causation and Free Will after Libet and Soon: Reclaiming Conscious Agency*. In Batthyany & Avshalom Elitzur. *Irreducibly Conscious: Selected Papers on Consciousness*, Universitätsverlag Winter Heidelberg 2009, p.135ff.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74, 1252–1265.

- Bealer, G. (1997). Self-Consciousness. *Philosophical Review*, 106: 69–117.
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press.
- Berlyne, D.E. (1960). *Conflict, arousal, and curiosity*. New York: McGrawHill.
- Berthoz, A. (2000). *The brain's sense of movement* (G. Weiss, Trans.). Cambridge, MA: Harvard University Press.
- Bickle, J. (2012). *A brief history of neuroscience's actual influences on mind-brain reductionism*. In Gozzano and Hill 2012, 43–65.
- Bisiach, E. (1988). *The (haunted) brain and consciousness*. In A. Marcel and E. Bisiach, eds. *Consciousness in Contemporary Science*. Oxford University Press.
- Blackmore, S. (1998) *Why psi tells us nothing about Consciousness*. In Hameroff, S. R.; Kaszniak, A. W. & Scott, C. (Eds). *Toward a Science of Consciousness II*. MIT Press. 701-707.
- Blanshard, B. (1964). *O acaso a favor do determinismo*. In *Determinismo e Liberdade na Era da Ciência Moderna*. Rio de Janeiro: Editora Fundo de Cultura.
- Block, N. (1980a). *Readings in the Philosophy of Psychology*, Volumes 1 and 2, Cambridge, MA: Harvard University Press.
- Block, N. (1980b). Troubles With Functionalism. In Block 1980a, 268–305.
- Block, N. (1980c). Are Absent Qualia Impossible? *Philosophical Review*, 89: 257–274.
- Block, N. (1981). Psychologism and Behaviorism. *Philosophical Review*, 90: 5–43.
- Block, N. (1986). Advertisement for a Semantics for Psychology. *Midwest Studies in Philosophy*, 10: 615–678.
- Block, N. (1990). Inverted Earth. *Philosophical Perspectives*, 4, Atascadero, CA: Ridgeview Press, 52–79.
- Block, N. (1997). Anti-Reductionism Slaps Back. *Philosophical Perspectives*, 11, Atascadero, CA: Ridgeview Press, 107–132.
- Block, N. & J. Fodor (1972). What Psychological States Are Not. *Philosophical Review*, 81: 159–181.
- Boden, M. A. (1988). *Computer Models of Mind*. Cambridge: Cambridge University Press.
- Boden, M. A. (2006). *Mind as machine: A history of cognitive Science*. Oxford: Oxford University Press.
- Boolos, G. S., Jeffrey, R. C. (1980). *Computability and Logic*. 2nd edition. Cambridge: Cambridge University Press.
- Brehm, J. W. (1966). *A theory of psychological reactance*. New York: Academic Press.
- Broad, C. D. (1952). *Determinism, Indeterminism, and Libertarianism in Ethics and the History of Philosophy*. London: Routledge & Kegan Paul.
- Burge, T. (1995). *Mind-Body Causation and Explanatory Practice*, in Heil and Mele 1995, 97–120.

- Bzuneck, J. A. (2004). A motivação do aluno: aspectos introdutórios. In: E. Boruchovitch e J. A. Bzuneck (Orgs.) *A motivação do aluno*, 3ª. Edição, pp. 9-36. Petrópolis: Vozes.
- Capovilla, A. G. S., Assef, E. C. S., & Cozza, H. F. P. (2007). Avaliação neuropsicológica das funções executivas e relação com desatenção e hiperatividade. *Aval. Psicol.*, 6(6), 51-60.
- Case, R. (1985). *Intellectual development: Birth to adulthood*. Orlando, FL: Academic Press.
- Castañon, G. A. (2010). O cognitivismo e o problema da cientificidade da psicologia. *Psicologia: teoria e prática*, 12(2), 223-253.
- Castañon, G. A.; Justi, F. R. R.; Araujo, S. F. (2014) A Psicologia Cognitiva. In: S. F. Araujo; F. Caropreso; G. Castañon; & R. T. Simanke (Eds.) *Fundamentos filosóficos da psicologia contemporânea* (pp.227-276) Juiz de Fora: Ed. UFJF.
- Castañon, G. A. (2015). O que é Construtivismo?. *Cadernos de História e Filosofia da Ciência (UNICAMP)*, v. 1, p. 160-180.
- Chalmers, D. (2011). A Computational Foundation for the Study of Cognition. *The Journal of Cognitive Science*, 12: 323–357.
- Chisholm, R. (1957). *Perceiving*. Ithaca: Cornell University Press.
- Church, A. (1936a). An Unsolvable Problem of Elementary Number Theory. *American Journal of Mathematics*, 58, 345-363.
- Church, A. (1936b). A Note on the Entscheidungsproblem. *Journal of Symbolic Logic*, 1, 40-41.
- Church, A. (1937). Review of Turing 1936. *Journal of Symbolic Logic*, 2, 42-43.
- Churchland, P. M. (1981). Eliminative Materialism and Propositional Attitudes. *Journal of Philosophy*, 78: 67–90.
- Churchland, P. S. (1981). On the Alleged Backwards Referral of Experiences and its Relevance to the Mind-Body Problem. *Philosophy of Science* 48 (2): 165–181.
- Churchland, P. M.; Churchland, P. S. (1983). Stalking the Wild Epistemic Engine. *Nous*, 17, 5-18.
- Churchland, P. M., Churchland, P. S. (1990). Could a Machine Think? *Scientific American*, 262 (Jan.), 32-37.
- Cofer, C. N. (1972). *Motivation and Emotion*. Glenview, Illinois: Scott, Foresman and Company.
- Coleman, S. R. & Mehlman, S. E. (1992). An empirical update (1969–1989) of D.L. Krantz's thesis that the experimental analysis of behavior is isolated. *Behavior Analyst*, 15, 43–49.
- Copeland, B. J. (2000a). Narrow Versus Wide Mechanism. *Journal of Philosophy*, 97, 5-32.
- Copeland, B. J. (2000b). The Turing Test. *Minds and Machines*, 10: 519–39.
- Copeland, B. J. & Proudfoot, D. (1999a). Alan Turing's Forgotten Ideas in Computer Science. *Scientific American*, 280 (April), 76-81.
- Copeland, B. J. & Proudfoot, D. (1999b). The Legacy of Alan Turing. *Mind*, 108, 187-195.

- Copeland, B. J. & Proudfoot, D. (2000). What Turing Did After He Invented the Universal Turing Machine. *Journal of Logic, Language, and Information*, 9, 491-509.
- Costelloe, T. (2014). Giambattista Vico. *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2014/entries/vico/>>.
- Cover, T. & J. Thomas (2006). *Elements of Information Theory*. Hoboken: Wiley.
- Damásio, A. R. (1996). *O erro de Descartes: emoção, razão e o cérebro humano*. São Paulo: Companhia das Letras, 2ª Ed.
- Davidson, D. (1980). *Essays on Actions and Events*. New York: Oxford University Press.
- Dennett, D. C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. Brighton: Harvester.
- Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little Brown.
- Dennett, D. C. (2003). *Freedom Envolves*. New York: Viking.
- Descartes, R. (1997). *Princípios da Filosofia*. Lisboa, Edições 70.
- Descartes, R. (2001). *Discurso do Método*. 2ª ed. São Paulo: Martins Fontes.
- D'Esposito, M. & Gazzaley, A. (2006). *Neurorehabilitation and executive function*. In M. E. Selzer, I. Cohen, F. H. Cage, S. Clarke, & P. W. Duncan (Eds). *Neural rehabilitation and repair*. Cambridge: Cambridge University Press, pp. 475-487.
- Deutsch, D. (1985). Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. *Proceedings of the Royal Society, Series A*, 400, 97-117.
- Dewsbury, D. A. (2000). Comparative cognition in the 1930s. *Psychonomic Bulletin and Review*, 7, 267-283.
- Dietrich, E. (1989). Semantics and the Computational Paradigm in Cognitive Psychology. *Synthese*, 79: 119-141.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. Oxford: Blackwell.
- Driver-Linn, E. (2003). Where is psychology going? Structural fault lines revealed by psychologists' use of Kuhn. *American Psychologist*, 58, 269-278.
- Eccles, J. C. (1985). Mental summation: The timing of voluntary intentions by cortical activity. *Behavioral and Brain Sciences* 8: 542-543.
- Eccles, J. C. (2000) *Cérebro e consciência: o self e o cérebro*. Lisboa: Instituto Piaget.
- Eysenck, M. W. (ed.) (1990). *The Blackwell Dictionary of cognitive psychology*. Cambridge: Blackwell Publishers.
- Eysenck, M. W., & Keane, M. T. (1994). *Psicologia cognitiva: Um manual introdutório*. Porto Alegre: Artes Médicas.
- Ferrater-Mora, J. (2000) *Dicionário de filosofia*. 2.ed. São Paulo: Edições Loyola.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

- Fischer, J. M. (1994). *The Metaphysics of Free Will*. Oxford: Blackwell.
- Fischer, J. M. & Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906–911.
- Frankfurt, H. G. (1969). *Alternate Possibilities and Moral Responsibility*. *The Journal of Philosophy*, 66(23) 829-839.
- Frankfurt, H. G. (1971). *Freedom of the Will and the Concept of a Person*. *The Journal of Philosophy*, 68(1) 5-20.
- Freud, S. (1900) *A interpretação dos sonhos*. In: Edição Standard das Obras psicológicas completas de Sigmund Freud (vol. IV e V). Rio de Janeiro: Imago, 1996.
- Freud, S. (1901) *Sobre a Psicopatologia da Vida Cotidiana*. In: Edição Standard das Obras psicológicas completas de Sigmund Freud (vol. VI). Rio de Janeiro: Imago, 1996.
- Freud, S. (1905) *Os Chistes e sua Relação com o Inconsciente*. In: Edição Standard das Obras psicológicas completas de Sigmund Freud (vol. VIII). Rio de Janeiro: Imago, 1996.
- Freud, S. (1900) *Esboço de Psicanálise*. In: Edição Standard das Obras psicológicas completas de Sigmund Freud (vol. XXIII). Rio de Janeiro: Imago, 1996.
- Fried, I.; Mukamel, R; Kreiman, G. (2011). Internally Generated Preactivation of Single Neurons In Human Medial Frontal Cortex Predicts Volition. *Neuron* 69(3): 548-562.
- Fodor, J. A. (1975). *The Language of Thought*. New York: Crowell.
- Fodor, J. A. (1968). *Psychological Explanation*. New York: Random House.
- Fodor, J. A. (1981). The Mind-Body Problem. *Scientific American*, 244 (Jan.), 124-32.
- Fodor, J. A. (1990). Making Mind Matter More. In J. Fodor, *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press, 137–159.
- Fodor, J. A. (1994). *The Elm and the Expert*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1998). *Concepts*. Oxford: Clarendon Press.
- Fuster, J. M. (2002). Frontal lobe and cognitive development. *J Neurocytol*, 31(3-5), 373-385.
- Fuster, J. M. (2003). *Cortex and mind: unifying cognition*. New York: Oxford University Press.
- Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., Plant, E. A., Tice, D. M., Brewer, L. E., & Schmeichel, B. J. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, 92, 325–336.
- Gandy, R. (1980). Church's Thesis and Principles for Mechanisms. In Barwise, J., Keisler, H.J., Kunen, K. (eds) 1980. *The Kleene Symposium*. Amsterdam: North-Holland.
- Gandy, R. (1988). The Confluence of Ideas in 1936. In Herken, R. (ed.) 1988. *The Universal Turing Machine: A Half-Century Survey*. Oxford: Oxford University Press.
- Gardner, H. (1996). *A nova ciência da mente*. São Paulo: EDUSP.

- Gazzaniga, M. S, Ivry, R. B. & Mangum, G. R. (2006). *Neurociência cognitiva: a biologia da mente*. Porto Alegre: Artmed.
- Geach, P. (1957). *Mental Acts*. London: RKP.
- Geroch, R. & Hartle, J. B. (1986). Computability and Physical Theories. *Foundations of Physics*, 16, 533-550.
- Ginet, C. (1966). *Might We Have No Choice*. In Keith Lehrer, ed., *Freedom and Determinism*. (205-224). New York: Random House.
- Godefroy, O. & Rousseaux, M. (1997). Novel decision making in patients with prefrontal or posterior brain damage. *Neurology* 49(3): 695-701.
- Goldberg, E. (2002). *O cérebro executivo: lobos frontais e a mente civilizada*. Rio de Janeiro: Imago.
- Gomes, G. (2005). O problema mente-cérebro em Freud. *Psicologia: Teoria e Pesquisa*. 21(2) 149-155.
- Green, C. (1996). Where did the word “cognitive” come from anyway? *Canadian Psychology*, 37(1), 31-39.
- Gregory, R.L. (1987). *The Oxford Companion to the Mind*. Oxford: Oxford University Press.
- Grice, P. (1989). *Studies in the Ways of Words*. Cambridge: Harvard University Press.
- Guthrie, E. R. (1935). *The Psychology of Learning*. New York: Harper.
- Guttenplan, S. (1994). *A Companion to the Philosophy of Mind*. Oxford: Blackwell.
- Haggard, P. (2011). Decision Time for Free Will. *Neuron* 69(3): 404-406.
- Haynes, J.-D.; Sakai, K.; Rees, G.; Gilbert, S.; Frith, C & Passingham, R. E. (2007). Reading Hidden Intentions in The Human Brain. *Current Biology* 17(4): 323-328.
- Hebb, D. O. (1955). Drives and the CNS conceptual nervous system. *Psychological Review*, 62, 245-254.
- Heckhausen, H. (1977). Achievement motivation and its constructs: A cognitive model. *Motivation and Emotion*, 1, 283-329.
- Hobbes, T. (1973) *Leviatã ou matéria, forma e poder de um estado eclesiástico e civil*. Tradução de João Paulo Monteiro e Maria Beatriz Nizza da Silva. São Paulo: Abril Cultural.
- Hobbs, S. & Burman, J. T. (2009). Is the 'cognitive revolution' a myth? *The Psychologist*. 22(9), 812-814.
- Hodgson, D. (2007). Why I (still) believe in free will and responsibility, edited version published under the title ‘Partly free’ in *The Times Literary Supplement* on 6 July, 2007. Disponível em http://users.tpg.com.au/raeda/website/why.htm#_ftn12
- Honderich, T. (ed.). *The Determinism and Freedom*. Philosophy Website. Disponível em URL = <<http://www.ucl.ac.uk/~uctytho/dfwIntroIndex.htm/>> acessado em 13/01/2015.
- Horgan, T. & Woodward, J. (1985). Folk Psychology is Here to Stay. *Philosophical Review*, 94: 197–226.

- Houdé, O. (ed.) (2004). *Dictionary of cognitive science: Neuroscience, Psychology, Artificial Intelligence, Linguistics, and Philosophy*. New York: Psychology Press.
- Hume, D. (2003). *Investigações sobre o Entendimento Humano e sobre os Princípios da Moral*. São Paulo: UNESP São Paulo.
- Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly*, 32: 127–136.
- Jackson, F. (1996). Mental Causation. *Mind*, 105(419): 377–413.
- Jackson, F.; Pargetter, R.; Prior, E. (1982). Functionalism and Type-Type Identity Theories. *Philosophical Studies*, 42: 209–225.
- James, W. (1880). The Feeling of Effort. *Mind*, 5(20): 582.
- Kane, R. (1998). *The Significance of Free Will*. Oxford: Oxford University Press.
- Kane, R. (2005). *Introduction: The Contours of Contemporary Free Will Debate*. In: Robert Kane, ed., *The Oxford Handbook of Free Will*. (1-24). Oxford: Oxford University Press.
- Kane, R. (2014). Acting of One's Own Free Will: Modern Reflections on an Ancient Philosophical Problem. *Proceedings of the Aristotelian Society*, 114(1) 35-55.
- Kant, I. (1987). *Prolegómenos a toda metafísica futura - que queira apresentar-se como ciência*. Lisboa: Edições 70.
- Kant, I. (2001a). *Crítica da Razão Pura*. 5.ed. Lisboa: Fundação Calouste Gulbenkian.
- Kant, I. (2001b). *Fundamentação da metafísica dos costumes*. Lisboa: Edições 70.
- Kiesel, A., Wagener, A., Kunde, W., Hoffmann, J., Fallgatter, J. A. & Stöcker, C. (2006). *Unconscious manipulation of free choice in humans*. *Consciousness and Cognition*. 15(2006) 397-408.
- Kim, J. (1989). Mechanism, Purpose, and Explanatory Exclusion. In J. Kim. *Supervenience and Mind*. Cambridge, Cambridge University Press.
- Kim, J. (1998). *Mind in a Physical World*. Cambridge, MA: Bradford.
- Kim, J. (2007). *Causation and Mental Causation*. In McLaughlin and Cohen 2007, 227–242.
- Kleene, S.C. (1967). *Mathematical Logic*. New York: Wiley.
- Kleene, S.C. (1952). *Introduction to Metamathematics*. Amsterdam: North-Holland.
- Krench, D. & Crutchfield, R. S. (1959). *Elements of psychology*. New York: Alfred A. Knopf.
- Kuhl, J. (1985). Volitional mediators of cognition behavior consistency: Selfregulatory processes and action versus state orientation. In: J. Kuhl & J. Beckman (Eds.), *Action control: From cognition to behavior*. Nueva York: springer-Verlag.
- Kuhl, J. (1986). Motivation and informatio processing: A new look at decisionmaking, dynamic change, and action control. In: R.M. Sorrentino & E.T. Higgins (Eds.), *Handbook of motivation and cognition. Foundations of social behavior*. New York: Wiley.
- Lachman, R.; Lachman, J.; & Butterfield, E. (1979). *Cognitive psychology and information processing: An introduction*. Hillsdale, NJ: Lawrence Erlbaum.

- Leahey, T. (1992). The mythical revolutions of American psychology. *American Psychologist*, 47(2), 308-318.
- Levin, Janet. (2013) Functionalism. *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2013/entries/functionalism/>>.
- Leite, A. & Simon, S. (2010). Werner Heisenberg e a Interpretação de Copenhague: a filosofia platônica e a consolidação da teoria quântica. *Sci. Stud.* 8(2) 213-241.
- Levy, N. (2013). *Addiction and self-control: Perspectives from Philosophy, Psychology, and Neuroscience*. 1.ed. Oxford: Oxford University Press.
- Lewin, K. (1935) *A dynamic theory of personality*. New York: McGraw-Hill.
- Lewis, D. (1963). *Scientific Principles of psychology*. Englewood Cliffs: Prentice-Hall.
- Lewis, D. (1966). An Argument for the Identity Theory. *Journal of Philosophy*, 63: 17–25.
- Lewis, D. (1972). Psychophysical and Theoretical Identifications, in Block 1980, 207–215.
- Lewis, D. (1994) Reduction of Mind. In S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Oxford: Blackwell.
- Libet, B. (1981). The Experimental Evidence for Subjective Referral of a Sensory Experience Backwards in Time: Reply to P.S. Churchland. *Philosophy of Science* 48: 182–197.
- Libet, B. (1985). *Unconscious cerebral initiative and the role of conscious will in voluntary action*. *The Behavioral and Brain Sciences* .8(4) 529-566.
- Libet, B. (2004). *Mind Time - The Temporal Factor in Consciousness*. Harvard: Harvard University Press.
- Libet, B. (2006). Reflections on the Interaction of the Mind and Brain. *Progress in Neurobiology* 78: 322–326.
- Libet, B.; Wright, J., Elwood, W.; Feinstein, B.; Pearl, D. K. (1979). Subjective Referral of the Timing for a Conscious Sensory Experience - A Functional Role for the Somatosensory Specific Projection System in Man. *Brain* 102: 193–224.
- Libet B.; Wright, E. W., Gleason (1983). Readiness potentials preceding unrestricted spontaneous pre-planned voluntary acts. *Electroencephalographic and Clinical Neurophysiology* 54: 322–325.
- Loar, B. (1981). *Mind and Meaning*. Cambridge: Cambridge University Press.
- Loewer, B. (2002). Comments on Jaegwon Kim's Mind and the Physical World. *Philosophy and Phenomenological Research*, 65(3): 555–662.
- Loewer, B. (2007). *Mental Causation, or Something Near Enough*. In McLaughlin and Cohen 2007, 243–264.
- Lopes, E. J.; Lopes, R.; & Teixeira, J. F. (2004). A psicologia cognitiva experimental cinquenta anos depois: a crise do paradigma do processamento de informação. *Paidéia* (Ribeirão Preto), 14(27), 17-26.
- Lovie, A. D. (1983). Attention and behaviourism – Fact and fiction. *British Journal of Psychology*, 74, 301–310.

- Malcolm, N. (1968). The Conceivability of Mechanism. *Philosophical Review*, 77: 45–72.
- Mandler, G. (2002). Origins of the cognitive revolution. *Journal of the History of the Behavioral Sciences*, 38(4), 339–353.
- Marcellos, C. F. (2012). *Objeto e método na psicologia de Wilhelm Wundt e Edward Titchener: um estudo comparativo*. (Dissertação de Mestrado). UFJF, BR.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.
- Maslow A (1968). *Introdução à psicologia do ser*. Rio de Janeiro: Eldorado.
- McCauley, R. (2012). About face: philosophical naturalism, the heuristic identity theory, and recent findings about prosopagnosia. In Gozzano and Hill 2012, 186–206.
- McClelland, J. L. (1979). On the time relations of mental processes. An examination of processes in cascade. *Psychological Review*, 86, 287-330.
- McClusky, C. (2007). *Medieval Theories of Free Will*. Internet Encyclopedia of Philosophy, 2007. Disponível em URL = <http://www.iep.utm.edu/freewi-m/>> acessado em 13/01/2015.
- McCulloch, W. & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
- McFadden, J. (2002). The Conscious Electromagnetic Information (Cemi) Field Theory: The Hard Problem Made Easy? *Journal of Consciousness Studies* 9 (8): 45–60.
- McKenna, M. (2009). *Compatibilism*. The Stanford Encyclopedia of Philosophy. Disponível em URL = <<http://plato.stanford.edu/archives/win2009/entries/compatibilism/>> acessado em 13/01/2015.
- McLaughlin, B. & Cohen, J. (2007). *Contemporary Debates in the Philosophy of Mind*. Malden, MA: Blackwell Publishing.
- Mille, D. E. (2014). Freedom and Resentment and Consequentialism: Why ‘Strawson’s Point’ Is Not Strawson’s Point. *Journal of Ethics & Social Philosophy*. 8(2) 1-22.
- Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Miller, G. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, 7(3), 141-144.
- Miller, G; Galanter, E.; & Pribram, K. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston.
- Miller, P. H. (2012). *Theories of developmental psychology*. New York: Worth Publishers.
- Melnyk, A. (2003). *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge: Cambridge University Press.
- Miller, J. (1988). Discrete and continuous models of human information processing: Theoretical distinctions and empirical results. *Acta Psychologica*. 67, 191-257.
- Mole, Christopher (2013) Attention. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2013/entries/attention/>>.

- Mook, D. G. (1987). *Motivation: the organization of action*. New York: W. W. Norton & Company.
- Mourão Junior, C. A., & Melo, L. B. R. (2011). Integração de três conceitos: função executiva, memória de trabalho e aprendizado. *Psicologia: Teoria e Pesquisa*, 27(3), 309-314.
- Nadel, L. (ed.) (2005). *Encyclopedia of cognitive science*. University of Arizona: Wiley.
- Nagel, T. (1974). What Is It Like To Be a Bat? *Philosophical Review*, 83: 435–450.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Neisser, U. (1976). *Cognition and Reality*. San Francisco: Freeman And Company.
- Newell, A. (1980). Physical Symbol Systems. *Cognitive Science*, 4, 135-183.
- Newell, A.; Shaw, J.C. & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, 65, 151-166.
- Norman, D.A. & Bobrow, D.G. (1975). On datalimited and resource-limited processes. *Cognitive Psychology*, 7, 44-64.
- Nuttin, J. (1985). *Future time perspective and motivation*. New Jersey: LEA.
- O'Connor, T. (2000). *Causality, Mind and Free Will*. In: Philosophical Perspectives 14: Action and Freedom. (105-117). Cambridge, MA: Blackwell.
- O'Connor, T. (2014). *Free Will*. In Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2014 Edition)*. Disponível em URL = <<http://plato.stanford.edu/archives/fall2014/entries/freewill/>> acessado em 13/01/2015.
- O'Connor, T. & Wong, H. Y. (2005). *The metaphysics of emergence*. *Noûs* 39(4) 658–678.
- O'Donohue, W. (1993). The Spell of Kuhn on psychology: Na exegetical elixir. *Philosophical Psychology*, 6, 267-287.
- Olds, J. & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*. 47(6): 419-27.
- Palmer, D. (2014). *Free Will, Libertarianism and Kane*. In: Palmer, D. *Libertarian Free Will: Contemporary Debates*. Oxford: Oxford University Press.
- Palmer, S.E. & Kimchi, R. (1986). *The information processing approach to cognition*. In T. Knapp & L.C. Robertson (Orgs.), *Approaches to cognition: Contrasts and controversies*. Hillsdale: Lawrence Erlbaum.
- Penna, A. G. (1984). *Introdução à psicologia cognitiva*. São Paulo: EPU.
- Penna, A. G. (2001). *Introdução à motivação e emoção*. Rio de Janeiro: Imago.
- Pereboom, D. (2001). *Living Without Free Will*. Cambridge: Cambridge University Press.
- Piaget, J. (1967). *Logique et connaissance scientifique*. Paris: Gallimard.
- Piaget, J. (1973). *Psicologia e epistemologia: por uma teoria do conhecimento*. Rio de Janeiro: Forense.

- Piaget, J. (1987). *Discussão*. In: M. Piatelli-Palmarini (Ed.), *Teorias da linguagem, teorias da aprendizagem* (pp. 219-230). Lisboa: Edições 70.
- Piccinini, G. (2004). Functionalism, computationalism, and mental states. *Stud. Hist. Phil. Sci.*, 35: 811–833.
- Place, U. T. (1956). Is Consciousness a Brain Process? *British Journal of Psychology*, 47: 44–50.
- Platão (2002). *República*. Rio de Janeiro: Editora Best-Seller.
- Post, E.L. (1936). Finite Combinatory Processes - Formulation 1. *Journal of Symbolic Logic*, 1, 103-105.
- Poland, J. & Graham, G. (2011). *Addiction and Responsibility*. Cambridge, MA, United States & London, United Kingdom: The MIT Press.
- Polger, T. (2011). Are Sensations Still Brain Processes? *Philosophical Psychology*, 24(1): 1–21.
- Posner, M.I. (1992). Attention as a cognitive and neural system. *Current Directions in Psychological Science*, 1, 11-14.
- Posner, M.I. & Dehaene, S. (1994). Attentional networks. *Trends in neurosciences*, 17, 75-79.
- Posner, M.I. & Petersen, S.E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25-42.
- Putnam, H. (1975). *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
- Putnam, H. (1988). *Representation and Reality*. Cambridge, MA: MIT Press.
- Quine, W.V.O. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Raynor, J. O. (1981). Future orientation and achievement motivation: Toward a theory of personality functioning and change. In: G. Ydewalle & W. Lens (Eds.) *Cognition in human motivation and learning*. New Jersey: Lawrence Erlbaum.
- Reeve, J. (2006). *Motivação e emoção*. 4.ed. Rio de Janeiro: LTC.
- Rey, G. (1997). *Contemporary Philosophy of Mind*. Cambridge, MA: Blackwell.
- Rupert, R. (2006). Functionalism, Mental Causation, and the Problem of Metaphysically Necessitated Effects. *Noûs*, 40: 256–283.
- Ryle, G. (1949). *The Concept of Mind*. London: Hutheson.
- Samelson, F. (1986). History, origin myth, and ideology: The ‘discovery’ of social psychology by Auguste Comte. In K.S. Larsen (Ed.) *Dialectics and ideology in psychology* (pp.14–29). Norwood, NJ: Ablex.
- Schifffrin, R.M., & Schneider, W. (1977). Controlled and automatic human processing: II. Perceptual learning, automatic attention and a general theory. *Psychological Review*, 84, 127–190.
- Schopenhauer, A. (2005). *O Mundo Como Vontade e Representação*. São Paulo: Editora UNESP.
- Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press.

- Searle, J. (1990) Is the Brain a Digital Computer? *Proceedings and Addresses of the American Philosophical Association*, 64: 21–37.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, England; New York: Cambridge University Press.
- Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27: 379–423, 623–656.
- Shapiro, L. & Polger, T. (2012). Identity, variability, and multiple realization in the special sciences. In Gozzano and Hill 2002, 264–287.
- Shima, K. & Tanji, J. (2000) Neuronal Activity in the Supplementary and Presupplementary Motor Areas for Temporal Organization of Multiple Movements. *Journal of Neurophysiology* 84:2148 –2160.
- Shoemaker, S. (1984). *Identity, Cause, and Mind*. Cambridge: Cambridge University Press.
- Silva Neto, S. A. (2001) As Críticas de Vico e Horkheimer ao Racionalismo: a negação da razão naturalista em nome da filosofia humanista. *Educação e Filosofia*, Uberlândia, 15 (29), 175-198.
- Singer, W. (1990). Search for coherence: A basic principle of cortical selforganization. *Concepts in Neurosciences*, 1, 1–26.
- Sipser, M. (2007) *Introdução à Teoria da Computação*. Editora Thomson Pioneira.
- Smart, J. J. C. (1959). Sensations and Brain Processes. *Philosophical Review*, 68: 141–156.
- Smart, J. J. C. (2014). The Mind/Brain Identity Theory. *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2014/entries/mind-identity/>>.
- Smith, K. (2011). Neuroscience vs philosophy: Taking aim at free will. *Nature* 477 (7362): 23–5.
- Smolensky, P. (1988). On the Proper Treatment of Connectionism. *Behavioral and Brain Sciences*, 11, 1-23.
- Skinner, F. B. (1973). *O Mito da Liberdade*. 2.ed. Rio de Janeiro: Bloch Editores.
- Soon, C. S.; Brass, M.; Heinze, H.-J.; Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience* 11 (5): 543–5.
- Sousa, C. E. B. (2014). Neurofilosofia da racionalidade: críticas e propostas a partir da filosofia e das neurociências. *Kínesis*, v.6, n.12, p. 156-183.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153, 652-654.
- Sternberg, S. (1969). The discovery of processing stages: extensions of Donders' method. *Acta Psychologica*, 30, 276-315.
- Sternberg, S. (1975). Memory scanning: new findings and current controversies. *Quarterly Journal of Experimental Psychology*, 27, 1-32.
- Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica*, 106, 147-246.
- Stewart, I. (1991). Deciding the Undecidable. *Nature*, 352, 664-5.

- Stich, S. (1981). Dennett on Intentional Systems. *Philosophical Topics*, 12: 39–62.
- Stich, S. (1983). *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.
- Strawson, G. (1994). *Mental Reality*. Cambridge, MA: MIT Press.
- Strawson, G. (1994). *The Impossibility of Moral Responsibility*. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*. 75 (1) 5-24.
- Strawson, P. (1982). Freedom and Resentment. In: Gary Watson ed. *Free Will*. Oxford: Oxford University Press. (59-80).
- Timpe, K. (2006). *Free Will*. Internet Encyclopedia of Philosophy. URL = <http://www.iep.utm.edu/freewill/> acessado em 13/01/2015.
- Todorov, J. C., & Moreira, M. B. (2005). O conceito de motivação na psicologia. *Revista Brasileira de Terapia Comportamental e Cognitiva*, 7(1), 119-132.
- Tolman, E. C. (1948) Cognitive maps in rats and men. *Psychological Review*, 55(4), 189-208.
- Turing, A. (1937a). On computable numbers, with a application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, series 2, 42 (1936-37), 230-65.
- Turing, A. (1937b). Computability and λ -Definability. *The Journal of Symbolic Logic*, 2: 153–163.
- Turing, A. (1946). Proposal for Development in the Mathematics Division of an Automatic Computing Engine (ACE). In: Carpenter, B.E., Doran, R.W. (eds) 1986. *A.M. Turing's ACE Report of 1946 and Other Papers*. Cambridge, Mass.: MIT Press.
- Turing, A. (1948). *Intelligent Machinery*. *National Physical Laboratory Report*. In: Meltzer, B., Michie, D. (eds) 1969. *Machine Intelligence*. Edinburgh: Edinburgh University Press. (Digital facsimile viewable at http://www.AlanTuring.net/intelligent_machinery.)
- Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, 59, 433-460.
- Turing, A. (1951a). *Can Digital Computers Think?* In Copeland, B.J. (ed.) 1999. ‘A Lecture and Two Radio Broadcasts on Machine Intelligence by Alan Turing’. In Furukawa, K., Michie, D., Muggleton, S. (eds) 1999. *Machine Intelligence 15*. Oxford: Oxford University Press.
- Turing, A. (1951b). *Intelligent Machinery, A Heretical Theory*. In Copeland, B.J. (ed.) 1999. ‘A Lecture and Two Radio Broadcasts on Machine Intelligence by Alan Turing’. In Furukawa, K., Michie, D., Muggleton, S. (eds) 1999. *Machine Intelligence 15*. Oxford: Oxford University Press.
- Velmans, M. (2000). *Understanding Consciousness*. London: Routledge.
- Van Inwagen, P. (1983). *An Essay on Free Will*. Oxford: Clarendon Press.
- Van Inwagen, P. (2008). How to Think about the Problem of Free Will. *J. Ethics*. 12, 327–341.
- Vihvelin, K. Arguments for Incompatibilism. In: Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition). Disponível em URL = <http://plato.stanford.edu/archives/spr2011/entries/incompatibilism-arguments/> acessado em 13/01/2015.

- Vila, J. & Fernández, C. (1990). Activación y conducta. In: S. Palafox & J. Vila (Eds.) *Motivación y emoción*. Madrid: Alhambra Universidad.
- Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology*, 94, 883–898.
- Vroom, V. H. (1964). *Work and motivation*. New York: Wiley and sons.
- Wegner, D. M.; Sparrow, B. & Winerman, L. (2004). *Vicarious Agency: Experiencing Control Over the Movements of Others*. *Journal of Personality and Social Psychology*. 86(6) 38–848.
- Wegner, D. M. & Wheatley, T. P. (1999). *Apparent mental causation: Sources of the experience of will*. *American Psychologist*. 5(7) 480-492.
- White, R.W. (1959). Motivation reconsidered: the concept of competence. *Psychological Review*, 66, 297-333.
- Wiener, N. (1948). *Cybernetics or control and communication in the animal and the machine*. New York: John Wiley & Sons.
- Wilson, R. A.; Keil, F. C. (eds.). (1999). *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, Massachusetts: MIT Press.
- Wittgenstein, L. (1953). *Philosophical Investigations*. New York: Macmillan.
- Wittgenstein, L. (1980). *Remarks on the Philosophy of Psychology*. Vol.1. Oxford: Blackwell.
- Yablo, S. (1992). Mental Causation. *Philosophical Review*, 101: 245–280.
- Young, P. T. (1961). *Motivation and emotion: a survey of the determinants of human and animal activity*. New York: Wiley.
- Zhu, J. (2003). Reclaiming volition: An Alternative interpretation of Libet's experiment. *Journal of Consciousness Studies*, Vol.10, Issue 11, 61-77.
- Zhu, J. (2004a). Is conscious will an illusion? (A critical notice of Daniel Wegner's *The Illusion of Conscious Will* (MIT Press 2002)). *Disputatio*, 16, 59-69.
- Zhu, J. (2004b). Intention and Volition. *Canadian Journal of Philosophy*. 34(2) 175-194.
- Zubiri, X. (1992). *Sobre el sentimiento y la volición*. 1.ed. Madrid: Alianza Editorial.