

UNIVERSIDADE DE SÃO PAULO (USP)
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS (FFLCH)
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

GUILHERME TEIXEIRA MARTINS SCHETTINI

**UMA CONCEPÇÃO DE LIVRE-ARBÍTRIO
PARA A CIÊNCIA DA MENTE E DO ENCÉFALO**

SÃO PAULO

2022

GUILHERME TEIXEIRA MARTINS SCHETTINI

UMA CONCEPÇÃO DE LIVRE-ARBÍTRIO PARA A CIÊNCIA DA MENTE E DO ENCÉFALO

Tese de Doutorado submetida ao Programa de Pós-Graduação em Filosofia do Departamento de Filosofia da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, como requisito para a obtenção do título de Doutor em Filosofia, sob a orientação do Prof. Dr. Osvaldo Frota Pessoa Junior.

Linha de Pesquisa: Lógica, Filosofia da Linguagem e Filosofia das Ciências.

SÃO PAULO

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

S327c Schettini, Guilherme
Uma concepção de livre-arbítrio para a ciência da mente e do encéfalo / Guilherme Schettini; orientador Osvaldo Pessoa Junior - São Paulo, 2022.
113 f.

Tese (Doutorado)- Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo.
Departamento de Filosofia. Área de concentração: Filosofia.

1. Livre-arbítrio. 2. Ciência da mente e do encéfalo. 3. Sensação de livre-arbítrio. I. Pessoa Junior, Osvaldo, orient. II. Título.

AGRADECIMENTOS

Oswaldo Frota Pessoa Junior.

Beatriz Sorrentino Marques.

Gilberto Lourenço Gomes.

Saulo de Freitas Araujo.

Programa de Pós-Graduação em Filosofia da Universidade de São Paulo.

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Meu primeiro ato de livre-arbítrio será acreditar no livre-arbítrio

William James

RESUMO

SCHETTINI, Guilherme. Uma Concepção de Livre-Arbítrio para a Ciência da Mente e do Encéfalo. Tese de Doutorado – Faculdade de Filosofia, Letras e Ciências Humanas. Departamento de Filosofia, Universidade de São Paulo, 2022.

Esta tese terá como propósitos: (i) apresentar as principais posições filosóficas sobre o tema do livre-arbítrio, nomeadamente o compatibilismo e o incompatibilismo; (ii) apresentar o conceito de responsabilidade moral, e avaliar como ele se relaciona ao conceito de livre-arbítrio; (iii) apresentar o problema mente-cérebro, e avaliar como ele se relaciona ao conceito de livre-arbítrio; (iv) apresentar a noção de escolha racional, e avaliar como ela se relaciona ao conceito de livre-arbítrio; (v) apresentar um breviário dos principais estudos da “neurociência do livre-arbítrio”; (vi) propor as bases para o desenvolvimento de uma ciência para a “sensação de livre-arbítrio”, que chamarei de “a ciência da mente e do encéfalo”.

Palavras-Chave: Livre-Arbítrio; Ciência da Mente e do Encéfalo; Sensação de Livre-Arbítrio.

ABSTRACT

SCHETTINI, Guilherme. A Conception of Free Will for the MindBrain Science. 2022. PhD – Faculdade de Filosofia, Letras e Ciências Humanas. Departamento de Filosofia, Universidade de São Paulo, São Paulo, 2022.

This thesis aims to: (i) present the main philosophical positions on the subject of free will, namely compatibilism and incompatibilism; (ii) present the concept of moral responsibility, and evaluate how it relates to the concept of free will; (iii) present the mind-brain problem, and assess how it relates to the concept of free will; (iv) present the notion of rational choice, and evaluate how it relates to the concept of free will; (v) present a breviary of the main studies of the “neuroscience of free will;” (vi) propose the bases for the development of a science for the “subjective experience of free will,” which I will call “the MindBrain science”.

Keywords: Free Will; MindBrain Science; Subjective Experience of Free Will.

SUMÁRIO

Considerações Iniciais	p.12
Propósitos da tese	p.12
Estrutura da tese	p.12
Método da tese	p.12
Originalidade da tese	p.13
Bloco I: Livre-Arbítrio: Definição, Sentidos e Problemas	p.14
Definição de “livre-arbítrio”, origem do termo e sua função primitiva	p.14
Sentidos filosóficos de “livre-arbítrio”	p.15
Problemas tradicionais relacionados ao livre-arbítrio	p.15
Um “novo” problema relacionado ao livre-arbítrio	p.16
Comentários finais (Bloco I)	p.17
Bloco II: Compatibilismo e Incompatibilismo	p.20
Compatibilismo	p.20
O compatibilismo de Agostinho	p.21
O compatibilismo de Thomas Hobbes	p.22
O compatibilismo de David Hume	p.22
Liberdade e livre-arbítrio	p.23
Exemplos de Frankfurt contra o princípio das possibilidades alternativas	p.24
O compatibilismo de Gilberto Gomes	p.25
Incompatibilismo	p.27
O incompatibilismo de Robert Kane	p.29
O incompatibilismo de Derk Pereboom	p.31
Comentários finais (Bloco II)	p.32

Bloco III: Responsabilidade Moral **p.37**

Responsabilidade moral	p.37
Fatos sociais	p.39
Diferenças entre “necessidade” e “suficiência”	p.41
Comentários finais (Bloco III)	p.42

Bloco IV: O Problema Mente-Cérebro **p.44**

O problema mente-cérebro	p.44
Intuições dualistas na escrita científica: a crítica de Mudrik e Maoz	p.46
O eterno retorno do materialismo: a crítica de Saulo Araújo	p.48
Um modelo quântico para o interacionismo cartesiano	p.49
John Eccles e os “campos de influência” da mente sobre o cérebro	p.51
A complementaridade mente-cérebro	p.52
Comentários finais (Bloco IV)	p.56

Bloco V: Escolha Racional **p.62**

Escolha racional	p.62
O modelo de escolha da Câmara dos Representantes: um diálogo com Pessoa Jr.	p.63
Comentários finais (Bloco V)	p.69

Bloco VI: A Neurociência do Livre-Arbítrio **p.72**

O experimento de Libet et al.	p.72
A concepção de Libet sobre o livre-arbítrio	p.74
Os experimentos de Isham et al.	p.75
O experimento de Filevich et al.	p.76
O experimento de Haggard e Eimer	p.76
Libet versus Haggard e Eimer	p.78

Beatriz Sorrentino Marques e os diferentes tipos de decisões	p.78
Experimentos de Trevena e Miller	p.79
Crítica de Gilberto Gomes a Trevena e Miller	p.80
Comentários finais (Bloco VI)	p.81
Bloco VII: Uma Ciência para a Sensação de Livre-Arbítrio	p.84
Filosofia da neurociência	p.84
A neurociência da consciência	p.86
Dois defensores do status científico da consciência	p.89
Pessoa Jr. e a consciência como observação do cérebro	p.89
Metaestabilidade no cérebro	p.91
O problema da percepção	p.92
Daniel Wegner: a experiência de livre-arbítrio como resultado de uma ilusão causal	p.94
Crítica de Beatriz Sorrentino Marques a Daniel Wegner	p.94
Agência	p.95
Beatriz Sorrentino Marques, a sensação de agência e o controle regulador	p.97
A ciência da mente e do encéfalo	p.98
Comentários finais (Bloco VII)	p.99
A sensação de livre-arbítrio não precisa ser confirmada por movimentos	p.100
Autoria e agência são noções distintas	p.101
A sensação de livre-arbítrio é resultante de um processo em duas etapas	p.102
A sensação de livre-arbítrio demanda algum grau de reflexão racional	p.102
A sensação de livre-arbítrio ocorre em graus	p.102
Exemplo	p.103
Primeira leitura conjunta das proposições anteriores	p.104
A sensação de livre-arbítrio corresponde a uma alteração mental	p.104
A sensação de livre-arbítrio corresponde a uma alteração encefálica	p.104
A sensação de livre-arbítrio expressa os graus de dúvida e de certeza	p.105
Segunda leitura conjunta das proposições anteriores	p.106

Consideração Final **p.107**

A minha escolha por estudar o tema do livre-arbítrio foi livre? p.107

Referências Bibliográficas **p.108**

CONSIDERAÇÕES INICIAIS

Propósitos da tese

Esta tese terá como propósitos: (i) apresentar as principais posições filosóficas sobre o tema do livre-arbítrio, nomeadamente o compatibilismo e o incompatibilismo; (ii) apresentar o conceito de responsabilidade moral, e avaliar como ele se relaciona ao conceito de livre-arbítrio; (iii) apresentar o problema mente-cérebro, e avaliar como ele se relaciona ao conceito de livre-arbítrio; (iv) apresentar a noção de escolha racional, e avaliar como ela se relaciona ao conceito de livre-arbítrio; (v) apresentar um breviário dos principais estudos da “neurociência do livre-arbítrio”; (vi) propor as bases para o desenvolvimento de uma ciência para a “sensação de livre-arbítrio”, que chamarei de “a ciência da mente e do encéfalo”.

Estrutura da tese

Esta tese possuirá sete blocos temáticos: (i) Livre-arbítrio: definição, sentidos e problemas; (ii) Compatibilismo e incompatibilismo; (iii) Responsabilidade moral; (iv) O problema mente-cérebro; (v) Escolha racional; (vi) A neurociência do livre-arbítrio; (vii) Uma ciência para a sensação de livre-arbítrio.

Haverá ainda uma pequeníssima seção final, de título “Consideração Final”, em que será apresentada uma resposta para a seguinte questão: “A minha escolha por estudar o tema do livre-arbítrio foi livre?”.

Método da tese

Esta tese buscará primar pela clareza e pela concisão. Posições e problemas filosóficos, estudos científicos e os comprometimentos do autor serão expostos da forma mais objetiva possível. Serão evitadas digressões e rebuscamento. Não se encontrarão, nesta tese, seções demasiado longas – o mesmo valerá para frases e parágrafos. À semelhança de José Ortega y Gasset (1883-1955), “sempre acreditei que a clareza é a gentileza do filósofo” (2016).

A abrangência temática será outra marca desta tese. O livre-arbítrio servirá como mote para a discussão de uma série de temas – o problema determinismo *versus* indeterminismo, a responsabilidade moral, o problema mente-cérebro, a escolha racional, a neurociência do livre-arbítrio, a ciência da mente e do encéfalo etc. Para cada problema, uma pluralidade de posições será apresentada. Serão utilizadas, para tanto, fontes de prestígio, como a *Stanford Encyclopedia of Philosophy*.

Os comprometimentos do autor estarão expostos, de forma mais conspícua (e menos formal), nas seções “Comentários finais”, ao final de cada bloco temático.

Originalidade da tese

É sabido que a originalidade é um requisito indispensável para uma tese de doutoramento.

No caso desta tese, no último bloco temático, serão apresentadas oito proposições sobre a “sensação de livre-arbítrio”, que, tomadas em conjunto, constituem, na visão do autor, um todo vanguardista.

Por essa razão, a parte final do referido bloco (pp. 100-106) é provavelmente a mais original e interessante de toda a tese.

BLOCO I: LIVRE-ARBÍTRIO: DEFINIÇÃO, SENTIDOS E PROBLEMAS

Definição de “livre-arbítrio”, origem do termo e sua função primitiva

O dicionário online Priberam da língua portuguesa define o vocábulo “livre-arbítrio” como “faculdade da vontade para se determinar”. O dicionário online Dicio apresenta uma definição mais informativa: “oportunidade ou possibilidade de tomar decisões por vontade própria, seguindo o próprio discernimento e não se pautando numa razão, motivo ou causa estabelecida”. Também é esclarecedora a definição ofertada pelo dicionário online Michaelis: “faculdade que o homem tem de escolher ou decidir conforme sua própria vontade, sem que haja condicionamento ou qualquer interferência nessa escolha”. Em comum, todas essas definições contêm o termo “vontade”.

A origem do termo “livre-arbítrio” é latina (*liberum arbitrium*) – a rigor, o termo apareceu pela primeira vez na obra *De Libero Arbitrio*, de Agostinho de Hipona (Santo Agostinho) (354-430). O verbete *libre arbitre* da *Wikipedia* (edição francesa) informa que, na teologia patrística latina [*patrística* foi a filosofia prevalecente nos primeiros séculos da era cristã, que teve em Agostinho de Hipona sua mais célebre figura], “livre-arbítrio” era utilizado mais frequentemente como a contração da expressão técnica “livre-arbítrio da vontade” [*liberum arbitrium voluntatis*]. Derk Pereboom, em artigo publicado na revista *Nôus* (1995), avaliza essa informação. Nas línguas correntes, “livre-arbítrio” continua a ser indissociável de “vontade livre” – a tradução inglesa *free will* (vontade livre) para a expressão latina *liberum arbitrium* o atesta.

Em seus primórdios, a noção de “livre-arbítrio” foi utilizada como uma proposta de solução para o “Problema do Mal” (ou “Paradoxo de Epicuro”). Epicuro (341 a.C. - 270 a.C.) argumenta contra a existência de um deus ao mesmo tempo onisciente, onipotente e benevolente – em verdade, apenas duas dessas características poderiam ser simultâneas (e.g., se Deus for onisciente e onipotente, ele não poderá ser benevolente, dado que existe o mal no mundo).

O “livre-arbítrio” surge, então, como a ideia de que o mal ingressou no mundo através da desobediência de Adão e Eva, vale dizer, do uso (inadequado) do livre-arbítrio pelos seres humanos. Para o pensamento cristão, Deus criou o mundo à sua semelhança – i.e., perfeito, sem sofrimento e sem mal. Agostinho de Hipona avança a ponto de defender que o mal não existe – o mal, para o filósofo, seria a mera ausência ou corrupção do bem.

No início do século XVIII, Gottfried Wilhelm Leibniz (1646-1716) criou o termo *teodiceia* para designar um conjunto de argumentos favoráveis à existência de Deus, aceitando a presença do mal no mundo. Ainda hoje, o livre-arbítrio ocupa posição de destaque entre esses argumentos (embora a posição mais popular entre os filósofos da religião contemporâneos seja o “ceticismo teísta”, segundo o qual Deus existe e os seres humanos não são capazes de compreender o divino).

Sentidos filosóficos de “livre-arbítrio”

Timothy O’Connor e Christopher Franklin são os autores do principal verbete sobre o livre-arbítrio da prestigiosa *Stanford Encyclopedia of Philosophy* [*Free Will* (2018)]. No verbete, é afirmado que o termo “livre-arbítrio” tem sido utilizado, nos últimos dois milênios, como “um designador canônico para um tipo significante de controle sobre as nossas ações”.

Ainda no verbete, sustentados na breve exposição que fazem da história da ideia de livre-arbítrio na filosofia ocidental, O’Connor e Franklin apontam para o que significa, na prática, esse “controle”: “quando um agente possui livre-arbítrio, suas escolhas e decisões dependem dele”. Sucede que, à luz da tradição filosófica, a expressão “dependem dele” comporta, para os autores, dupla interpretação: (i) o agente encontra-se apto a agir de outra maneira, ou a não fazer o que efetivamente faz; (ii) o agente é a causa (a origem) de sua ação.

Têm-se esboçados os dois sentidos filosóficos de “livre-arbítrio”. Num primeiro sentido, livre-arbítrio associa-se à ideia de *possibilidades alternativas*, e define-se como “capacidade para agir de outro modo”. Num segundo sentido, livre-arbítrio remete às noções de *agência e controle*, e tem sentido de “origem ou autodeterminação das ações”. Em que pese esses dois sentidos não serem excludentes, infere-se de O’Connor e Franklin que a tradição filosófica privilegia ora um, ora outro.

Problemas tradicionais relacionados ao livre-arbítrio

Na visão dos especialistas Timothy O’Connor e Christopher Franklin, exposta no já mencionado verbete da *Stanford Encyclopedia of Philosophy*, “disputas sobre o livre-arbítrio envolvem inelutavelmente disputas sobre a metafísica e a ética”. Ao se discutir o tipo de controle envolvido no livre-arbítrio, por exemplo, irrompem questões sobre “causação, leis da natureza,

tempo, substância, reducionismo ontológico *versus* emergência, a relação entre as explicações causais e as explicações baseadas em razões, a natureza da motivação e, mais genericamente, dos seres humanos”, entre outras.

O debate sobre o significado do livre-arbítrio, por sua vez, suscita questões como “o certo e o errado, o bem e o mal, a virtude e o vício, a culpa e o mérito, a punição e a recompensa, e o merecimento”. Há, ainda, um debate eminentemente empírico, de que a ciência se ocupa, que diz respeito à questão fundadora de se os seres humanos realmente possuem livre-arbítrio e, em caso afirmativo, em que medida.

Poder-se-ia dizer, extrapolando as considerações dos autores no verbete, que essa plethora de questões denuncia uma tendência histórica à fuga do tema entre aqueles que discutem o livre-arbítrio. Em geral, análises sobre o livre-arbítrio descambam para considerações sobre a responsabilização moral, o determinismo ou o indeterminismo do universo, e a redutibilidade ou a não redutibilidade da mente ao corpo, para ficar numa lista mais parcimoniosa do que a oferecida por O’Connor e Franklin.

Por conseguinte, não é absolutamente claro se os que discutem a questão ética (em que circunstâncias o ser humano deve ser moralmente responsabilizado por seus atos?) estão a referir-se ao mesmo “livre-arbítrio” dos que discutem a questão física (o universo segue leis deterministas ou indeterministas?), ou se estes aludem ao mesmo conceito dos que estão interessados no problema mente-corpo (a mente é redutível ao corpo, ou a mente e o corpo são entidades independentes?).

Um “novo” problema relacionado ao livre-arbítrio

Em artigo publicado no jornal *O Estado de São Paulo* (2020), o filósofo português radicado no Brasil Desidério Murcho associa as dificuldades do tema do livre-arbítrio a três problemas. O primeiro problema, “temos realmente livre-arbítrio, ou é só uma ilusão?”, dificilmente permitiria uma abordagem filosófica rendosa, na visão do pensador. Porquanto constitui questão empírica aberta, esse problema deveria ser primeiramente dirigido a cientistas, não a filósofos.

O segundo problema, “quais são as componentes mais relevantes do livre-arbítrio, e que contributo dão para uma vida humana bem vivida?”, dependeria de esclarecimentos conceituais próprios da Filosofia. De acordo com Murcho, trata-se de problema com enorme impacto prático,

e nisso também se diferenciaria do anterior, de repercussões mais teóricas. Embora essencialmente filosófico, o segundo problema guardaria relação com os estudos científicos, uma vez que estes poderiam esclarecer as condições em que as pessoas deliberam melhor, decidem melhor e fortalecem seu autocontrole.

O terceiro problema, “o livre-arbítrio é compatível com a tese de que todos os acontecimentos são efeitos de causas anteriores?”, é o que prevaleceu no debate filosófico desde a eclosão da Revolução Científica, segundo Murcho. Trate-se de questão hipotética e teórica, sem nenhum impacto sobre os outros dois problemas – independentemente de o livre-arbítrio ser compatível com a tese de que todos os acontecimentos são efeitos de causas anteriores, ele pode existir e ser relevante para uma vida humana bem vivida, ou ele pode não existir.

Confrontando os três problemas levantados por Murcho com os problemas tradicionais do livre-arbítrio mencionados na seção anterior, sobressai, pela sua novidade, o segundo problema trazido pelo filósofo português. Na Antiguidade, Platão e Aristóteles manifestavam constante preocupação com a vida humana “bem vivida”. De forma revigorada, aberta a influências da ciência, o “novo” problema do livre-arbítrio – o segundo da lista de Murcho – resgata essa preocupação.

Comentários finais (Bloco I)

A respeito da seção “Definição de ‘livre-arbítrio’, origem do termo e sua função primitiva”, quero enfatizar a presença do vocábulo “vontade” em todas as definições apresentadas de “livre-arbítrio”. Também me parece relevante a informação de que “livre-arbítrio”, nos textos latinos, era apenas a contração da expressão “livre-arbítrio da vontade”. Essa informação me parece fiável, na medida em que, nos dicionários de latim que pude consultar, “livre-arbítrio” (*liberum arbitrium*) é sinônimo de “vontade livre” (*libera voluntas*).

O sinal de hífen em “livre-arbítrio” é obrigatório – trata-se de palavra composta. Como consequência, “livre-arbítrio” e “arbítrio livre” só designarão a mesma coisa se entendermos a palavra “arbítrio” em seu sentido literal, que é o de *decisão dependente apenas da vontade* (dicionário online Google). Em qualquer sentido conotativo do termo “arbítrio” (e.g., “arbítrio” enquanto “juízo”), “livre-arbítrio” e “arbítrio livre” designarão coisas distintas. Mesmo que

o meu “julgamento” seja livre, pode ser que a vontade que o conduziu não seja livre – e essa *vontade* é tudo o que importa para a definição de livre-arbítrio.

Nesta tese, serão exibidas tentativas de se distinguir “liberdade de ação” de “liberdade da vontade”. Essa distinção me parece absolutamente legítima, desde que não se queira, com ela, mostrar que o livre-arbítrio é compatível com o determinismo, pois a “liberdade de ação” o é. Os dicionários não nos autorizam a tratar “livre-arbítrio” como sinônimo de “liberdade de ação”, se o fizermos em detrimento da “liberdade da vontade”. “Livre-arbítrio” é, por definição, “liberdade da vontade”. Quem crê em “livre-arbítrio” sem crer em “liberdade da vontade” está provavelmente dando nome errado a sua crença.

No que toca à seção “Sentidos filosóficos de ‘livre-arbítrio’”, entre os dois sentidos de livre-arbítrio apresentados, a lembrar, “capacidade para agir de outro modo” e “origem (autodeterminação) das ações”, quero registrar, a princípio, minha preferência pelo segundo, que me parece mais facilmente aplicável à ciência. “Poder agir de outro modo” remete a *possibilidades alternativas*, vale dizer, a cenários contrafactuais que efetivamente não ocorrem e que, portanto, são menos propícios à investigação empírica (quando não totalmente inadequados). “Origem (autodeterminação) das ações”, por outro lado, remete às noções de *agência e controle*, que penso serem mais “amigáveis” à ciência.

Não obstante, parece-me desnecessário falar em “ação” (e mesmo em *agência*) para referir-se ao livre-arbítrio, pois nem toda escolha (ou decisão) pressupõe uma ação posterior que a confirme. Eu posso decidir hoje, por exemplo, o que farei na minha aposentadoria (e.g., em que cidade irei morar), ainda que esse dia esteja muito longe de chegar, e o fato de essa decisão não ser imediatamente confirmada por uma ação não me parece um impedimento para eu considerá-la livre.

Por esse motivo, proponho uma sutil alteração para o que chamei de “o primeiro sentido filosófico de ‘livre-arbítrio’”: (i) “livre-arbítrio” significa “poder *decidir* (ou *escolher*) de outro modo” (independentemente de essa decisão – ou escolha – ser ou não confirmada por uma ação posterior).

Gostaria, igualmente, de alterar o “segundo sentido filosófico” de livre-arbítrio: (ii) “livre-arbítrio” significa “origem ou determinação das decisões pela vontade consciente, o que remete às noções de *autoria* (em vez de *agência*) e *controle*” [*autoria*, no sentido de que a vontade consciente é a *autora* da decisão; e *controle*, no sentido de que a vontade consciente tem *controle* sobre ela].

Aqui, vale destacar que são duas as alterações que empreendi: substituí, uma vez mais, a noção de *ação* pela noção de *decisão*; e abdiquei da noção de “autodeterminação”, que me parece imprecisa.

A rigor, numa situação de livre-arbítrio, as decisões não são “autodeterminadas”. Se o fossem, poder-se-ia considerar que elas ocorrem “por acaso” (se as decisões fossem “autodeterminadas”, alguém poderia interpretar que elas “brotam do nada”), o que não favorece, em absoluto, as noções de *autoria* e *controle* embutidas no livre-arbítrio.

Decidir livremente (no sentido de decidir com livre-arbítrio) significa decidir a partir dos ditames da vontade consciente (e apenas deles). É a vontade consciente, pois, que é a *origem* das decisões e que as *determina* numa escolha com livre-arbítrio.

Recapitulando, quando se fala em “livre-arbítrio”, podemos discriminar uma definição para o termo e dois “sentidos” filosóficos. A definição de livre-arbítrio, exibida pelos dicionários, é algo próximo a “vontade consciente livre, isenta de condicionamentos, a partir da qual se delibera”. Os “sentidos” dessa definição [desde Gottlob Frege (1848-1925), sabemos que um mesmo “referente” pode possuir mais de um “sentido” – no célebre exemplo do lógico, ao “planeta Vênus” associam-se tanto o sentido de “estrela da manhã” quanto o sentido de “estrela da tarde”] são: (i) *possibilidades alternativas* de escolha, que implica “capacidade para decidir de outro modo”; e (ii) *autoria e controle* sobre a escolha, que implica “origem e determinação da escolha pela vontade consciente”.

Na sequência do texto, seguiremos com essa definição e com esses dois sentidos de livre-arbítrio (de resto, bastante consagrados pela literatura filosófica), enfatizando, sempre que possível, a nossa preferência pelo segundo sentido.

BLOCO II: COMPATIBILISMO E INCOMPATIBILISMO

Compatibilismo

O verbete “Compatibilismo” [*Compatibilism*], da *Stanford Encyclopedia of Philosophy*, assinado por Michael McKenna e D. Justin Coates, informa que o compatibilismo é uma posição que emerge como uma resposta ao problema do determinismo causal. O determinismo causal é a tese de que “tudo o que ocorre é o resultado inevitável das leis da natureza e do estado do mundo num passado distante”.

O compatibilismo é definido como a tese de que o livre-arbítrio é compatível com o determinismo causal. Dado que o livre-arbítrio costuma ser considerado uma condição para a responsabilidade moral, muitas vezes o compatibilismo também é definido como a tese de que a responsabilidade moral é compatível com o determinismo causal.

Mas não sucederia o contrário? Não seria mais razoável considerar o determinismo causal um empecilho para o livre-arbítrio, seja enquanto “liberdade para decidir de outro modo”, seja enquanto “origem e determinação das decisões pela vontade consciente”?

Os autores do verbete informam que Carl Ginet, Peter van Inwagen, Derk Pereboom e outros pensadores exploraram as linhas do raciocínio anterior, formalizando o incompatibilismo, isto é, a tese de que “nossa autopercepção como agentes livres e responsáveis ficaria seriamente prejudicada se o determinismo causal se mostrasse verdadeiro”. A posição incompatibilista será apresentada em mais detalhes em seção oportuna nesta tese.

No campo compatibilista, são variantes as estratégias de defesa da posição. Alguns acreditam que a verdade do determinismo causal não mina “nossa liberdade para decidir de outro modo” – McKenna e Coates mencionam Bernard Berofsky, C.A. Campbell e Kadri Vihvelin. Outros rejeitam que a “liberdade para decidir de outro modo” seja condição necessária para o livre-arbítrio – nesse time, segundo McKenna e Coates, estão Harry Frankfurt, Gary Watson e Daniel Dennett. Há ainda os que negam qualquer ligação entre a “liberdade para decidir de outro modo” e a responsabilidade moral – à frente, John Martin Fischer, Mark Ravizza, T. M. Scanlon, B. Allan Wallace e Carolina Sartorio.

No que toca à evolução histórica, de acordo com McKenna e Coates, o compatibilismo se desenvolveu em três estágios. O primeiro estágio, que os autores chamam de “a forma clássica do

compatibilismo”, é tributário das contribuições de dois filósofos empiricistas da era moderna: Thomas Hobbes (1588-1679) e David Hume (1711-76).

O segundo estágio envolve três contribuições distintas dos anos 1960 – o argumento da consequência [*consequence argument*] de Carl Ginet, de teor incompatibilista, o argumento de Harry Frankfurt contra o princípio das possibilidades alternativas (PAP), e a concepção de responsabilidade moral de Galen Strawson, com ênfase nas relações interpessoais e nas atitudes reativas.

O argumento da consequência (defendido, reforce-se, por incompatibilistas) seria recuperado na década de 1980 por Peter van Inwagen (*An Essay on Free Will*, 1983), que o apresenta, em linhas gerais, deste modo: partindo da premissa de que o determinismo é pura e simplesmente “a tese de que o passado determina um futuro único” (van Inwagen, p.2, tradução nossa), caso o determinismo seja verdadeiro, nós não poderemos ser moralmente responsabilizados por nossos atos, pois estes serão, no caso, a mera *consequência* de eventos do passado distante que, em sua origem, “não dependiam de nós” [*were not up to us*].

O terceiro estágio do compatibilismo contempla várias formas contemporâneas de compatibilismo – McKenna e Coates mencionam seis delas –, todas influenciadas por ao menos uma das contribuições do estágio anterior. Entre essas novas formas de compatibilismo, está a posição de Susan Wolf.

Wolf (1993) defende que o livre-arbítrio é a capacidade para agir de acordo com a razão, com o pleno conhecimento da verdade e da bondade. A posição da autora, de clara inspiração platônica, assevera que os seres humanos são livres na medida em que fazem a coisa “certa”, e não livres (i.e., escravos dos desejos e das paixões) quando fazem a coisa “errada”.

O compatibilismo de Agostinho

Agostinho de Hipona (Santo Agostinho) (354-430), criador do termo “livre-arbítrio” [empregado de forma pioneira na obra *De Libero Arbitrio*], defende, no início do terceiro livro da obra, que a onisciência e a presciência de Deus são compatíveis com o livre-arbítrio dos humanos. Para o filósofo medieval, o fato de Deus conhecer previamente todas as ações humanas não invalida o livre-arbítrio, pois as ações humanas não são *causadas* pelo conhecimento de Deus.

De forma análoga, poderia haver alguém no mundo (um outro ser humano) que soubesse exatamente o que estamos fazendo agora (e.g., através da observação), mas não diríamos que o seu conhecimento é a *causa* de nossa ação, ou que compromete nossa liberdade. Com efeito, o mero fato de eu saber o que você está fazendo agora (no caso, o fato de eu saber que você está lendo esta tese) não é um impedimento para você agir livremente.

Para Agostinho, sucede o mesmo em nossa relação com Deus (que nos observa a todo instante). Se compreendêssemos que Deus existe de forma independente do tempo, entenderíamos que nós só enxergamos o presente, mas Deus enxerga o conjunto da história, sem comprometer nossa liberdade. Afinal, ainda que Deus saiba quais serão as nossas escolhas amanhã, as escolhas continuarão sendo *nossas*, e não dele.

O compatibilismo de Thomas Hobbes

Thomas Hobbes (1588-1679) sustenta posição materialista e determinista, e explica a natureza humana a partir de pressupostos da ciência. Além disso, defende que os atos da vontade são causados por processos físicos – por essa razão, a vontade não seria livre ou espontânea.

Mas, para Hobbes, a liberdade enquanto “ausência de todos os impedimentos para a ação, que não estejam contidos na natureza e na qualidade intrínseca do agente” (Hobbes, 1841, p. 367, tradução nossa), vale dizer, a liberdade enquanto ausência de restrições externas, é inequivocamente experimentada pelos humanos.

A rigor, seria correto observar que Hobbes não crê exatamente no livre-arbítrio (no sentido clássico, de “liberdade da vontade”), mas é patente a preocupação do filósofo em conciliar o determinismo com algum tipo de liberdade – no caso, com a “liberdade de ação”.

O compatibilismo de David Hume

É célebre a distinção de David Hume (1711-76) entre “liberdade de espontaneidade” e “liberdade de indiferença”: uma pessoa que não esteja submetida à violência pode dizer-se livre no sentido de ser capaz de agir sem constrangimentos externos (essa liberdade, Hume batizou-a de “liberdade de espontaneidade”). Mas isso não diz nada sobre a sua vontade. Em particular, o fato

de a ação ser livre não significa que a vontade que a engendra também o seja (essa outra liberdade, Hume batizou-a de “liberdade de indiferença”).

Para Hume, adepto do empiricismo radical, a liberdade de espontaneidade não implica uma negação da necessidade e das causas da ação. Em outras palavras, os seres humanos podem ter liberdade de espontaneidade (a capacidade para agir sem a imposição ou a restrição física de outrem), mas não possuem liberdade de indiferença (a capacidade para agir extrapolando as leis deterministas da vontade).

Ao reputar como “absurda e ininteligível” a “doutrina da liberdade” (Hume, 1739, p. 297, tradução nossa), i.e., a doutrina que não reconhece a necessidade das cadeias causais da natureza, Hume defende que a origem do problema do livre-arbítrio está na ambiguidade do termo. Se os filósofos entendessem por “livre-arbítrio” apenas a “liberdade de espontaneidade”, não haveria mais dificuldades em conciliá-lo com o determinismo causal, na visão do pensador.

Liberdade e livre-arbítrio

A partir da distinção entre “liberdade de ação” e “liberdade da vontade” (ou entre “liberdade de espontaneidade” e “liberdade de indiferença”), percebe-se uma diferença entre os conceitos de “liberdade” e “livre-arbítrio”.

Com efeito, o conceito de “liberdade” é mais abrangente do que o de “livre-arbítrio”, pois engloba, além da “liberdade da vontade” (que é a própria definição de livre-arbítrio), a “liberdade de ação”.

Por consequência, uma pessoa pode agir *livremente* mesmo sem exercer seu livre-arbítrio. Por outro lado, sempre que uma pessoa exercer seu livre-arbítrio (supondo que ela o possua), ela agirá *livremente*.

Isaiah Berlin (1909-1997), filósofo britânico de origem russa, propõe, em célebre palestra na Universidade de Oxford, em 1957, uma distinção adicional, entre “liberdade positiva” e “liberdade negativa”. Sua distinção, é bom frisar, aplica-se mais ao campo da ciência política – o filósofo refere-se a *liberdades políticas*, portanto.

Por “liberdade negativa”, Berlin entende a liberdade de agir sem a “obstrução” (ou a “coerção”) de outros homens. Por “liberdade positiva”, o filósofo alude à liberdade que deriva da vontade do indivíduo “de ser seu próprio mestre” [*to be his own master*] – i.e., de ser sujeito, e

não objeto, de ser movido pela consciência, e não por razões externas, etc. Para Berlin, a “liberdade positiva”, na medida em que pressupõe a *consciência de si*, é um dos fundamentos da ideia de racionalidade humana.

Exemplos de Frankfurt contra o princípio das possibilidades alternativas

Em verbete da *Stanford Encyclopedia of Philosophy* dedicado ao princípio das possibilidades alternativas [*Principle of Alternative Possibilities* (PAP)], David Robb menciona dois “casos de estilo Frankfurt” [*Frankfurt style cases*], referindo-se a exemplos desenvolvidos pelo nonagenário filósofo estadunidense Harry Frankfurt, que invalidariam esse princípio.

O princípio das possibilidades alternativas (PAP) afirma o seguinte: uma pessoa é moralmente responsável por um ato que praticou se, e somente se, pudesse ter agido de outra forma no momento da ação.

No primeiro exemplo mencionado por Robb, Frankfurt apela a um “vício adormecido” [*dormant addiction*], isto é, a um vício inconsciente que, em algum momento, fará determinado indivíduo fazer uso de uma droga. Para o indivíduo, que não tem consciência de seu vício, o uso da droga decorrerá de sua livre vontade, e parecerá absolutamente evitável.

Para Frankfurt, o vício não abstém o “viciado dormente” da responsabilidade moral. Isso ocorre porque, na perspectiva do “viciado dormente” – e essa perspectiva é a que conta para a responsabilidade moral –, não tomar a droga é uma opção. Tem-se um caso em que o princípio das possibilidades alternativas (PAP) é ferido: o indivíduo é moralmente responsabilizado mesmo não podendo agir de outra forma.

No segundo exemplo mencionado por Robb, recuperado de John Martin Fischer, Frankfurt apela a um neurocirurgião que implanta um dispositivo no cérebro de um paciente, sem o seu conhecimento, para monitorar e controlar suas preferências. Mais precisamente, o dispositivo monitora e controla as preferências políticas do paciente. Se o desejo espontâneo do paciente é votar no Partido Democrata, o dispositivo altera as suas preferências, fazendo-o querer votar no Partido Republicano. Se o desejo espontâneo do paciente é votar no Partido Republicano, o dispositivo não altera suas preferências.

Para Frankfurt, se o desejo espontâneo do paciente é votar no Partido Republicano, o dispositivo instalado em seu cérebro não o abstém da responsabilidade moral, caso ele venha a

confirmar esse voto. Tem-se mais um caso em que o princípio das possibilidades alternativas (PAP) é ferido: mesmo na ausência de possibilidades alternativas, o indivíduo é moralmente responsabilizado por seu ato.

Em comum, os dois “casos de estilo Frankfurt” mencionados por Robb apontam para um agente que atua “por conta própria” [*for his own reasons*]. Nesses casos, para Frankfurt, há responsabilidade moral, independentemente de haver possibilidades alternativas.

Harry Frankfurt notabilizou-se, ademais, pelos conceitos de *desejos (ou volições) de primeira e de segunda ordens*. Um desejo de primeira ordem é simplesmente o desejo de se fazer algo. Um desejo de segunda ordem, ao contrário, é o desejo de se ter (ou de não se ter) certo desejo.

O livre-arbítrio, para Frankfurt, corresponde à situação em que o desejo (ou *volição*) de segunda ordem determina o desejo de primeira ordem. De acordo com o filósofo, os seres humanos são livres na medida em que conseguem determinar quais são as suas vontades.

O compatibilismo de Gilberto Gomes

Gilberto Gomes, em *Free Will, the Self, and the Brain* (2007), observa que o termo livre-arbítrio não pode ser meramente definido como “propriedade ou capacidade para decidir de outro modo”: há também que levar em consideração as circunstâncias, os desejos e as crenças que contribuem para as ações.

De fato, se nos fosse facultado alterar a cada instante as circunstâncias em que deliberamos, possivelmente não decidiríamos de forma sempre idêntica, quero dizer, nossas decisões provavelmente mudariam à medida que alterássemos suas circunstâncias; mas é evidente que, no sentido forte, não é isso o que nos torna livres. Possuiremos livre-arbítrio se, e somente se, nossa “propriedade ou capacidade para decidir de outro modo” persistir mesmo com circunstâncias, desejos e crenças fixados.

Em artigo mais recente (2019), Gilberto Gomes oferece uma resposta a Marcelo Fischborn, que o definira como “incompatibilista” (e não como “compatibilista”, que é a forma como o próprio Gomes se apresenta no debate sobre o livre-arbítrio).

Gomes admite não endossar o “determinismo estrito”. Ainda assim, o autor advoga pela compatibilidade entre a “determinação causal local e probabilística”, de um lado, e a “vontade

livre”, de outro, razão pela qual se vê como um “compatibilista”. O seu compatibilismo, porém, constitui “visão mais flexível” do que o compatibilismo tradicional.

Via de regra, o determinismo é definido como “a tese de que o estado do universo em um determinado momento, conjugado às leis da natureza, implica o estado do universo em qualquer outro momento” (Gomes, 2019, p. 114). Para Gomes, essa é uma definição de “determinismo estrito”. Haveria, igualmente, outras versões de determinismo disponíveis.

Jay E. Gould, por exemplo, lançou mão do conceito de “determinismo probabilístico (ou estatístico)” – Gomes entende que a expressão “causalidade probabilística” seria mais oportuna, evitando contradições – para designar “a tese de que, sob condições especificáveis, um certo resultado ocorrerá, mas só em uma certa porcentagem dos casos” (Gomes, 2019, p. 114). Aparentemente, é com essa posição que o autor brasileiro se compromete.

Nesse ponto, um incompatibilista poderia argumentar que o “determinismo probabilístico” é uma forma de indeterminismo, e não de determinismo, a não ser pelo nome – i.e., é um indeterminismo no qual se podem prever distribuições estatísticas.

Gilberto Gomes rejeita taxativamente que o “determinismo estrito” seja uma consequência necessária do conhecimento científico atual. Segundo o autor, nem a mecânica clássica (newtoniana) nem a mecânica relativista (einsteiniana) teriam essa forma de determinismo como corolário. Considerações de John Earman – e.g., de que “a física newtoniana não é um paraíso para o determinismo”, e de que a Teoria Geral da Relatividade (TGR) não é favorável ao determinismo – endossam a visão de Gomes.

Por outro lado, deve-se observar que a Teoria da Relatividade Restrita (TRR) favorece o determinismo, segundo argumenta Earman. O problema na Mecânica Clássica é que, em sistemas com velocidades infinitas, poderia haver uma quebra do determinismo, mas a TRR exclui tal possibilidade.

Para Gomes, o “determinismo estrito” não é uma pressuposição necessária para a ciência. O fato de muitos cientistas o subscreverem seria, antes, expressão de uma “ideologia”, de um “preconceito”, em que pese Bitsakis referir-se ao “otimismo gnosiológico”, i.e., à “confiança na possibilidade de nosso intelecto descobrir as leis da natureza” (Gomes, 2019, p. 116), como decorrente da crença no “determinismo estrito”.

Essa crença de muitos cientistas (mas não de todos) no “determinismo estrito” seria justificada por razões psicológicas, segundo Gomes: em particular, pela ideia de que, a partir do

“determinismo estrito”, é possível “amenizar nossa angústia diante do desconhecido, pois a ordem é mais tranquilizadora do que o desconhecido” (Gomes, 2019, p.117).

Fundamentalmente, para Gilberto Gomes, se o “determinismo estrito” for verdadeiro, será inútil argumentar em seu favor, pois a argumentação de que o determinismo é verdadeiro será, ela própria, pré-determinada; aliás, qualquer argumentação com vistas ao convencimento de outrem será inócua, pois o fato de as pessoas se convencerem ou não já estará, desde sempre, pré-determinado.

Não por outra razão, o autor brasileiro observa, com felicidade, que “se o determinismo [estrito] é verdadeiro, é impossível determinar que ele é verdadeiro” (Gomes, 2019, p.118).

Reputando o “determinismo estrito” como uma visão filosófica “particular e extremada”, Gilberto Gomes não abdica da crença de que as ações humanas possuem explicações causais. Para o filósofo brasileiro, a “possibilidade de decidir de outra forma” – mesmo com circunstâncias (externas) fixas [Gomes entende que as condições internas ao processo decisório podem ser variáveis, sem que isso comprometa a ideia de livre-arbítrio] – é compatível com a determinação probabilística.

Por defender a compatibilidade entre causalidade e liberdade, Gilberto Gomes é um legítimo representante da posição compatibilista.

Incompatibilismo

Kadri Vihvelin é a autora do verbete “Argumentos para o incompatibilismo” [*Arguments for Incompatibilism*], da *Stanford Encyclopedia of Philosophy*. No verbete, o incompatibilismo é apresentado como a seguinte tese condicional: se o determinismo for verdadeiro, a crença de que possuímos livre-arbítrio será falsa.

Em suas considerações preliminares, a filósofa faz pertinentes esclarecimentos. Em primeiro lugar, reconhece a abrangência do termo “determinismo”. O sentido que ela procura fixar no verbete é o de “determinismo causal” (ou “determinismo nomológico”), isto é, a tese de que “tudo o que acontece, inclusive tudo o que você escolhe e faz, é determinado por fatos do passado e pelas leis da natureza”.

O “determinismo causal” contrasta com outros dois determinismos, ao menos: o “determinismo lógico” – a tese de que o princípio da bivalência, pelo qual uma sentença é

verdadeira ou é falsa, se aplica a todas as proposições, inclusive às proposições sobre o futuro –, e o “determinismo teológico” – a tese de que Deus existe e tem perfeito conhecimento de tudo, inclusive de nosso futuro.

“Determinismo lógico” e “determinismo teológico” também ensejam problemas de compatibilidade com o livre-arbítrio: a existência de sentenças verdadeiras sobre o futuro e a onisciência de Deus parecem entrar em choque com a ideia de que podemos agir livremente. Não obstante, as atenções de Vihvelin no verbete estão voltadas para o “determinismo causal”, que é mais presente no debate filosófico.

A autora também procura distinguir o livre-arbítrio da responsabilidade moral, reconhecendo que a tradição filosófica, muitas vezes, trata aquele como condição para esta. Para Vihvelin, o problema da compatibilidade diz respeito ao livre-arbítrio e ao determinismo causal, apenas. Considerações sobre a responsabilidade moral, se utilizadas, devem ser meros meios auxiliares na argumentação sobre a compatibilidade (ou a incompatibilidade) entre livre-arbítrio e determinismo causal.

Os argumentos de incompatibilidade são então apresentados pela filósofa. Em síntese, são dois argumentos: um para mostrar que o determinismo causal é incompatível com a ideia de que a origem das decisões humanas está na vontade consciente, i.e., de que a vontade consciente possui autoria e controle sobre elas; outro para mostrar que o determinismo causal é incompatível com a ideia de que o ser humano tem possibilidades alternativas de escolha, i.e., de que o ser humano pode “decidir de outro modo”.

No primeiro caso, vigora a premissa de que uma “decisão livre” é aquela causada e controlada pela vontade consciente, vale dizer, de que livre-arbítrio é “origem ou determinação das decisões pela vontade consciente”. Partindo dessa definição, o argumento procura mostrar que o determinismo causal, ao postular que todas as decisões decorrem de fatos do passado e das leis da natureza, tira da vontade consciente a autoria sobre as escolhas, inviabilizando o livre-arbítrio.

No segundo caso, pressupõe-se que, em contraste com a natureza fixa do passado, o ser humano seria “livre” na medida em que seu futuro estivesse em aberto – em síntese, que livre-arbítrio significa “poder decidir de outro modo”. Partindo dessa definição, o argumento aduz que, se o determinismo causal for o caso, haverá uma única continuação possível do passado consistente com as leis da natureza; portanto, se o determinismo causal for o caso, “nós escolheremos, sempre, a única coisa que poderemos escolher”, e não haverá livre-arbítrio.

A favor da tese incompatibilista, Katri Vihvelin apresenta ainda argumentos baseados na intuição. Inicialmente, é exibido o “argumento da inexistência de caminhos bifurcados” [*no forking paths argument*]: sempre que fazemos uma escolha, agimos como um viajante que está diante de uma bifurcação. As únicas estradas disponíveis para o viajante são as continuadoras da estrada atual. Se o determinismo causal for falso, o viajante realmente terá duas ou mais estradas à sua frente, e poderá escolher uma delas. No entanto, se o determinismo causal for verdadeiro, toda a jornada do viajante ocorrerá necessariamente numa só estrada, sem bifurcações. Esse argumento tem por objetivo mostrar que, se o determinismo causal for o caso, não poderá existir livre-arbítrio, pois só haverá um único futuro possível.

Por fim, é examinada uma família de “argumentos da manipulação” [*manipulation arguments*]. Esses argumentos apelam a dois personagens: a vítima e o manipulador. Em geral, este é o criador daquela, possuindo controle sobre suas vontades e ações. A vítima desconhece que suas vontades e ações são manipuladas, e tampouco sabe da existência do manipulador. Os argumentos procedem, então, da seguinte maneira: a vítima não pode ser moralmente responsabilizada por seus atos, pois não age de maneira livre. Suas ações são produto da vontade do manipulador. Se o determinismo causal for verdadeiro, todos os seres humanos farão o papel da vítima. Nessa hipótese, as ações humanas não terão origem nos seres humanos, mas nos fatos do passado e nas leis da natureza. Logo, se o determinismo causal for o caso, não poderá haver livre-arbítrio.

O incompatibilismo de Robert Kane

Robert Kane é um dos mais aclamados filósofos incompatibilistas contemporâneos. No artigo *Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism* (1999), o professor aposentado da Universidade do Texas reage ao “princípio da sorte” [*the luck principle (LP)*], formulado em termos próximos aos seguintes: “se uma ação é *indeterminada* no instante *t*, então o fato de ela ocorrer ou não no instante *t* é uma mera questão de *acaso* ou *sorte*, pelo que não se trata de uma ação *livre* ou de uma ação que enseja *responsabilidades*”.

Para Kane, o princípio da sorte, comumente empregado contra a posição incompatibilista, é falso. Enquanto “acaso” e “sorte” são termos da linguagem ordinária que designam eventos “fora de nosso controle”, “indeterminismo” é termo técnico que se opõe à ideia de “causação

determinista”. O princípio da sorte contém uma inferência inválida, na medida em que é possível um evento ser indeterminado e, ao mesmo tempo, estar sob nosso controle. É o que sucede, no exemplo do filósofo, quando uma pessoa tenta resolver um problema matemático num ambiente barulhento. Ainda que a resolução do problema torne-se indeterminada em função do barulho, caso a pessoa obtenha êxito, não será razoável atribuir seu êxito ao acaso ou à sorte.

Desse modo, se, por um lado, “o indeterminismo funciona como um obstáculo ao sucesso” (Kane, 1999, p.227, tradução nossa), por outro lado, “o indeterminismo, em si mesmo, não necessariamente mina o controle e a responsabilidade” (Kane, 1999, p.226, tradução nossa). Essa seria, aliás, uma ideia bastante natural, porquanto “a maior parte das pessoas é naturalmente incompatibilista” (Kane, 1999, p.218, tradução nossa). Na filosofia contemporânea, entretanto, as posições estão invertidas. O compatibilismo é a “posição reinante” (Kane, 1999, p.218, tradução nossa). Para Kane, a prevalência do princípio da sorte explicaria essa inversão.

O conceito-chave do filósofo estadunidense são as “ações de autoformação” [*self-forming actions* (SFA)], que garantem o livre-arbítrio ao ser humano. As SFAs ocorrem em situações de impasse, em que “se deve superar a tentação de se fazer algo que nós também desejamos muito” (Kane, 1999, p.224, tradução nossa). Nessas circunstâncias, “nossos cérebros afastam-se do equilíbrio termodinâmico”, e estão sensíveis a “microindeterminações no nível neural”. Por óbvio, o resultado do impasse não é predeterminado, ainda que seja “desejado”, “racional” e “voluntário” (num impasse, todas as opções são “desejadas”, “racionais” e “voluntárias”, pois são igualmente fortes).

Numa demonstração de perspicácia sobre a natureza humana, Robert Kane observa que, nos casos em que as SFAs operam, “enquanto nós desejamos superar a tentação, nós também desejamos falhar, por razões incomensuráveis e bastante diferentes” (Kane, 1999, p.225, tradução nossa). No entanto, ao tomarmos uma deliberação, as indeterminações iniciais transformam-se em escolha determinada – no nível cerebral, a via neural correspondente à opção vitoriosa alcança um “limite de ativação, superando o indeterminismo gerado pelas demais” (Kane, 1999, p.226, tradução nossa).

As SFAs moldam o caráter do ser humano. Nas palavras de Kane, elas marcam “o início de um ‘experimento de valor’, cuja justificação está no futuro e não no passado” (Kane, 1999, p.238, tradução nossa). São, em realidade, atos arbitrários (embora “desejados”, “racionais” e “voluntários”), pelos quais “assumem-se riscos sem prévia garantia de sucesso” (Kane, 1999,

p.238, tradução nossa). Por meio dessas ações, os seres humanos tornam-se autores de suas histórias. Na feliz distinção proposta por Jan Branson, recuperada por Kane na última nota de seu artigo, num ato de livre-arbítrio, os seres humanos não escolhem apenas “alternativas para si” [*alternatives for oneself*]; escolhem, fundamentalmente, “alternativas de si” [*alternatives of oneself*].

O incompatibilismo de Derk Pereboom

Em artigo de 1995, o filósofo neerlandês Derk Pereboom traça o “perfil demográfico” dos participantes do debate sobre o livre-arbítrio, atestando a prevalência de “deterministas suaves” [*soft determinists*]. O termo “determinismo suave” foi cunhado pelo psicólogo estadunidense William James (1842-1910) para designar a tese de que “nós possuímos a liberdade requerida para a responsabilidade moral; o determinismo é verdadeiro; e essas coisas são compatíveis” (Pereboom, 1995, p.21, tradução nossa). O termo antípoda “determinismo duro” [*hard determinism*], também cunhado por James, designa a tese de que “o determinismo é verdadeiro; e nós não possuímos a liberdade requerida para a responsabilidade moral” (Pereboom, 1995, p.21, tradução nossa). Essa é a posição de Pereboom, reputada, pelo próprio, como “raramente defendida”.

Entre o “popular” determinismo suave e o “impopular” determinismo duro, haveria ainda uma posição intermediária: o “libertarianismo” – a tese de que nós possuímos a liberdade requerida para a responsabilidade moral, e o determinismo é falso. O libertarianismo é posição minoritária, se comparada ao determinismo suave, e majoritária, se comparada ao determinismo duro. Para Pereboom, “esse foi sempre o padrão” (Pereboom, 1995, p.21, tradução nossa) na história da filosofia ocidental.

Na visão do professor da Universidade Cornell (NY), algum “livre-arbítrio” é necessário para a responsabilidade moral (no que discordamos frontalmente), e o libertarianismo, o determinismo suave e o determinismo duro constituem posições mutuamente excludentes. Pereboom reconhece que todas essas posições apresentam vulnerabilidades. Contra o libertarianismo, pesam supostas violações das leis da física “sempre que uma ação livre é praticada”. Contra o determinismo suave, pesa a atribuição de culpa ao indivíduo que pratica ato errado, a despeito de, por hipótese, esse ato não estar sob seu controle. Por fim, contra o

determinismo duro, pesa a ausência de responsabilização moral em quaisquer circunstâncias, mesmo em face dos atos mais atrozés.

Pereboom, um “determinista duro”, entende que essa é a posição “mais fácil de ser aceita”. Em defesa de sua posição, o filósofo argumenta que “o determinismo duro não precisa ser do tipo mais duro”. Isso significa que ele “não precisa subverter o compromisso a fazer o que é certo”, e é compatível com a manutenção de bons relacionamentos interpessoais. Seja como for, a posição de Pereboom mereceria, ao menos, uma maior consideração por parte dos filósofos, na visão do próprio.

Comentários finais (Bloco II)

Há três comentários que gostaria de fazer acerca deste bloco temático.

Em primeiro lugar, na seção “Compatibilismo”, é afirmado que o livre-arbítrio é muitas vezes entendido como condição para a responsabilidade moral, razão pela qual a tese compatibilista costuma ser indistintamente apresentada como “compatibilidade entre livre-arbítrio e determinismo causal” e “compatibilidade entre responsabilidade moral e determinismo causal”.

Ora, essa informação me parece pertinente, pois de fato constato essa tendência na literatura filosófica. De todo modo, é preciso tomar muito cuidado com o significado da palavra “condição”. Quando alguém diz que o livre-arbítrio é “condição” para a responsabilidade moral, o que está querendo dizer?

Não pode querer dizer, penso eu, que, sem “livre-arbítrio”, jamais haverá responsabilidade moral. A responsabilidade moral é um fato social: as sociedades responsabilizam moralmente os indivíduos. É bom frisar que esse processo nem sempre é justo: algumas vezes, somos responsabilizados por atos que sequer praticamos. Logo, para haver responsabilização moral, não é, em princípio, sequer necessário que haja “ação” do indivíduo (quanto mais “livre-arbítrio”).

Pode-se dizer, entretanto, que, se uma ação for livre (no sentido de ser a consequência de uma escolha com livre-arbítrio), isso já será suficiente para que o indivíduo seja moralmente responsabilizado. Se uma pessoa poderia não praticar um ato errado e mesmo assim decidiu praticá-lo (a partir de uma decisão livre e espontânea, em que ela apenas seguiu os ditames de sua vontade consciente), é razoável que isso constitua uma causa suficiente para responsabilizá-la em termos morais.

Há, portanto, que distinguir “condições suficientes” de “condições necessárias”. O livre-arbítrio não é uma “condição necessária” para a responsabilidade moral (pois pode haver responsabilização moral mesmo sem haver livre-arbítrio), mas é uma “condição suficiente” (se houver livre-arbítrio, é natural que haja responsabilização moral).

Se é assim, resulta claro, por um lado, que há um vínculo entre livre-arbítrio e responsabilidade moral. Mas é igualmente claro, por outro lado, que esse vínculo não justifica a substituição dos vocábulos no enunciado da tese compatibilista. Uma coisa é dizer que o livre-arbítrio (suposto fato da natureza humana) é compatível com o determinismo causal. Outra coisa, bem distinta, seria dizer que a responsabilidade moral (fato da sociedade) é compatível com o determinismo causal.

Essa substituição de vocábulos na tese compatibilista só faria sentido se o livre-arbítrio fosse condição necessária para a responsabilidade moral (se uma coisa não pode existir sem a outra, é natural que elas sejam intercambiáveis quando queremos analisar sua compatibilidade com um terceiro elemento). No entanto, como o livre-arbítrio é uma condição suficiente (e não necessária) para a responsabilidade moral, não me parece legítimo avaliar sua compatibilidade com o determinismo causal à luz da compatibilidade deste com a responsabilidade moral.

Por isso, na seção “Incompatibilismo”, parece-me que Kadri Vihvelin procede com correção quando afasta a responsabilidade moral das análises centrais sobre a tese compatibilista, tratando-a como mero meio auxiliar. A distinção entre “condições suficientes” e “condições necessárias”, porque me parece muito relevante, será abordada em mais detalhes na sequência da tese.

Em segundo lugar, gostaria de recuperar a distinção entre as três formas de determinismo, levadas a cabo por Vihvelin e apresentadas na seção “Incompatibilismo”.

É lamentável que filósofos interpretem “determinismo”, no mais das vezes, apenas como “determinismo causal” (nomológico), ignorando outros sentidos possíveis do termo. Se Aristóteles afirmou que “pode-se dizer do ‘ser’ em vários sentidos”, talvez fosse o caso de acrescentarmos: “pode-se dizer do ‘determinismo’ em vários sentidos”.

Um teólogo abordará o tema do livre-arbítrio interessado em avaliar a compatibilidade (ou não) da presciência divina com a liberdade humana. Um matemático investigará o livre-arbítrio à luz das verdades “eternas” da lógica (como essas verdades podem existir, se o ser humano é capaz de alterar a realidade?). Os filósofos, que deveriam abordar a questão da maneira mais abrangente

possível, se dedicam, com muita insistência, ao exame do determinismo causal e de sua compatibilidade com o livre-arbítrio. Por que ignoram os demais sentidos de determinismo?

Essa preferência dos filósofos pelo “problema do determinismo causal” torna o debate sobre o livre-arbítrio um tanto maçante e repetitivo. Perde-se, ademais, a oportunidade de discutir questões assaz interessantes.

Teólogos, por exemplo, questionam se Deus pode ou não possuir livre-arbítrio. À primeira vista, essa questão se afigura tão relevante quanto a discussão sobre o “sexo dos anjos” [este foi um debate real havido em Constantinopla no século XV, às vésperas da tomada da cidade pelos turco-otomanos, mobilizando importantes autoridades clericais da época, que não chegaram a um consenso sobre se os anjos possuíam sexo].

Não obstante, por trás da questão de se Deus possui ou não livre-arbítrio, há um questionamento lógico pertinente. Se Deus é necessariamente bom, sua ação é determinada – Deus não pode fazer o mal. Logo, Deus não possui livre-arbítrio. Mas como os seres humanos (criados por Deus e, portanto, mais imperfeitos do que Deus) podem possuir livre-arbítrio, se quem os criou não possui?

Pode-se aceitar que a “necessidade” seja um subconjunto da “liberdade”, mas o contrário não parece razoável. Se, na origem do universo (Deus), havia apenas a “necessidade”, de onde veio a “liberdade” (supondo que exista liberdade)? Num universo criado “livremente”, seria plausível que alguns eventos fossem necessários (afinal, em meio ao *caos*, alguns *padrões* podem ser encontrados). Mas, num universo criado “necessariamente”, como podem existir eventos livres?

Anselmo de Cantuária (1033-1109), na obra *De Libertate Arbitrii* [*On Freedom of Choice*, 2002], procura resolver o problema do livre-arbítrio de Deus propondo a distinção entre *libertas* e *liberum arbitrium*. Por um lado, há uma capacidade que permite ao homem diferenciar o certo do errado, o bom do mau, o verdadeiro do falso etc. A essa capacidade, Anselmo dá o nome de *libertas*. Nesse sentido, o homem pode dizer-se “livre” apenas na medida em que segue uma lei moral divina (i.e., apenas quando faz o que é certo, bom ou verdadeiro). Para Anselmo, apenas Deus (e os anjos) possuem *libertas* no estado puro.

Por outro lado, há a capacidade que permite ao homem escolher entre *pecar* e *não pecar*, que nada mais é do que “a capacidade para decidir de outro modo”, como chamamos

anteriormente. A essa capacidade, Anselmo dá o nome de *liberum arbitrium*. Por óbvio, a Deus (que é um ser perfeito e sumamente bom) não é possível pecar.

De todo modo, de Anselmo até hoje, houve o que considero um “progressivo empobrecimento do debate” sobre o livre-arbítrio: *ad nauseam*, procurou-se argumentar sobre a compatibilidade (ou não) do livre-arbítrio (e da responsabilidade moral) com o “determinismo causal”, em detrimento de outras questões.

Em terceiro lugar, gostaria de me posicionar no debate “compatibilismo x incompatibilismo”, reforçando, porém, que esse debate, a meu juízo, não é nem de longe o mais relevante para o livre-arbítrio.

A princípio, se eu tudo ignorasse sobre o debate, preferiria o termo “compatibilismo” ao termo “incompatibilismo”. Aprendi, com Aristóteles, que o filósofo deve perseguir o “caminho do meio”, e ser “compatibilista” (em qualquer assunto que desperte polêmicas) talvez seja uma boa forma de perseguir aquele caminho.

No entanto, no debate sobre o livre-arbítrio, os termos são enganosos. O compatibilista esposa, a meu ver, posição menos razoável do que o incompatibilista. Explico-o, apelando a Robert Kane. Como sustenta o filósofo estadunidense, “a maior parte das pessoas é naturalmente incompatibilista”. De fato, se pararmos uma pessoa na rua e lhe explicarmos os conceitos de livre-arbítrio e determinismo causal, ela muito provavelmente dirá que essas coisas caminham em sentidos opostos – pois, num caso (livre-arbítrio), nosso futuro está *em aberto*; noutro (determinismo causal), nosso futuro está *fechado*.

Penso, pois, que, para ser compatibilista, é preciso treinamento filosófico, é necessário que se negue essa impressão natural, imediata, que os seres humanos têm do livre-arbítrio e da causalidade física, segundo a qual uma coisa está em desarmonia com a outra, em nome de uma “satisfação cognitiva”.

Essa satisfação perseguida pelos compatibilistas é resultado da conjunção de duas crenças desejáveis, mas excludentes (em minha visão). Se o universo fosse apenas uma sucessão de causas e efeitos, a humanidade poderia, algum dia, ter conhecimento sobre tudo (e possuir essa crença é algo bom). Por outro lado, a crença de que eu sou livre também é prazerosa (mas, para eu ser livre, o universo não pode ser apenas uma sucessão de causas e efeitos – a minha vontade precisa ser *espontânea*, e não *causada*).

Nesse ponto, o compatibilista poderia argumentar: “há uma razão para eu crer, ao mesmo tempo, no livre-arbítrio e no determinismo causal. E se for causalmente necessário que eu seja livre?”

Ora, se for causalmente necessário que eu seja livre, isso significará que a necessidade é anterior à liberdade, isto é, que a necessidade *causa* a liberdade, o que não me parece razoável. Do necessário só se chega ao necessário. Do livre é que se chega ao livre ou ao necessário (pois a “necessidade” é uma das possibilidades da “liberdade”, como argumentado no debate sobre o livre-arbítrio de Deus).

Se até hoje levei uma vida completamente desregrada, sem seguir nenhuma rotina, dormindo e acordando cada dia num horário diferente, passando por lugares diferentes todos os dias, nada me impede que, a partir de amanhã, eu me torne um “outro homem”, e passe a viver tal como Kant em Königsberg. Com efeito, eu posso renunciar à minha “liberdade” inicial e adotar, doravante, uma vida de “necessidades”.

O contrário, porém, não seria possível. Se eu já tivesse nascido “programado” para deitar-me todos os dias às 22h e levantar-me às 6h (i.e., se eu fosse um *robô*), não seria possível, a partir de amanhã, alterar minha rotina, pois, a partir da “necessidade”, não se pode chegar à “liberdade”.

Assim, penso que os incompatibilistas sustentam posição mais moderada no debate sobre o livre-arbítrio, e por isso me incluo entre os que consideram que “livre-arbítrio” e “determinismo causal” caminham em sentidos opostos.

A rigor, de minha parte, das duas “crenças desejáveis” (acreditar que, em tese, a humanidade pode conhecer tudo; e acreditar que eu sou livre), não acredito piamente em nenhuma. Nem acho que as regularidades da natureza permitam à humanidade chegar, algum dia, à “Teoria de Tudo” (pelo que poderia ser chamado de *cético*), nem acredito que os seres humanos sejam absolutamente livres, no sentido de terem pleno controle sobre suas ações (embora, algum controle, eu presumo que nós tenhamos).

BLOCO III: RESPONSABILIDADE MORAL

Responsabilidade moral

O principal verbete da *Stanford Encyclopedia of Philosophy* sobre o tema da responsabilidade moral é de autoria de Matthew Talbert. Já no início do verbete, o filósofo aponta para o que está pressuposto quando julgamos alguém moralmente responsável por seus atos: numa primeira aproximação, estamos reconhecendo que essa pessoa possui certos “poderes e capacidades”.

A maior parte da tradição filosófica associará esses poderes e capacidades a uma forma particular de controle: o controle que permite à pessoa “decidir de outro modo”, i.e., a não escolher o que efetivamente escolhe. No artigo, Talbert empenha-se em mostrar que a responsabilidade moral não precisa pressupor esse tipo de controle, associado ao livre-arbítrio e ao problema do determinismo causal.

Sejam quais forem suas condicionantes, a responsabilidade moral é normalmente atribuída apenas a seres humanos adultos mentalmente sadios. Animais não humanos, crianças muito novas e pessoas com sérios comprometimentos mentais não costumam ser moralmente responsabilizados por seus atos.

Outro aspecto singular da responsabilidade moral, mencionado por Talbert no verbete, é a relação intrínseca que ela guarda com as noções de “culpa” [*blame*] e “elogio” [*praise*]. A culpa decorre do julgamento de que a pessoa é moralmente responsável por uma ação má ou errada. Muitas vezes, ela antecede uma sanção. O elogio é a resposta a uma ação considerada boa ou correta, sobre a qual se pode imputar responsabilidade moral.

No verbete, Talbert também procura marcar certas distinções. Em primeiro lugar, o autor distingue “responsabilidade moral” de “responsabilidade por deveres e obrigações”. Possuir “deveres e obrigações”, em certo sentido, é algo que enseja responsabilidades à pessoa, mas essas responsabilidades não necessariamente envolvem aspectos morais. A responsabilidade moral, que é o assunto que ocupa Talbert no verbete, diz respeito a um pretenso controle que a pessoa exerce sobre suas ações.

A responsabilidade moral também é distinguida da “responsabilidade causal”. Eventos mais simples, como o fogo da lareira, são “causalmente responsáveis” pela sensação de calor em

uma pessoa pessoa que esteja próxima. Ainda assim, não faria sentido asserir que o fogo da lareira é “moralmente responsável” pela sensação de calor na pessoa que esteja próxima.

Poder-se-ia dizer, indo além do que é afirmado no verbete, que a responsabilidade moral, quando comparada a seu análogo causal, requer condições adicionais para ocorrer (em particular, a condição de haver um agente com controle sobre suas ações).

No ponto alto do verbete, Talbert apresenta três abordagens sobre a responsabilidade moral que não pressupõem o livre-arbítrio enquanto “capacidade para decidir de outro modo”. A primeira abordagem apresentada é a “prospectiva” [*forward-looking approach*], que o autor atribui principalmente a Moritz Schlick (1882-1936), filósofo e físico alemão, fundador do positivismo lógico e do Círculo de Viena.

Em linhas gerais, a abordagem prospectiva é aquela que justifica a responsabilidade moral apelando às consequências benéficas que uma pessoa pode obter ao engajar-se em determinadas práticas. Nessa abordagem, está muito presente a noção de “incentivo”, que pode ser positivo ou negativo, e que tem o condão de encorajar a pessoa a fazer o que é considerado certo e a desincentivá-la a fazer o que é considerado errado. Para os adeptos da abordagem prospectiva, uma pessoa pode ser moralmente responsabilizada na medida em que suas escolhas e seu comportamento podem ser moldados pela lógica de recompensas e punições. Como se nota, nenhuma palavra é dita sobre determinismo causal ou livre-arbítrio.

Como abordagem alternativa, Talbert cita a “concepção das atitudes reativas” de P.F. Strawson [*reactive attitudes approach*]. A abordagem de Strawson tampouco apela à Metafísica para justificar a responsabilidade moral, que teria origem nas próprias relações interpessoais e nas emoções que elas suscitam. Para essa abordagem, uma pessoa pode ser moralmente responsabilizada na medida em que vive em sociedade, em que é afetada pelos julgamentos dos outros e em que age reativamente.

A terceira abordagem mencionada por Talbert é a de “resposta às razões” [*reasons-responsive approach*]. John Martin Fischer e Mark Ravizza são os principais expoentes dessa abordagem, segundo Talbert. A dupla distingue o “controle regulador” [*regulative control*] do “controle de orientação” [*guidance control*]. O controle regulador consubstancia-se no poder para decidir de outro modo. O controle de orientação, ao revés, apenas indica que o “agente guia seu comportamento numa direção particular”, ainda que não possa guiá-lo numa direção diferente.

Fischer e Ravizza comprometem-se com a existência desse último tipo de controle, suficiente para assegurar, na visão dos filósofos, a responsabilidade moral.

No verbete, são ainda apresentadas três concepções contemporâneas de responsabilidade moral. São elas: a *attributability-responsibility*, para a qual a origem da responsabilidade moral está no caráter do agente; a *accountability-responsibility*, que enfatiza o papel da sociedade na constituição da responsabilidade moral; e a *answerability-responsibility*, que aloca a responsabilidade moral nos julgamentos avaliativos do agente.

Talbert encerra o verbete com considerações sobre a “competência moral”, isto é, a “habilidade para reconhecer e responder a considerações morais”. Trata-se de competência não universal – isso é amplamente aceito. Em Filosofia, a figura do “psicopata”, por exemplo, é tipicamente caracterizada como “inteiramente incapaz de responder a considerações morais”.

Os céticos da responsabilidade moral são justamente aqueles que extrapolam esses casos particulares para defenderem que a responsabilidade moral não é possível em circunstância nenhuma.

Fatos sociais

Paul Gingrich, professor emérito da Universidade de Regina (Canadá), disponibiliza um rico acervo de notas de aula em seu sítio eletrônico profissional. Em uma de suas aulas do curso de outono de 1999, Gingrich abordou o tema “fatos sociais e suicídio”. As notas iniciam-se com a definição que o sociólogo francês Émile Durkheim (1858-1917) oferece para “fatos sociais”: são “coisas externas e coercitivas” ao agente – na paráfrase de Jeffrey Hadden (1937-2003), são fatos “criados por forças coletivas e que não emanam do indivíduo”.

Para Durkheim, as crenças religiosas são exemplos de “fatos sociais”, porquanto impõem uma conduta ao indivíduo, independentemente de sua vontade. Enfatize-se que fatos sociais não se deduzem da razão pura ou do pensamento. Eles demandam estudos empíricos, e não filosóficos, como ensina George Ritzer.

Hadden, enfatizando a diferença entre o estudo dos fatos sociais e a Filosofia, observa que aqueles são expressamente “manifestos em indicadores externos de sentimentos como as doutrinas religiosas, as leis e os códigos morais”.

Nas palavras do professor canadense, fatos sociais “existem como resultado de interações sociais e desenvolvimentos históricos (...) [eles] são aprendidos (através da socialização) e geralmente aceitos, mas o indivíduo não tem nenhum poder para estabelecê-los”.

Note-se, aqui, que fatos sociais constituem, por assim dizer, o “reverso da medalha” do livre-arbítrio. Num caso (livre-arbítrio), os atos emanam da vontade do indivíduo. Noutro caso (fatos sociais), os atos são alheios a sua vontade – não se descarta, é evidente, a hipótese de o indivíduo não ter consciência do fato social e incorporá-lo de forma tão “natural” a ponto de considerá-lo fruto de sua vontade.

Parecem subsistir, na Sociologia, as crenças de que a sociedade é mais do que a soma dos indivíduos e de que “fatos novos” surgem no nível social, merecendo análises próprias. No entanto, os fatos sociais muitas vezes estão na mente dos indivíduos, e “são também estados mentais”. Cumpre reforçar, então, a diferença entre “fatos sociais” e “fatos psicológicos”.

A esse respeito, Ritzer nota, nas expressões de Gingrich, que “os fatos sociais podem ser considerados fenômenos mentais que são externos e coercitivos aos fatos psicológicos, tais como os instintos humanos”. Em outras palavras, fatos sociais são mais do que fatos psicológicos, na medida em que surgem externamente e condicionam a mente do indivíduo, trabalhando sobre seus instintos.

Gingrich, numa frase inspirada, observa que “os fatos sociais regulam a ação social humana e atuam como restrições sobre o comportamento e a ação individual”. Antes de iniciar suas considerações sobre o suicídio (que não serão expostas aqui), o professor alude a duas distinções. Em primeiro lugar, os “fatos sociais materiais”, tais como as estruturas e as instituições sociais, que apresentam realidade material, são diferenciados dos “fatos sociais não materiais”, vale dizer, das normas, dos valores e dos sistemas de moralidade, sem existência material.

Em segundo lugar, os “fatos sociais normais” são discernidos dos “fatos sociais patológicos”. Enquanto aqueles constituem os fatos sociais “mais amplamente distribuídos e úteis, contribuindo para a manutenção da sociedade e da vida social”, estes “podem ser associados a problemas sociais e doenças de vários tipos”.

Enfatize-se, quanto aos fatos sociais, que eles podem ter aplicabilidade apenas local, isto é, “pode haver diferentes normas e valores para diferentes grupos da sociedade”. Com efeito, em seus trabalhos finais, Durkheim abjurou a doutrina da “consciência coletiva única”, sem prejuízo para a tese de que os “fatos sociais” não são nunca redutíveis à consciência individual.

Diferenças entre “necessidade” e “suficiência”

Viceja, na tradição filosófica anglófona dos últimos cem anos, uma constante preocupação com a busca de “condições necessárias e suficientes” para a definição precisa de termos. Um “filósofo analítico” provavelmente já terá ouvido expressões do tipo “ser solteiro é condição necessária e suficiente para não ser casado”, ou “estar livre é condição necessária e suficiente para não estar preso”.

Talvez com menos entusiasmo, fala-se também, na filosofia contemporânea, em “condições necessárias, mas não suficientes”. Essa definição tem o mérito de capturar uma diferença entre “necessidade” e “suficiência”. Com efeito, nem tudo o que é “necessário” é também “suficiente”. Como exemplo, pode-se dizer que “ser humano é condição necessária para ser filósofo, mas não é condição suficiente”, ou “possuir 35 anos ou mais é condição necessária para ser Presidente do Brasil, mas não é condição suficiente”.

No entanto, é de causar estranheza o fato de os filósofos analíticos raramente explorarem uma terceira possibilidade lógica, a saber, a das “condições suficientes, mas não necessárias”. Essa definição também capturaria uma diferença importante entre “necessidade” e “suficiência”, pois nem tudo o que é “suficiente” é “necessário”. À guisa de ilustração, tome-se a frase “escrever uma tese sobre o livre-arbítrio é condição suficiente para obter o título de ‘doutor em Filosofia’, mas não é condição necessária” (de fato, muitas pessoas obtêm o mesmo título escrevendo sobre temas distintos).

“Necessidade” e “suficiência” são, portanto, conceitos em princípio independentes, mas que podem ser compatíveis. Do ponto de vista matemático, diz-se que são “duais” em relação ao outro. Isso significa que, dadas duas sentenças [A e B], se A é necessário para B, B é suficiente para A.

A imagem a seguir, retirada do verbete *Necessity and Sufficiency* da *Wikipedia* (de onde, de resto, retiramos as considerações do parágrafo anterior), ilustra as diferenças entre “necessidade” e “suficiência”:

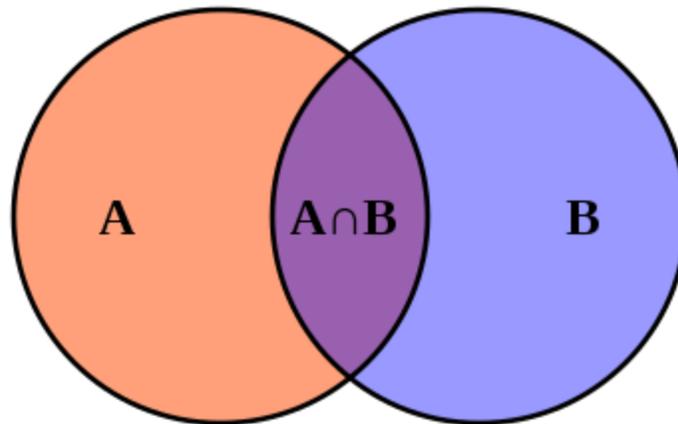


Figura 1: Estar na zona roxa (centro) é condição suficiente para ser membro do conjunto “A”, mas não é condição necessária [com efeito, os pontos na zona marrom (esquerda) também são membros do conjunto “A”, mesmo não estando na zona roxa (centro)]. Ser membro do conjunto “A” é condição necessária para estar na zona roxa (centro), mas não é condição suficiente [com efeito, todos os membros da zona roxa (centro) pertencem ao conjunto “A” (e também ao conjunto “B”), mas pertencer ao conjunto A não é suficiente para estar na zona roxa (centro), como o revela a zona marrom (esquerda)]. Ser membro dos conjuntos “A” e “B” é condição necessária e suficiente para estar na zona roxa (centro). Estas explicações seguem, com adaptações, as observações em inglês que podem ser encontradas abaixo da mesma imagem no já mencionado verbete da *Wikipedia*.

Comentários finais (Bloco III)

Há dois comentários que gostaria de fazer acerca deste bloco temático, ambos referentes à seção “Fatos sociais”.

Em primeiro lugar, parece-me interessante enfatizar a distinção que propus entre “fatos sociais” e “livre-arbítrio”. Não se trata de uma distinção convencional, decerto. Mas, ainda assim, penso que se trata de uma distinção pertinente: se os “fatos sociais” são aqueles que independem da vontade do indivíduo, e as manifestações de “livre-arbítrio” são, por definição, aquelas decorrentes da “vontade espontânea” do indivíduo, nada mais natural do que tratar um conceito como o “reverso da medalha” do outro (é verdade que, além dos fatos sociais, há também fatos físicos, como a gravidade, capazes de constranger a vontade humana).

O oposto de agir “livremente” seria, assim, agir de acordo com normas que não reconheço, mas que predominam no meu meio social. Há, é evidente, a hipótese de os fatos sociais não soarem

tão estranhos a mim (i.e., a hipótese de eu “introjetar”, com certa naturalidade, os fatos sociais). Nesse caso, o contraste entre livre-arbítrio e fatos sociais não será tão evidente.

É sobre isso que quero falar em segundo lugar. Os fatos sociais, em alguma medida, são também fatos psicológicos – isso significa que, em algum momento, nós “interiorizamos” os fatos sociais. Como se dá esse processo de “interiorização”? Tem-se aqui uma boa pergunta a ser dirigida a neurocientistas.

De todo modo, a mim não deixa de parecer extraordinário que fatos gerados externamente tenham essa capacidade de se “introjetar” em mim, sendo inclusive capazes de guiar as minhas ações, a ponto de, em alguns casos, ser impossível distingui-los da minha própria vontade espontânea. Esse raciocínio, se levado às últimas consequências, pode comprometer seriamente a noção de “indivíduo”.

De resto, cumpre reforçar aquilo que já havia antecipado no bloco temático anterior. A responsabilidade moral, a meu juízo, é apenas uma “chave” secundária para a compreensão do livre-arbítrio.

As noções de determinismo causal e responsabilidade moral são muito discutidas em Filosofia. Para mim, o que há de mais relevante no tema do livre-arbítrio é o que será abordado na sequência da tese.

BLOCO IV: O PROBLEMA MENTE-CÉREBRO

O problema mente-cérebro

O filósofo britânico Stephen Priest é o autor do verbete “interação mente-cérebro” [*mind-brain interaction*], da *Encyclopedia of Science and Religion*. No verbete, Priest introduz o problema mente-cérebro enfatizando a aparente inconsistência entre a suposição do senso comum de que eventos mentais causam eventos físicos e eventos físicos causam eventos mentais, e a suposição científica de que o universo físico é um sistema determinista fechado.

Se o universo físico for, de fato, um sistema determinista fechado, todo evento físico terá como causa eventos físicos prévios, suficientes para a sua ocorrência. A ideia de causação mental será, então, redundante. Priest reforça esse ponto mencionando a Primeira Lei da Termodinâmica, que estabelece que a quantidade de energia no universo é constante. Ou bem a causação mental é redundante, ou bem ela implica uma adição de energia ao universo, ferindo a Primeira Lei da Termodinâmica.

Por outro lado, o filósofo britânico reconhece que não seria trivial explicar, de forma satisfatória, como a consciência é produzida a partir do cérebro, supondo ser esse o caso. Em última instância, o cérebro é um arranjo de “bilhões e bilhões de átomos”, e é de causar perplexidade, a nós humanos, a hipótese de nossa consciência não ser mais do que o resultado desse arranjo.

Levando em conta as permutações lógicas, Priest alega serem quatro as possibilidades de interação mente-cérebro: (i) a mente afeta o cérebro, mas o cérebro não afeta a mente; (ii) o cérebro afeta a mente, mas a mente não afeta o cérebro; (iii) o cérebro afeta a mente, e a mente afeta o cérebro; (iv) o cérebro não afeta a mente, e a mente não afeta o cérebro.

O autor do verbete apresenta brevemente, então, os posicionamentos de grandes nomes da Filosofia a respeito do tema, começando pelo filósofo pré-socrático Anaxágoras, a quem é atribuída a frase “a mente causa todas as coisas”. Na sequência, Priest menciona Aristóteles, que, em sua “Ética a Nicômaco”, afirma ser a escolha (mental) a causa eficiente da ação (física).

Priest também cita René Descartes, defensor da interação, nos dois sentidos, da alma imaterial (mente) com o corpo humano (cérebro); e Gottfried Leibniz, adepto da tese do “paralelismo psicofísico”, que nega a interação mente-cérebro em ambos os sentidos. O autor do

verbete alude ainda a Baruch Spinoza, que trata pensamento (mente) e extensão (cérebro) como atributos de uma mesma substância.

O debate contemporâneo sobre a interação mente-cérebro é apresentado à parte. Destaque é dado a Donald Davidson, cujo “monismo anômalo” é reputado, por Priest, como a teoria de causalção mental mais influente do último quarto do século passado.

Em linhas gerais, o monismo anômalo afirma que eventos mentais são “eventos físicos anômalos”, não descritíveis por leis. Eventos mentais causam eventos físicos pela simples razão de serem, eles também, físicos. No entanto, a existência de leis psicofísicas, que explicariam esse mecanismo de causalção mental, é rejeitada.

Priest extrai uma consequência da posição de Davidson: a tese de que inexitem leis psicofísicas estritamente deterministas implica a tese de que não há explicação física para o mental. Abre-se, assim, um espaço para o livre-arbítrio.

O autor do verbete informa que o monismo anômalo pode também ser caracterizado como uma “teoria da superveniência”. *Grosso modo*, a teoria da superveniência afirma que o mental sobrevém ao físico. São múltiplos os sentidos dessa afirmativa, mas o primeiro mencionado por Priest no verbete é o seguinte: se dois eventos mentais diferem, eles também devem diferir em algum aspecto físico. Por consequência, uma pessoa não pode “abandonar” um estado mental sem “abandonar” um estado físico, ou, mais genericamente, não pode alterar seus aspectos mentais sem alterar seus aspectos físicos.

A teoria da superveniência, quando aplicada apenas ao mundo real, é caracterizada por Priest como “fraca”; quando estendida a mundos possíveis não existentes, é reputada por Priest como “forte”. De um modo ou de outro, mais do que em evidências científicas, a superveniência estaria ancorada em intuições: a intuição de que eventos mentais dependem de eventos físicos, mas eventos físicos não dependem de eventos mentais; e a intuição de que a mente não é redutível ao cérebro, embora dependa dele. Em que pese essas intuições poderem estar equivocadas, para Priest, a teoria da superveniência tem o mérito de capturá-las.

O verbete se encerra com considerações sobre as teorias epifenomenistas contemporâneas, que reagem à teoria da superveniência. Em primeiro lugar, Priest apresenta a objeção do filósofo canadense Ted Honderich à posição de Davidson: numa leitura mais atenta do “monismo anômalo”, os eventos mentais não causariam eventos físicos “em virtude de serem” mentais, mas “em virtude de serem” físicos. As propriedades mentais a que Davidson faz referência não

desempenhariam nenhum papel causal, e sua “teoria da causação mental” seria, em última análise, falsa.

O epifenomenismo é apresentado como a tese de que eventos mentais são causados por eventos físicos, mas eventos físicos não são causados por eventos mentais. Essa posição aceita a existência de leis psicofísicas, ainda a serem descobertas. Do ponto de vista empírico, a Priest parece incontroverso que o neural seja necessário para o mental, ao menos no caso humano – o autor lembra que lesões neurais costumam acarretar comprometimentos mentais. Do ponto de vista lógico, Priest enfatiza que, se A é necessário para B, B é suficiente para A. Logo, se o neural é necessário para o mental, o mental é suficiente para o neural.

Por fim, em resposta ao epifenomenismo, são apresentadas as posições de Karl Popper (1902-93), para quem o mental, embora seja um produto evolucionário do cérebro, tem poderes causais sobre ele; e de filósofos que rejeitam a premissa de que o universo físico é um sistema determinista fechado, como os ingleses Tim Crane e D.H. Mellor. Pouca atenção, no entanto, é dedicada a essas posições.

Intuições dualistas na escrita científica: a crítica de Mudrik e Maoz

Em artigo publicado na revista *Journal of Cognitive Neuroscience* (2014), Liad Mudrik e Uri Maoz conceitualizam a “falácia do sujeito duplo” [*double-subject fallacy (DSF)*]: o raciocínio que atribui divergentes (ou mesmo opostos) estados psicológicos ao cérebro e ao sujeito. Essa distinção entre “mim” e “o meu cérebro” [*me x my brain*] estaria presente não apenas em livros de divulgação de neurociência, dirigidos ao público leigo, mas também em artigos científicos, revisados por pares.

Na visão dos autores, os textos neurocientíficos estão contaminados por intuições dualistas, ainda que sejam poucos os pesquisadores na área que endossem abertamente o dualismo cartesiano. A neurociência teria promovido uma revolução nos estudos da relação mente-cérebro, por meio da “neuralização da mente” [*neuralization of the mind*], mas intuições cotidianas, assentadas em visões dualistas, ainda influenciariam a forma de escrita de muitos neurocientistas.

A escrita baseada na DSF não seria meramente metafórica. Metáforas, via de regra, estabelecem um paralelo implícito entre termos utilizados no sentido denotativo (literal) e termos utilizados no sentido conotativo (não literal). Na metáfora “o tempo é um rio”, por exemplo, a

palavra “tempo” é utilizada no sentido literal, e a palavra “rio”, no sentido conotativo, para designar “fluidez”. Mas em frases do tipo “parece que o cérebro sabe de nossas decisões antes de nós” (Mudrik e Maoz, p.145, tradução nossa), tanto “cérebro” como “nós” seriam, na visão de Mudrik e Maoz, utilizados no sentido denotativo.

A escrita baseada na DSF tampouco seria inofensiva. Mesmo que a opção por “personificar” o cérebro e tratá-lo como substância independente do “eu” fosse meramente estilística ou metafórica, e mesmo que metáforas pudessem ser úteis às explicações científicas [a esse respeito, os autores mencionam um livro de Theodore Brown (cf. Referências Bibliográficas)], no caso da neurociência, a “falácia do sujeito duplo” aludiria a uma intuição dualista que, depreende-se de Mudrik e Maoz, precisa ser superada.

A neurociência atravessaria um período de transição, em que intuições dualistas são substituídas por intuições materialistas. Para Mudrik e Maoz, enquanto essa transição não estiver consolidada, será preciso tomar ciência da escrita baseada na DSF e evitá-la ao máximo, tanto mais porque “o pensamento dualista é ao mesmo tempo errado e socialmente perigoso” [os autores mencionam um livro de Richard Dawkins e um artigo de Joshua Greene para endossar o último ponto (cf. Referências Bibliográficas)].

As motivações para a visão de mundo dualista, sustentada de forma implícita (e mesmo inconsciente) por pesquisadores da neurociência, se encontrariam na própria experiência humana. Os autores citam Chris Frith, que assume “parecer às vezes um dualista”, distinguindo o “eu” do “meu cérebro”, pois “assim é a experiência” (Frith, p. 23, tradução nossa). A escrita baseada na DSF resultaria da conjunção desse “dualismo de armário” [*closet dualism*] de certos neurocientistas com a busca pela compreensão das relações mente-cérebro.

Anexo a seu artigo, Mudrik e Maoz apresentam uma tabela em que são propostas reescritas de textos baseados na DSF. A título de ilustração, uma frase de Libet et al. (1983) é assim reelaborada:

“... o cérebro evidentemente 'decide' iniciar ou, pelo menos, se prepara para iniciar o ato antes de haver qualquer consciência subjetiva relatável de que tal decisão tenha ocorrido” (Libet et al., p. 640, tradução nossa) → “... processos cerebrais inconscientes evidentemente iniciam ou, pelo menos, se preparam para iniciar o ato antes de haver qualquer consciência subjetiva relatável de que tal decisão tenha ocorrido.” (Mudrik e Maoz, p. 217, tradução nossa).

Com alterações desse tipo, Mudrik e Maoz acreditam colaborar para o desenvolvimento de uma neurociência menos presa a intuições equivocadas.

O eterno retorno do materialismo: a crítica de Saulo Araújo

Em artigo publicado na Revista de Psiquiatria Clínica (2013), Saulo Araújo constata que o avanço das tecnologias de neuroimagem nas últimas décadas motivou a proliferação de explicações materialistas para os fenômenos mentais. Com efeito, aos anos 1990, chegou a ser proposto o epíteto de “a década do cérebro”. Versões mais arcaicas do materialismo, no entanto, como o materialismo francês do século XVIII e o materialismo alemão do século XIX, já antecipavam muitas das considerações dos neurocientistas contemporâneos.

Na visão de Saulo Araújo, há presentemente uma tendência de se considerar o materialismo como uma consequência inescapável das investigações científicas. Desde o início de seu texto, o autor advoga, ao contrário, pela independência entre materialismo e ciência. A neurociência contemporânea estaria a eliminar a autonomia da experiência subjetiva e a prometer, em troca, um “novo futuro”, em que mistérios da mente seriam elucidados por explicações fisiológicas. Concretamente, a realidade apenas exibiria, para Araújo, o “eterno retorno do materialismo”.

Remontam ao século XVIII, na França, e ao século XIX, na Alemanha, o “entusiasmo epistêmico” e o “otimismo explícito” com o materialismo, bem como a insatisfação com a linguagem psicológica tradicional, fundada em “ilusões” dualistas e religiosas. Carl Vogt, mencionado por Araújo em seu artigo, o ilustra: “é inegável que a sede da consciência, da vontade, do pensamento deve ser enfim procurada unicamente no cérebro” (Vogt, p.17, tradução de Saulo Araújo).

Tampouco é nova a estratégia de se apelar a um futuro desconhecido em favor da posição materialista – de forma talvez espirituosa, Karl Popper (1902-1994) qualificaria o materialismo como posição “promissória”.

À semelhança de Mudrik e Maoz, Saulo Araújo, mesmo que sob uma perspectiva não materialista, critica a linguagem metafórica, “quase ficcional”, de que têm se valido neurocientistas para se referirem ao cérebro, atribuindo a este (parte) “capacidades e realizações” de seres humanos (todo).

A abstração do sujeito e o relato em terceira pessoa de experiências subjetivas em primeira pessoa, presentes em textos neurocientíficos, seriam, para Araújo, faces de uma mesma moeda: a moeda acrítica e ideológica de materialistas, sem verdadeiro lastro na ciência.

O “eterno retorno do materialismo” seria, em suma, representação do “eterno esquecimento dos limites do conhecimento”. Ao defenderem o materialismo como a consequência inescapável dos avanços da ciência, materialistas endossariam visão utópica e ilusória. Utópica, porque aponta para um “futuro promissor”, edênico, em que o problema mente-cérebro seria finalmente esclarecido, em benefício da humanidade. Ilusória, porque desconsidera as limitações do conhecimento humano – já se fala desse “futuro promissor” há séculos, sem que nunca se chegue a ele.

Inversamente, seria preciso lembrar que a ciência demanda exame crítico da realidade.

Um modelo quântico para o interacionismo cartesiano

O dualismo interacionista tem sua versão mais conhecida na filosofia de René Descartes (1596-1650), que chega a apontar o local onde a interação mente-cérebro se daria: a glândula pineal – uma pequena glândula no epitélamo (centro do cérebro) que, pelo que se sabe hoje, possui função meramente endócrina.

Embora simples e intuitiva, a explicação cartesiana é muitas vezes submetida a críticas de implausibilidade: o seu modelo supostamente não respeitaria as leis de conservação de energia e de momentos linear e angular da Física.

Cientes disso, Osvaldo Pessoa Jr. e Luma Melo (2015) desenvolveram um modelo quântico para a glândula interacionista. A dupla considerou um aniquilamento de duas partículas (um elétron e^- e um pósitron e^+), e a consequente geração de um par de raios gama (γ).

Sendo simétrica a configuração inicial do elétron (e^-) e do pósitron (e^+), a soma dos seus momentos seria nula. A quantidade de momento dos raios gama (γ) resultantes também seria nula, dado que estes rumariam em sentidos opostos. Logo, entre a configuração inicial e a final, haveria conservação de energia e de momentos linear e angular.

Não seria possível saber previamente, entretanto, a direção exata em que os raios gama rumariam (apesar de se saber que eles rumariam em sentidos opostos).

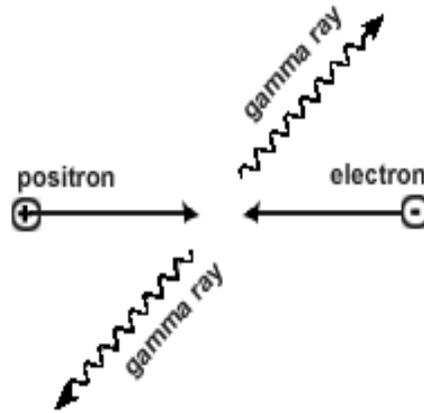


Figura 2: Diagrama representativo de um aniquilamento de partículas (pósitron e elétron) e consequente formação de raios gama. Observe-se que, aqui, os raios gama se formaram na direção nordeste-sudoeste. Não seria possível conhecer essa direção antes do aniquilamento das partículas.

Os autores propuseram, então, a seguinte analogia. Se o pósitron (e^+) representasse uma percepção oriunda do ambiente externo (por exemplo, uma instrução para o indivíduo flexionar um dos dedos indicadores, em um experimento de neurociência, assim que sentisse vontade), e o elétron (e^-) representasse as predisposições corporais internas do indivíduo (por exemplo, uma disposição do indivíduo para contribuir com o experimento de neurociência), o par de raios gama (γ) resultante do aniquilamento poderia representar a escolha do indivíduo (por exemplo, flexionar o dedo indicador direito).

Na analogia proposta, a escolha do indivíduo seria imprevisível até o instante em que a “consciência” (ou o “espírito”, para falar como Descartes) determinasse em quais quadrantes os raios rumariam. Assim, a detecção dos raios gama (γ) no primeiro e no terceiro quadrantes da figura acima seria atribuída, pelo modelo interacionista, à ação direta do espírito sobre a glândula pineal.

Fundamentalmente, o que o modelo de aniquilamento de partículas mostra é que essa interpretação interacionista, ainda que excêntrica para os padrões científicos atuais, não viola nenhuma lei de conservação da Física. Não há, portanto, como provar, de forma definitiva, que ela está errada, apelando-se apenas a leis de conservação de energia.

John Eccles e os “campos de influência” da mente sobre o cérebro

O neurofisiologista australiano John Eccles (1903-97), prêmio Nobel de Medicina em 1963, desenvolveu uma hipótese neurofisiológica sobre a interação mente-cérebro em livro publicado em 1953 pela *Oxford Press*. Dois anos antes, Eccles publicara um artigo na revista *Nature* sobre o mesmo tema. Conjeturar o modo pelo qual a mente alcança ligação com o cérebro é objetivo de ambas as publicações. A hipótese central de Eccles é de que a mente exerceria “campos de influência” sobre o cérebro, que seriam detectados pelo córtex cerebral quando este se encontrasse em “nível alto de atividade”. Esses “campos de influência” alterariam os padrões espaço-temporais da atividade neural.

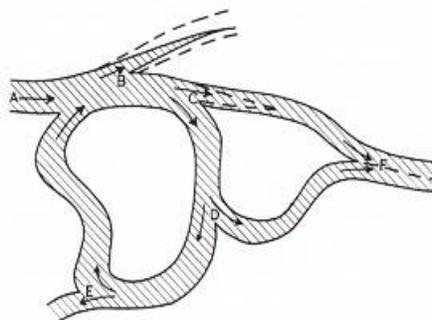


Figura 3: Diagrama representativo de um padrão espaço-temporal de uma atividade neural, utilizado por Eccles em 1953.

Na figura acima, que representa, de forma arbitrária, um padrão espaço-temporal de uma atividade neural, as setas indicam a direção de propagação dos conjuntos neurais, e as linhas quebradas, os “campos de influência” da mente sobre o cérebro – percebe-se que elas alteram os caminhos originais.

Desconsiderando as linhas quebradas, vale dizer, não levando em conta a atuação da mente sobre o cérebro, é possível observar (i) que os conjuntos neurais vindos de A encontram ramificação em B, e que um dos caminhos de B é abortivo; (ii) que, em F, os conjuntos neurais vindos de C e D unem-se, reforçando-se (fenômeno denominado “somatório sináptico” por Eccles); (iii) que a ramificação em E pode levar a dois caminhos – um deles, o caminho de volta a B, que implicaria o fechamento da cadeia.

Considerando as linhas quebradas, vale dizer, levando em conta a atuação da mente sobre o cérebro, constata-se (i) que o antigo caminho abortivo depois de B deixa de sê-lo; (ii) que o

antigo caminho depois de C que levava a F passa a ser abortivo; (iii) que não há mais “somatório sináptico” em F – os conjuntos neurais em F vindos pelo caminho D não são mais reforçados pelos conjuntos neurais vindos pelo caminho C.

Em que pese o diagrama de Eccles ter o mérito de exibir a plasticidade cerebral, i.e., a capacidade de o sistema nervoso central modificar sua organização estrutural e funcional em resposta a estímulos ambientais (nisso reside a essência do processo de aprendizagem), e poder suscitar, em pesquisadores materialistas, a busca pelo correlato físico dos “campos de influência” da mente, deve-se reconhecer que, nele, a relação mente-cérebro é mais postulada do que explicada.

O próprio Eccles o admite: “É evidente que as hipóteses aqui desenvolvidas têm caráter fragmentário e provisório (...) estas postuladas ‘influências da mente’ ainda não foram detectadas por nenhum instrumento físico existente” (Eccles, 1953, p. 283, tradução nossa).

A complementaridade mente-cérebro

Em artigo de 1969 publicado em *The Journal of Philosophy*, Nathan Brody e Paul Oppenheim (químico e filósofo alemão, colega de caminhada de Albert Einstein nas manhãs de domingo) propõem uma aplicação do “princípio da complementaridade” de Niels Bohr (1885-1962) ao problema mente-cérebro. Bohr, lembre-se, foi o célebre físico dinamarquês que contribuiu decisivamente para o entendimento da estrutura atômica e para o desenvolvimento da Física Quântica, tendo sido agraciado com o Prêmio Nobel de Física em 1922.

Para explicar a complementaridade na Física, Brody e Oppenheim partem da constatação de que princípios fundamentais da Física Clássica parecem não se aplicar à Física Quântica. A Física Clássica estuda os objetos de nível macro. Em teoria, é possível saber simultaneamente a posição precisa e o momento exato desses objetos. No entanto, de acordo com o “princípio da indeterminação” de Heisenberg, o mesmo não se aplica aos objetos de nível micro, estudados pela Física Quântica: não é possível saber simultaneamente a posição precisa e o momento exato de objetos microscópicos.

Os autores aludem, em seguida, à chamada “dualidade entre matéria e luz”. Para a Física Clássica, um objeto de nível macro “corpuscular” não poderá, em hipótese nenhuma, tornar-se em algum momento ulterior “ondulatório”. Por óbvio, o inverso também se aplica. Em outras palavras,

para a Física Clássica, “matéria” e “luz” são noções mutuamente excludentes, e as propriedades “corpuscular” e “ondulatória” são incompatíveis.

Na Física Quântica, a questão é mais sutil. De acordo com a teoria de Bohr, um mesmo objeto de nível micro pode possuir tanto a *aparência* “corpuscular” quanto a “ondulatória”, embora isso nunca ocorra simultaneamente (mesmo na Física Quântica, um dado objeto jamais será, *ao mesmo tempo*, corpuscular e ondulatório). O arranjo experimental utilizado pelo físico para medir o objeto quântico é que determinará *se ele se comporta*, num dado instante, como matéria ou como luz (lembre-se que a posição de Bohr é antirrealista, pelo que juízos categóricos do tipo “o objeto é uma partícula” ou “o objeto é uma onda” devem ser evitados). Nas expressões de Brody e Oppenheim, na Física Quântica (interpretação de Bohr), as propriedades “corpuscular” e “ondulatória”, ao invés de incompatíveis, são apenas *não compatíveis*.

Com efeito, a unidade conceitual da Física não é violada pela interpretação de Bohr. No caso quântico, as propriedades “corpuscular” e “ondulatória” não se aplicam apenas ao objeto de nível micro, mas ao *conjunto* formado pelo objeto de nível micro *e* o arranjo experimental utilizado pelo físico para medi-lo. Ou seja: ainda que o objeto de nível micro possa assumir, individualmente e em diferentes momentos, *feições* corpusculares e ondulatórias, quando se fixa um conjunto {objeto de nível micro, arranjo experimental}, será observada, a qualquer tempo e de forma excludente, ou a propriedade corpuscular, ou a propriedade ondulatória, à semelhança do que ocorre na Física Clássica. A “dualidade entre matéria e luz” é, por essa via, preservada.

Bohr sugeriu que seu “princípio da complementaridade” poderia ser aplicado, com igual sucesso, em outras áreas do conhecimento, inclusive na Psicologia. Coube a Brody e Oppenheim, no artigo que nos ocupa, a missão de fazê-lo. A dupla partiu da distinção entre *entidades do tipo M* (mentais) e *entidades do tipo C* (corpóreas) [*bodily (B) entities*, em inglês].

Por *entidades do tipo M*, Brody e Oppenheim entendem aquelas que possuem, basicamente, duas características: (i) permitem acesso privilegiado (exclusivo) ao indivíduo que as experimenta; (ii) não estão localizadas em nenhuma parte do corpo. Quanto ao segundo ponto, os autores chegam a explorar a hipótese de as entidades mentais estarem localizadas no cérebro – de toda forma, para eles, seria irrelevante conhecer essa localização.

As *entidades de tipo C* são aquelas de características opostas: (i) não permitem acesso privilegiado (exclusivo) ao indivíduo – i.e., ou podem ser observadas por mais de uma pessoa, ou não podem ser observadas por ninguém; (ii) estão localizadas em alguma parte do corpo.

Na sequência, os autores apelam a uma segunda distinção, independente da primeira. Diferenciam as *entidades de tipo 1* (referidas pela psicologia do senso comum) das *entidades de tipo 2* (referidas pela ciência). As *entidades de tipo 1* estão tipicamente presentes em relatos de experiências subjetivas e em observações prosaicas (e.g., “eu estou experimentando o medo” e “ele está corado”). O que as caracteriza é o fato de não terem nenhuma pretensão científica. Tanto *entidades de tipo M* quanto *entidades de tipo C* podem ser *entidades de tipo 1*. Em “eu estou experimentando o medo”, faz-se referência a uma *entidade de tipos M e 1*. Em “ele está corado”, faz-se referência a uma *entidade de tipos C e 1*.

As *entidades de tipo 2*, por oposição, são justamente aquelas referidas por sentenças embasadas em alguma teoria científica. Encontram-se em frases como “os conjuntos de células são formados lentamente” e “o sistema límbico controla as emoções”. Brody e Oppenheim argumentam que tanto *entidades de tipo M* quanto *entidades de tipo C* podem ser *entidades de tipo 2*. No entanto, como o revelam os exemplos deste parágrafo, é muito mais fácil encontrar uma *entidade de tipos C e 2* do que uma *entidade de tipos M e 2*.

Embora seja “bastante inusual” conceber uma entidade mental como científica, há literatura que o autoriza. *Entidades de tipos M e 2*, segundo os autores do artigo, seriam abordadas nas teorias de Dulaney e de Snygg e Combs. Infelizmente, não são oferecidas explicações pormenorizadas dessas teorias. Para sorte do leitor, Brody e Oppenheim abordam, na sequência do texto, o exemplo da “fome”, entidade que pode ser mental (M) ou corpórea (C), não científica (1) ou científica (2), dando mais concretude a suas ideias.

Em primeiro lugar, é avaliado o caso em que a fome é uma entidade corpórea (tipo C). De acordo com a teoria glicostática, há dois centros no hipotálamo (parte do diencéfalo situada na base do cérebro) que controlam o comportamento alimentar dos indivíduos: o “centro de saciedade” [*satiety center*] e o “centro de alimentação” [*eating center*]. A fome pode ser concebida como uma relação entre atividades nesses dois centros. Nesse sentido, ela tem uma localização específica no corpo – os núcleos ventromedial e lateral do hipotálamo – e não permite acesso privilegiado (exclusivo) ao indivíduo que a experimenta. Trata-se, em suma, de entidade corpórea (tipo C). Dada a importância da teoria glicostática para essas explicações, a fome também seria, no caso, entidade científica (tipo 2).

Brody e Oppenheim avaliam, em segundo lugar, o caso em que a fome é uma entidade mental (tipo M). De saída, referem-se à “Teoria da Dissonância Cognitiva” (TDS). A “dissonância

cognitiva” é um estado psicológico que o indivíduo deseja superar. Ela está presente, por exemplo, na seguinte conjunção de crenças: (i) “eu sei que fumar é prejudicial à saúde”; e (ii) “eu sei que eu fumo”. Em situações desse tipo, segundo a TDS, o indivíduo tentará eliminar uma de suas crenças, a fim de superar a “dissonância cognitiva”.

Brody e Oppenheim aplicam a TDS ao caso da fome. Os autores aludem ao seguinte experimento: indivíduos são levados a um laboratório, onde lhes é requisitado que fiquem em jejum durante um determinado período. Encerrado esse período, outra requisição lhes é feita: que continuem em jejum por mais um novo período. No momento dessa segunda requisição, os indivíduos são separados em dois grupos. Para um grupo, poucas explicações são oferecidas para a continuação do experimento. Para outro grupo, são prometidas recompensas financeiras.

Observa-se, então, que os indivíduos do grupo que recebeu a promessa de recompensas financeiras passam a reclamar mais frequentemente da fome dos que os que não receberam promessa alguma. Essa estranha situação é prevista pela TDS. Com efeito, o indivíduo que não tem motivações para continuar no experimento entra em um estado de “dissonância cognitiva”, com a conjunção das seguintes crenças: (i) “eu sei que estou com fome”; e (ii) “eu sei que não tenho nenhuma razão para continuar neste experimento”. Nessa situação, de acordo com a TDS, para escapar da “dissonância cognitiva”, o indivíduo eliminará uma de suas crenças. No caso de eliminar a primeira, ele passará a reclamar menos da fome.

Nesse exemplo, a fome é claramente uma entidade mental (tipo M). Isso é dizer: ela permite acesso privilegiado (exclusivo) ao indivíduo que a experimenta – os pesquisadores dependem do relato subjetivo do indivíduo (no caso, de suas reclamações) para “acessá-la” – e ela não tem uma localização específica no corpo – para Brody e Oppenheim, mesmo que existisse essa localização, ela seria irrelevante.

Mas a fome também é, nesse exemplo, uma entidade científica (tipo 2). De fato, ela é descrita pela “Teoria da Dissonância Cognitiva” (TDS), não pelo senso comum. Essa teoria é capaz de explicar, numa situação específica como a do experimento, o aparecimento e o desaparecimento da fome. À semelhança da teoria glicostática, ela tem capacidade preditiva e merece crédito.

Em suma, se a fome é, numa dada circunstância, uma *entidade de tipos C e 2*, ela pode ser, numa outra circunstância, uma *entidade de tipos M e 2*. Brody e Oppenheim argumentam que o mesmo não sucederia para uma *entidade de tipo 1* (não científica). Ou seja: se uma entidade for, num primeiro momento, de *tipos C e 1*, ela não poderá ser, num momento ulterior, de *tipos M e 1*.

Esse é um ponto crucial do artigo, que talvez mereça análise mais detida. Com efeito, é aqui que Brody e Oppenheim aplicam o “princípio da complementaridade” ao problema mente-cérebro.

Consideremos, de novo, as frases “eu estou experimentando o medo” e “ele está corado”. Essas frases, enunciadas ao léu, não têm quaisquer pretensões científicas, nem estão fundadas numa teoria específica. Referem-se, portanto, a *entidades de tipo 1*. No primeiro caso, apela-se a uma entidade mental (i.e., cujo acesso é “privado” e cuja localização é inexistente ou irrelevante). No segundo caso, apela-se a uma entidade corpórea (i.e., cujo acesso é “público” e cuja localização é precisa). Nessas circunstâncias, as propriedades “mental” e “corpórea” são *incompatíveis*: a primeira frase fará sempre referência a uma entidade “mental”; a segunda, a uma “corpórea”.

Mas, quando consideradas as *entidades de tipo 2*, as propriedades “mental” e “corpórea” se aplicam consoante à teoria empregada. Ou seja: uma mesma entidade pode ser, num primeiro momento, mental (e.g., a fome, de acordo com a teoria da dissonância cognitiva), e, num momento ulterior, corpórea (e.g., a fome, de acordo com a teoria glicostática). Tudo dependerá da escolha do pesquisador pelo seu “aparato teórico”. Diz-se, portanto, que as propriedades “mental” e “corpórea” são apenas *não compatíveis* (em oposição a “incompatíveis”) para as *entidades de tipo 2*.

Da mesma forma que, para a Física Clássica, um objeto será ou “partícula” ou “onda”, porque essas coisas são incompatíveis, para uma *entidade de tipo 1*, aplicar-se-á ou a propriedade “mental” ou a propriedade “corpórea”, porque essas coisas também são incompatíveis.

E, da mesma forma que, para a Física Quântica, um objeto *pode parecer* “partícula” e *pode parecer* “onda”, dependendo do aparato instrumental utilizado, para uma *entidade de tipo 2*, aplicar-se-á a propriedade “mental” e a propriedade “corpórea”, dependendo do aparato conceitual utilizado.

Em ambos os casos, a complementaridade é uma solução para as aparentes contradições.

Comentários finais (Bloco IV)

Há três comentários que gostaria de fazer a respeito deste bloco temático.

Em primeiro lugar, no que toca à última seção (“A complementaridade mente-cérebro”), acho que a teoria de Brody e Oppenheim não funciona completamente. Isso porque, a meu juízo,

uma entidade de tipo 1 (não científica) pode ser, em momentos distintos, mental e corpórea (tal como ocorre com as entidades de tipo 2). Ou seja: para mim, as propriedades mental e corpórea são sempre “não compatíveis” (ao invés de “incompatíveis”), tanto para as entidades de tipo 1 quanto para as entidades de tipo 2.

Para ficar no exemplo apresentado pelos autores, consideremos novamente a “fome”. Eu posso dizer “eu estou experimentando a fome”. Nesse caso, a fome é claramente uma entidade “não científica” e mental (tipos 1 e M). No entanto, eu também posso dizer “a minha barriga está roncando” – e, nesse caso, a fome será claramente uma entidade “não científica” e corpórea (tipos 1 e C).

Em outras palavras, a “fome”, enquanto entidade de tipo 1 (não científica), pode também ser ou mental ou corpórea, dependendo do contexto. Logo, não há uma simetria perfeita com o que ocorre na Física Clássica e na Física Quântica, como querem Brody e Oppenheim.

Mesmo assim, julgo meritória a tentativa dos autores de desenvolver uma teoria da complementaridade para o problema mente-cérebro. A meu ver, a ideia de complementaridade é capaz de abarcar os elementos em questão (mente e cérebro) com o equilíbrio *desejável* – lembremos que a chamada *teoria da identidade mente-cérebro*, desenvolvida por J. J. C. Smart (1920-2012) e Ullin Place (1924-2000) em meados do século passado, a despeito de seu nome, acabou por privilegiar em demasia o elemento cerebral em detrimento do mental.

Ademais, parece prevalecer, na clínica médica, o entendimento de que certas psicopatologias requerem tanto tratamento medicamentoso (e, em alguns casos, cirúrgico) quanto terapias “não invasivas”, baseadas no diálogo e na autoanálise. Esses dois tratamentos são complementares e eficazes, o que indica que tanto o cérebro quanto a mente desempenham um papel relevante na cura do paciente.

Em segundo lugar, gostaria de fazer um comentário genérico a respeito deste bloco temático. O problema mente-cérebro, a meu ver, é muito mais relevante para o tema do livre-arbítrio do que as questões discutidas anteriormente nesta tese (determinismo causal e responsabilidade moral).

Com efeito, se eu tivesse que adivinhar o posicionamento de uma pessoa sobre o tema do livre-arbítrio e só pudesse lhe fazer uma pergunta, minha pergunta seria esta (nesse caso, estou supondo que não posso fazer uma pergunta direta sobre o livre-arbítrio): qual é o seu posicionamento no problema mente-cérebro?

Se a pessoa me respondesse que é dualista, eu arriscaria dizer que ela crê no livre-arbítrio; se a resposta da pessoa fosse o materialismo reducionista (ou o fisicalismo), eu diria que ela não crê no livre-arbítrio. Por quê? Ora, simplesmente porque a noção de livre-arbítrio pressupõe a ideia de *controle* sobre as decisões, e esse controle é, antes de tudo, um atributo da mente.

Seria desarrazoado dizer que o meu cérebro *controla* as minhas decisões (se eu dissesse isso, incorreria na “falácia do sujeito duplo”, como ensinam Mudrik e Maoz). Por outro lado, faz todo sentido dizer que a minha mente *controla* as minhas decisões, supondo que eu acredite em mente da mesma forma que um dualista.

É claro que uma pessoa pode dizer-se materialista reducionista e, ainda assim, acreditar no livre-arbítrio. Nesse caso, é provável que ela tenha inventado uma nova definição para o termo “livre-arbítrio” (diferente da definição fixada nos dicionários, de “vontade espontânea, livre de condicionamentos”). Ou pode ser que ela realmente creia na definição “clássica” de livre-arbítrio. Aqui, são duas as hipóteses: ou ela está sendo incoerente, ou o meu raciocínio (de que materialismo reducionista e livre-arbítrio são incompatíveis) está errado.

Mas, mesmo que o meu raciocínio esteja errado, acredito que ele me conduziria a um índice de acertos muito superior ao alcançado com a seguinte pergunta: você é compatibilista ou incompatibilista?

Pois uma pessoa pode ser incompatibilista e crer ou não crer no livre-arbítrio. Ao dizer-se incompatibilista, ela está apenas afirmando: “eu não acho que as ideias de livre-arbítrio e determinismo causal sejam compatíveis”. Como, a partir dessa informação, eu conseguirei adivinhar se ela crê ou não crê no livre-arbítrio, ou se ela crê ou não crê no determinismo causal?

Da mesma forma, uma pessoa pode ser compatibilista e crer ou não crer no livre-arbítrio (embora, na literatura filosófica, por algum motivo, o termo “compatibilista” esteja sendo reservado para aqueles que creem no livre-arbítrio – ou na pertinência da responsabilidade moral – e creem no determinismo causal). Em princípio, eu posso achar que as noções de livre-arbítrio e de determinismo causal são compatíveis, mas não crer nem em uma coisa nem em outra (exatamente como posso achar que as ideias de Deus e de *vida após a morte* são compatíveis, e não crer nem em uma coisa nem em outra).

Em resumo, para mim, o problema mente-cérebro é muito mais relevante para o livre-arbítrio do que o problema do determinismo causal. Não me causaria surpresa, entretanto, se interlocutores filósofos alegassem não ver diferença entre esses problemas, tamanha a frequência

com que se associam, no debate anglo-saxão, as noções de materialismo reducionista e determinismo, por um lado, e dualismo e indeterminismo, por outro.

Aqui, será preciso lembrar o óbvio: nem todo materialista reducionista é obrigado a crer no determinismo [e.g., eu posso ser materialista reducionista e acreditar na interpretação de Copenhague da Mecânica Quântica], e nem todo dualista é obrigado a crer no indeterminismo [e.g., eu posso ser dualista e acreditar na “harmonia pré-estabelecida” entre alma e corpo, defendida por Leibniz].

Em terceiro lugar, não posso deixar de comentar a excelente tese do “eterno retorno do materialismo”, de Saulo Araújo.

Chama-me a atenção o fato de, presentemente, ser bastante incomum encontrar um filósofo de tradição analítica que se declare abertamente *dualista*. Foi bom aprender com Araújo que o entusiasmo com o materialismo reducionista é uma tendência que remonta aos séculos XVIII e XIX, na França e na Alemanha, respectivamente – sem que, até hoje, esse entusiasmo tenha sido referendado por algum resultado conclusivo.

Há, decerto, filósofos que não endossam o materialismo reducionista – mas nem por isso eles se apresentam como dualistas. Em filosofia da mente, é razoavelmente difundida, por exemplo, a posição *emergentista*.

De uma forma genérica, uma propriedade de um sistema é *emergente* quando ela é mais do que a soma das propriedades dos componentes, i.e., quando ela representa um resultado novo em relação às propriedades iniciais. John Stuart Mill (1806-1873), C. D. Broad (1887-1971), C. Lloyd Morgan (1852-1936), Samuel Alexander (1859-1938) e Ludwig von Bertalanffy (1901-1972) contribuíram para o desenvolvimento de diferentes versões do emergentismo, em diferentes áreas do conhecimento.

No problema mente-cérebro, o emergentismo é a tese de que não se pode reconstruir o estado mental a partir do estado fisiológico [no que os dualistas provavelmente estariam de acordo], ainda que o estado mental tenha a sua origem no estado fisiológico [essa é uma suposição que os dualistas não costumam fazer].

Mais especificamente no tema sobre o livre-arbítrio, há filósofos que endossam posição indeterminista e que, além disso, aceitam a tese da “causação por agente”, em que “agente” significa *substância persistente* [*persisting substance*] [definir exatamente o que significa *substância persistente* é um desafio para os adeptos da posição].

De acordo com a tese da “causação por agente”, o agente (ou “substância persistente”) é um produtor de decisões livres, sendo também uma “causa incausada”. Ou seja: esse agente (ou “substância persistente”) não seria efeito de nada; ele seria apenas a origem de novas cadeias causais. Procuram-se garantir, por essa via, o livre-arbítrio e a responsabilidade moral para os seres humanos.

George Berkeley (1685-1753) e Thomas Reid (1710-1796), mais remotamente, e Roderick Chisholm (1916-1999), Randolph Clarke e Timothy O’Connor, mais recentemente, endossaram visões desse tipo. Como é sabido, Berkeley desenvolveu filosofia bastante singular e contraintuitiva (sem deixar de ser robusta), denominada *idealismo*, que negava a existência de coisas materiais. Por óbvio, Berkeley não pode ser chamado de *dualista*, pois os dualistas creem na existência da matéria.

Um caso emblemático da resistência que o termo *dualismo* provoca em filósofos analíticos contemporâneos é o de Timothy O’Connor, que teria todos os motivos para se apresentar como um *dualista*, uma vez que abraça a teoria da “causação por agente”.

Surpreendentemente, O’Connor (2000) se compromete apenas com um *dualismo de propriedades*. Em sua visão, a teoria da “causação por agente” requereria “um tipo muito distinto de substância em relação à substância material postulada pelo dualismo cartesiano” (O’Connor, p.121, tradução nossa).

Enquanto o *dualismo de substância* (defendido por Descartes) é a doutrina de que as coisas que possuem propriedades mentais são distintas das coisas que possuem propriedades físicas, além de serem irreduzíveis umas às outras (e.g., a alma é distinta do cérebro e é irreduzível ao cérebro), o *dualismo de propriedades* é a doutrina (bem mais modesta) de que as propriedades mentais são distintas das propriedades físicas, além de serem irreduzíveis umas às outras, mesmo que (e isto é relevante) as propriedades mentais e as propriedades físicas pertençam exclusivamente ao cérebro.

Para evitar confusões, penso que o uso do termo *dualista* deveria ser reservado àqueles que pensam como Descartes – o que não parece ser o caso de O’Connor, considerando a forma como ele se apresenta.

No livro que comemorou o centenário do Prêmio Nobel [*100 Years of Nobel Prizes*], é informado que, do lançamento do Prêmio (1901) até o ano de publicação da obra (2002), 65% dos laureados eram cristãos. Para citar apenas físicos, fazem parte dessa lista figuras como o católico romano Guglielmo Marconi (que contribuiu para o desenvolvimento da telegrafia sem fio), o

presbiteriano Arthur Holly Compton (que descobriu o efeito que leva o seu nome) e o luterano Werner Heisenberg (que criou a Mecânica Quântica).

Não há nenhuma evidência de que a crença no dualismo mente-cérebro (comum a todos os cristãos) constitua um óbice para o desenvolvimento da ciência. O descrédito do dualismo cartesiano entre filósofos contemporâneos que discutem o problema do livre-arbítrio parece-me absolutamente despropositado, tanto mais porque Descartes oferece uma possível solução para esse problema.

BLOCO V: ESCOLHA RACIONAL

Escolha racional

Chris Swoyer e Stephen Ellis assinam o verbete “Escolha Racional” [*Rational Choice*], do *New Dictionary of the History of Ideas*. No verbete, é apresentada a “teoria da decisão clássica”, desenvolvida nos primeiros sessenta anos do século XX. Os autores informam que o foco dessa teoria está na “racionalidade instrumental”, isto é, em “como os agentes podem alcançar seus desejos em face de suas crenças”.

As primeiras palavras-chave da teoria da decisão clássica são, portanto, “desejos” e “crenças”. Como notam Swoyer e Ellis, essas noções estão muitas vezes entrelaçadas – é frequentemente mais racional fazer algo com uma recompensa moderada, mas provável, do que algo com uma recompensa elevada, mas improvável. Desse modo, as crenças racionais exercem uma espécie de limitação sobre os desejos.

Os autores do verbete afirmam serem três as condições em que uma decisão pode ser tomada: a condição de certeza, a condição de risco e a condição de incerteza. Na condição de certeza, os resultados da ação são certos e previamente conhecidos; na condição de risco, os resultados da ação não são certos, mas suas probabilidades são conhecidas; na condição de incerteza, nem os resultados da ação são certos, nem suas probabilidades são conhecidas. Na teoria da decisão clássica, predominam análises sobre a condição de risco, segundo Swoyer e Ellis.

“Ações”, “condições” e “resultados” complementam a lista de palavras-chave da teoria da decisão clássica. A ação é uma opção sobre a qual o agente pondera e decide; a condição expressa o estado de coisas que se afigura ao agente imediatamente antes da ação; o resultado é o estado final de coisas, posterior à ação.

A teoria da decisão clássica, segundo Swoyer e Ellis, afirma que as decisões racionais são aquelas que geram o maior nível de satisfação (utilidade) ao indivíduo. Ainda de acordo com os autores do verbete, para essa teoria, todos os indivíduos cujas crenças e desejos respeitam certas premissas (em particular, a premissa da transitividade das preferências, segundo a qual se A é melhor do que B, e B é melhor do que C, A é melhor do que C) são agentes maximizadores de utilidade, i.e., agentes que visam sempre ao maior grau de satisfação.

O grande debate em torno da teoria da decisão clássica diz respeito à natureza de sua tese central. Alguns a tratam como uma tese descritiva – as pessoas de fato agiriam como preconiza a teoria; outros, como uma tese prescritiva – a teoria não captaria o modo real de decisão das pessoas; ela apenas indicaria que um agente racional deve agir maximizando sua satisfação (utilidade).

Entre os argumentos favoráveis à concepção descritiva, destaque-se o da seleção natural, mencionado por Swoyer e Ellis no verbete: a seleção natural deve ter favorecido indivíduos maximizadores de utilidade, de modo que, pouco a pouco, a espécie humana se tornou mais racional no processo de tomada de decisões.

Como objeção à tese descritiva, é apontado que as pessoas raramente têm conhecimento completo sobre suas preferências (utilidades). Mesmo que sempre o tivessem, isso não implicaria, *per se*, aptidão para realizar os cálculos de probabilidade demandados pela teoria da decisão clássica. Em resposta, os adeptos da tese descritiva apontam para a eficácia da teoria em prever a decisão em nível agregado (decisões de grupos econômicos, por exemplo), ainda que, individualmente, sejam mesmo de difícil alcance as competências computacional e de memória requeridas pela teoria.

Entre os argumentos favoráveis à concepção normativa, Swoyer e Ellis dão importância aos que associam a irracionalidade a uma espécie de “autosubversão”. Se as pessoas não agirem com vistas ao maior grau de satisfação (utilidade), elas estarão suscetíveis a um “livro holandês” [*Dutch book*], i.e., a uma série de apostas em que elas certamente perderão, independentemente de como as coisas acabem.

O modelo de escolha da Câmara de Representantes: um diálogo com Pessoa Jr.

Pessoa Jr.: Eu atualmente ando pensando em um modelo de escolha humana inspirado em uma Câmara de Representantes. Meu corpo seria composto por vários “eus” (Nietzsche foi nessa linha, inspirado por Roux), e cada um contribuiria para a decisão final que eu tomo. Isso seria análogo a o que ocorre na Câmara de Deputados, em que uma votação ocorre, e cada parte do corpo contribui com um voto. Um deputado pode estar seguindo o desejo de sua base eleitoral, outro, o desejo de outros deputados, outro recebeu suborno, outro jogou uma moeda ou seguiu suas emoções na hora da votação. O ponto é que a decisão foi tomada pela Câmara, “autonomamente”, no sentido de que as causas em nível corporal são as partes do corpo. Há causas externas, mas há um sentido em que

quem tomou a decisão foi a Câmara. Essa seria, então, uma tentativa de definir uma “autonomia” que, no fundo, se alia com uma visão mecanicista do universo.

(...)

Enfim, você pergunta sobre a noção de graus de liberdade. Sobre a liberdade de ação, acho que não há problema nenhum. Dado que tomo uma decisão, a capacidade de efetivá-la pode ser quantificada, e alguém já deve ter feito isso. Mas você está explorando a liberdade da vontade, e aí o terreno fica pantanoso para mim. Vou tentar usar o modelo da Câmara para responder.

(...)

Uma questão é apresentada à Câmara. Um sentido em que a questão é “simples” é se todos votarem imediatamente, mesmo que com divisão de votos. Outro sentido em que a questão é “simples” é se todos votarem igual, mesmo que haja muita discussão. Então parece que há pelo menos dois sentidos de “dificuldade”: no primeiro caso, a dificuldade é homogeneizar a opinião das partes, mas o mecanismo de decisão (voto em urna, no caso da Câmara; no caso humano, está em aberto a questão do mecanismo de decisão – é uma questão empírica da neurociência o que acontece no corpo quando uma decisão é tomada) é eficaz e rápido. No segundo caso, a decisão demora muito, pois o mecanismo de decisão não propicia um resultado rápido, mas a decisão global finalmente é tomada. O “asno de Buridan” é a situação em que não há deliberação.

(...)

Agora, a questão de se a decisão é “determinada” ou “indeterminada” depende dos mecanismos de decisão de cada parte. Se jogo moeda, posso dizer que há um elemento de indeterminismo. Se pergunto para uma pessoa externa, isso é análogo a um “criptodeterminismo”, e, em relação ao corpo, constitui uma quebra do determinismo (ou seja, considerando o sistema de interesse apenas o corpo).

(...)

O interessante de trabalhar com um modelo como o da Câmara é que ele permite analisar o problema com mais particularidades. Aí a tarefa de quem não considera o modelo verdadeiro é traduzir as questões para uma outra concepção. Esse é o valor do debate interparadigmático.

(...)

O modelo da Câmara aceita que possa haver elementos inconscientes em uma decisão, ou que mesmo toda ela seja inconsciente ou automática. Se ela é “livre”, aí depende da definição de “liberdade”. Creio que, no modelo mecanicista que apresentei, ter-se-ia que traçar a estrutura

causal de como cada parte chegou em sua decisão individual (dado que os mecanismos de decisão são endógenos à Câmara – por exemplo, voto em urna). Em outras palavras, os mecanismos causais se dividem em dois: (1) aqueles que partem das partes e que geram as decisões (processo de voto na Câmara), e (2) aqueles que contribuem para a “decisão” de cada parte. Se houver interferência no mecanismo (1), tipo fraude, há uma quebra de autonomia do tipo (1). Mas, com relação a (2), não se define apropriadamente uma quebra de autonomia, pois cada político está aberto a muitas influências externas.

Guilherme: Achei o modelo de escolha que você apresentou interessantíssimo. Há um ponto que me chama a atenção: pelo paradoxo de Condorcet [ou “paradoxo do voto”, formulado por Marquês de Condorcet (1743-1794) no final do século XVIII, segundo o qual, diferentemente das preferências individuais, as preferências coletivas podem ser cíclicas (situação em que a opção A é julgada melhor do que a opção B, B é julgada melhor do que C e C é julgada melhor do que A)], e pelo teorema da impossibilidade de Kenneth Arrow (1921-2017) [que afirma que não é possível determinar uma ordem clara de preferências num processo de escolha social que respeite certos axiomas], o modelo que você apresenta abre a possibilidade para a ocorrência de escolhas individuais intransitivas, que não estão previstas nas teorias econômicas tradicionais (e.g., na teoria do consumidor) [pois essas teorias tratam a transitividade como um axioma, no caso das escolhas individuais], mas que me parecem ocorrer na vida real.

(...)

Para simplificar as coisas, imaginemos que “eu” (Guilherme) seja constituído por apenas três “eus” [o “Guilherme 1”, o “Guilherme 2” e o “Guilherme 3”], e que eu precise decidir em que cidade quero viver: Rio de Janeiro, São Paulo ou Salvador.

(...)

Para as teorias clássicas da escolha individual consagradas pela ciência econômica, “eu” (Guilherme) sou um sujeito racional, com preferências completas e transitivas. Isso significa que “eu” (Guilherme) sou capaz de comparar todas as opções disponíveis e, ao avaliar uma opção adicional, manter a minha consistência inicial. Por exemplo, se eu julgo que viver no Rio de Janeiro é melhor do que viver em São Paulo e, depois, que viver em São Paulo é melhor do que viver em Salvador, eu sou obrigado a concluir que viver no Rio de Janeiro é, a meu juízo, melhor do que

viver em Salvador [afinal, eu já havia julgado que viver no Rio de Janeiro é melhor do que viver em São Paulo, e que viver em São Paulo é melhor do que viver em Salvador].

(...)

Analisemos, agora, o modelo que você apresenta. A princípio, o “eu” (Guilherme) não é muito relevante para o processo decisório: tudo o que importa são os “eus” que constituem o Guilherme [na prática, o modelo que você apresenta acaba com a ideia de “escolha individual”: todas as escolhas passam a ser “coletivas”, já que todos nós somos compostos por vários “eus”]. Cada um dos “eus” tem autonomia para elencar suas preferências e está sujeito a influências externas. No final, temos a seguinte configuração hipotética:

(...)

O “Guilherme 1” prefere o Rio de Janeiro a São Paulo, e São Paulo a Salvador;

O “Guilherme 2” prefere São Paulo a Salvador, e Salvador ao Rio de Janeiro;

O “Guilherme 3” prefere Salvador ao Rio de Janeiro, e o Rio de Janeiro a São Paulo.

(...)

Agora, veja que interessante. Se “eu” (Guilherme) adotar a “regra da maioria” e fizer a votação dos “eus” que me constituem aos pares, terei os seguintes cenários:

(1) Coloco em disputa, primeiramente, as opções “Rio de Janeiro” e “Salvador”. Em seguida, a opção vitoriosa disputa a “final” com “São Paulo”;

(2) Coloco em disputa, primeiramente, as opções “São Paulo” e “Salvador”. Em seguida, a opção vitoriosa disputa a “final” com “Rio de Janeiro”;

(3) Coloco em disputa, primeiramente, as opções “Rio de Janeiro” e “São Paulo”. Em seguida, a opção vitoriosa disputa a “final” com “Salvador”.

(...)

Pelo cenário 1, a opção vitoriosa seria São Paulo [na disputa inicial, Salvador ganharia do Rio de Janeiro (2 votos contra 1); mas, na disputa final, São Paulo ganharia de Salvador (2 votos contra 1)].

Pelo cenário 2, a opção vitoriosa seria o Rio de Janeiro [na disputa inicial, São Paulo ganharia de Salvador (2 votos contra 1); mas, na disputa final, o Rio de Janeiro ganharia de São Paulo (2 votos contra 1)].

Pelo cenário 3, a opção vitoriosa seria Salvador [na disputa inicial, o Rio de Janeiro ganharia de São Paulo (2 votos contra 1); mas, na disputa final, Salvador ganharia do Rio de Janeiro (2 votos contra 1)].

(...)

Ou seja, o “Guilherme 1”, o “Guilherme 2” e o “Guilherme 3” me deixaram na seguinte situação (um tanto desagradável): “eu” (Guilherme) acredito que São Paulo é melhor do que Salvador; que Salvador é melhor do que o Rio de Janeiro; mas que o Rio de Janeiro é melhor do que São Paulo!

(...)

É claro que, no meu exemplo, a combinação de preferências que gerou a intransitividade da escolha é expressão de um caso particular e improvável.

Mas quanto mais “eus” e opções considerarmos, mais frequentemente lidaremos com a intransitividade.

(...)

Que conclusão tirar disso? Arrisco o seguinte: i) o “eu” (Guilherme) não é o somatório dos diversos “eus” que o constituem; ii) o “eu” (Guilherme) não tem poder de voto; quem vota são apenas os diversos “eus”; iii) no entanto, o “eu” (Guilherme) é o “presidente da Câmara”: ele tem o poder de escolher algumas regras de votação, em particular a ordem em que serão votadas as matérias, quando houver mais de duas (embora seja obrigado a respeitar a “regra da maioria” – nesse sentido, diferentemente de você, não acredito que possa haver fraude no “mecanismo de tipo 1”); iv) as combinações intransitivas podem estar associadas a certas manifestações psicopatológicas (ainda mais se considerarmos que elas são a exceção, e não a regra); v) nas situações de intransitividade, o “eu” (Guilherme), que é o “presidente da Câmara”, tem um “poder moderador” fundamental: a opção vitoriosa será sempre a que ele deixar para votar por último, vide o exemplo das cidades (isso pode explicar a eficácia do autoconhecimento no combate a certas manifestações psicopatológicas).

Pessoa Jr.: Interessante. Considere agora o “problema do asno de Buridan” [problema atribuído ao filósofo e religioso francês do século XIV Jean Buridan (mas que, em verdade, já havia sido abordado por Aristóteles), e que se refere a uma situação hipotética em que um asno é colocado à igual distância de um recipiente com água e um fardo de palha (frise-se que há versões alternativas para o mesmo problema – e.g., a versão que afirma que o asno estaria diante de dois maços de

feno). A premissa do paradoxo é de que o asno sempre procurará aquilo que está mais próximo dele. Se é assim, resulta claro que, na situação descrita pelo problema, a razão que levaria o asno a ir atrás do recipiente com água é a mesma que o incitaria a perseguir o fardo de palha – afinal, a distância entre o asno e o alvo é a mesma para ambos os objetos. Por não ter *razão suficiente* para decidir entre uma opção ou outra, o asno ficaria parado, e morreria de sede e de fome – eis o paradoxo].

Diferentemente do que conclui o paradoxo de Buridan, julgo que, em ocorrendo um empate nas avaliações decisórias, a seleção natural (no mínimo) impôs algum mecanismo de desempate, para impedir a imobilidade. Pode ser a existência de um presidente, como você sugeriu, ou o lance de uma moeda (no caso dos animais, talvez alguma flutuação emocional). Pergunta: esse mesmo mecanismo resolveria o problema da intransitividade? (O “asno de Buridan” não parece ser, a rigor, um caso especial de intransitividade, pois só envolve dois estados, mas talvez sejam todos um mesmo tipo de problema).

Guilherme: Eu acho que os problemas do “asno de Buridan” e da intransitividade têm uma importante característica em comum.

(...)

Nos dois casos, uma vez tomada a decisão, se nos fosse facultado repetir o processo decisório em circunstâncias exatamente idênticas (i.e., submetidos ao mesmo estado inicial e às mesmas leis da físico-química), seria possível decidir diferentemente.

(...)

No limite, em 100 repetições, teríamos algo como 50 escolhas por A e 50 escolhas por B, no caso do problema de Buridan (considerando que as opções fossem A e B), e 33 ou 34 escolhas por A, 33 ou 34 escolhas por B, e 33 ou 34 escolhas por C, no caso da intransitividade (considerando que as opções fossem A, B e C).

(...)

Isso pode significar que nossa decisão foi fruto da mera sorte ou do mero acaso (lançamento de uma moeda, digamos). Nesse caso, não há que falar em livre-arbítrio, penso eu.

(...)

Mas isso também pode significar, como acabei sugerindo, que há um “presidente da Câmara” em nossa mente que, num dado processo decisório, ou encerra a “sessão” antes de “computar todos os

votos” (no problema de Buridan) [nesse caso, o impasse, que tem potencial para ser eterno, é resolvido pela arbitrariedade do “presidente”, que não “ouve” todos os “eus” eleitores – se o ouvisse, certamente haveria empate e “imobilidade”], ou altera a ordem em que as opções são votadas (no problema da intransitividade) [nesse caso, em função da prerrogativa de poder votar por último a opção que mais lhe interessa, o “presidente” fixa uma ordem de preferências para a Câmara, ainda que, em teoria, essa ordem inexista]. Nesse sentido, eu acho que há livre-arbítrio, e que o grau de liberdade é máximo (é aqui que o ser humano tem mais *controle* sobre a decisão). Resumindo, eu acredito que os mecanismos que você mencionou resolvem o problema da intransitividade, no sentido de permitirem uma resposta.

Comentários finais (Bloco V)

Há dois comentários que gostaria de fazer sobre este bloco temático.

Em primeiro lugar, que foram as ideias de Desidério Murcho (apresentadas no início desta tese) que me levaram a estudar o tema da “escolha racional”. Murcho, vale lembrar, cogitou que o livre-arbítrio pode estar relacionado à capacidade de se tomar “melhores decisões, baseadas em melhores deliberações”.

Essa ideia, em princípio, me apraz. Mas, tão logo tomei contato com a teoria da decisão clássica (racional), percebi que não há nada mais distante dela do que o conceito de livre-arbítrio. A teoria da decisão clássica visa à maximização da utilidade do indivíduo (em outras palavras, ao aumento de sua satisfação). Dadas as preferências e as crenças do indivíduo, a teoria da decisão clássica indica, através de cálculos de probabilidade sofisticados, qual é a melhor decisão a ser tomada em cada caso.

Ora, uma escolha com livre-arbítrio não necessariamente será pautada por um critério racional. Se o ser humano decidisse sempre baseado nas “melhores razões”, ele não decidiria de maneira livre. Ele seria, em verdade, “escravo” das melhores razões. Isso é especialmente válido se entendermos livre-arbítrio como “capacidade para decidir de outro modo”.

No que se refere ao outro sentido de livre-arbítrio (origem ou determinação das decisões pela vontade consciente), não está claro para mim se ele se compatibiliza com a teoria da escolha racional. De todo modo, o livre-arbítrio me interessa muito mais em seus aspectos qualitativos do

que em sua dimensão quantitativa (em que pese eu advogar, na sequência do texto, pela tese de que a sensação de livre-arbítrio existe em graus, e que é possível medi-los).

Para mim, o que o livre-arbítrio teria a oferecer às pessoas (caso existisse) seria não a obtenção de um grau máximo de prazer (ou satisfação) em cada processo de escolha, mas a resolução de “problemas incomensuráveis”, de ordem eminentemente qualitativa, “humanos, demasiado humanos”, que uma máquina teria dificuldades para resolver.

Na “era da informação” e do “Big Data”, o livre-arbítrio seria um elemento valioso para os seres humanos. A própria dignidade da pessoa humana e seu valor no mundo do trabalho estariam assentados no livre-arbítrio. Há muitas tarefas que as máquinas podem executar, mas há aqueles problemas que apenas seres humanos, possuidores do livre-arbítrio, seriam capazes de resolver satisfatoriamente. Refiro-me a questões qualitativas, concernentes a valores humanos. Nesse sentido em que atribuo valor ao livre-arbítrio, a teoria da decisão clássica me parece de pouca valia.

Em segundo lugar, fiz questão de deixar registrado o enriquecedor diálogo que tive com Osvaldo Pessoa Júnior sobre o modelo de decisão da Câmara de Representantes. Esse modelo será muito útil na análise de experimentos da neurociência do livre-arbítrio. Por ora, quero fazer uma reflexão adicional sobre o problema do asno de Buridan, que sempre me fascinou.

À primeira vista, parece tratar-se de problema tolo: é óbvio que, ao contrário do que conclui o paradoxo, o asno não ficará parado até morrer de sede e de fome (em que pese o brocardo popular “pensando, morreu um burro!” aparentemente ratificar as ideias atribuídas a Buridan). O valor do problema de Buridan está, a meu ver, em sua lógica subjacente. Mais especificamente, o problema do asno de Buridan explicita as consequências necessárias (e *desagradáveis*) que se apresentam a quem endossa determinada linha de pensamento.

Refiro-me à aceitação do “princípio da razão suficiente”, segundo o qual, para cada entidade ou evento Z, se Z existe, há uma explicação suficiente para Z existir. Ora, esse é um princípio muito bem estabelecido em Filosofia desde Gottfried Wilhelm Leibniz (1646-1716). No entanto, se o utilizarmos na situação descrita pelo problema do asno de Buridan, não poderemos concluir nada de diferente do que conclui o paradoxo.

A conclusão de que o asno morrerá de sede e de fome é, em termos práticos, absurda; mas, em termos teóricos, necessária, se aceitarmos o princípio da razão suficiente. Dado que o problema do asno de Buridan é aquele em que, por hipótese, não há *razão suficiente* para se deliberar por

qualquer uma das opções, não haveria outro destino para o asno, a não ser morrer de fome e de sede.

BLOCO VI: A NEUROCIÊNCIA DO LIVRE-ARBÍTRIO

O experimento de Libet et al.

Libet e parceiros realizaram, em 1983, na Universidade da Califórnia, em *San Francisco*, o que viria a ser um dos experimentos seminais e de maior impacto na área da neurociência: sujeitos com atividades cerebrais monitoradas por um sistema de corrente contínua com eletrodos ativos sobre o couro cabeludo, especificamente sobre as áreas do córtex motor relacionadas aos movimentos da mão direita, colocados diante de um relógio cujo indicador de tempo perfazia o giro de 360 graus em apenas 2,56 segundos, foram instados a flexionar o punho direito, também mapeado por eletrodos, quando sentissem vontade consciente, e a memorizar o instante em que, pelo seu juízo, sentiram a “primeira vontade” de flexionar o punho direito. Ao final do processo, os participantes deveriam informar esse dado, que convencionou-se chamar de juízo *W*.

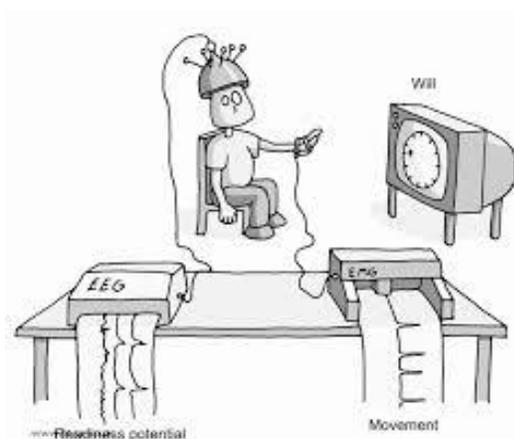


Figura 4: Ilustração do experimento de Libet e colaboradores (1983). Ao contrário do que sugere a imagem, foi o punho direito, e não o esquerdo, a ter seus movimentos aferidos. Imagem retirada da Internet.

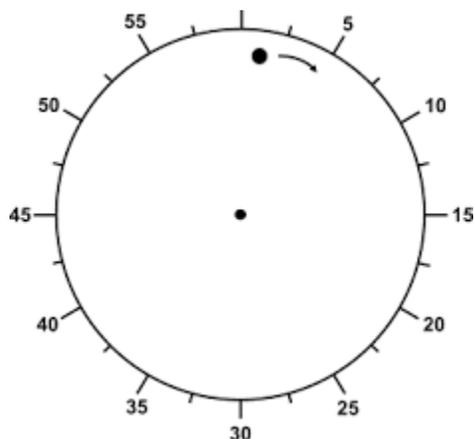


Figura 5: O “relógio de Libet”. O ponto mais próximo à circunferência perfaz os 360 graus em apenas 2,56 segundos, e continua a mover-se por um tempo randômico após o movimento do punho direito. Imagem retirada da Internet.

À época, já era conhecido o “potencial de prontidão” [*Bereitschaftspotential*, BP, ou *readiness potential*, RP], descoberto por Hans Kornhuber e Lüder Deecke em 1965. Trata-se de uma voltagem detectada no córtex motor e na área motora suplementar, e que está associada aos movimentos voluntários do corpo. Em seu experimento, Libet e colaboradores calcularam, para cada participante, a média dos instantes em que pela primeira vez apareceram os potenciais de prontidão (RPs), a fim de compará-la com os instantes associados aos juízos *W*, e com os instantes em que apareceram os sinais eletromiográficos (EMGs), que monitoram as membranas excitáveis das células musculares (isto é, que medem diretamente o instante do movimento).

Decerto, era esperado que os potenciais de prontidão (RPs) antecedessem os sinais eletromiográficos (EMGs), em consonância com os achados de Kornhuber e Deecke, que detectaram sinais elétricos no cérebro antes da realização, pelos participantes, de movimentos conscientes no corpo. O que não se sabia era a posição que o juízo *W* ocupava nessa linha cronológica. Seria razoável supor, seguindo uma concepção ordinária sobre a relação mente-cérebro, que *W* (a vontade de flexionar o punho direito) antecedesse – ou, por que não, *gerasse* – a atividade cerebral correspondente ao potencial de prontidão (RP), e que esta, por sua vez, antecedesse a detecção dos sinais eletromiográficos (EMGs), correspondentes ao movimento propriamente dito.

No entanto, os resultados do experimento indicaram uma ordem diversa. O instante reportado pelos participantes como sendo o da vontade de mexer o punho antecedeu em aproximadamente 200ms o seu movimento (detecção de EMGs), mas sucedeu, em pelo menos 350ms, o início [*the onset*] da detecção do potencial de prontidão (RPII). Nos casos em que os sujeitos revelaram ter planejado o movimento, o início da detecção do potencial de prontidão [RPI] antecedeu em cerca de 850ms o aparecimento da vontade consciente.

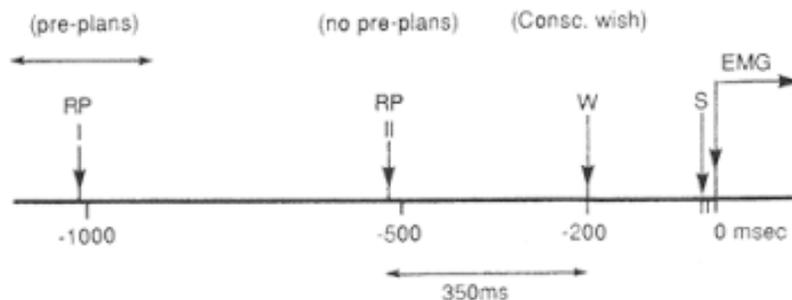


Figura 6: A sequência cronológica dos eventos cerebrais e mentais, de acordo com o experimento de Libet et al. (1983). A fim de atestar a correção dos dados informados pelos participantes, também lhes foi aplicado o estímulo elétrico S. A diferença entre a percepção do estímulo e a sua real aplicação foi de apenas 50ms. Imagem retirada da Internet.

Em suma, antes de os indivíduos sentirem vontade consciente de flexionar o punho, já estava em curso o processo fisiológico que culminaria no aparecimento dessa vontade e, em seguida, na realização do movimento. A que deveria, então, ser atribuída a flexão do punho: à vontade consciente do indivíduo ou à atividade cerebral que gerou essa vontade – se é que essas coisas são distintas?

A concepção de Libet sobre o livre-arbítrio

Nos 200 milésimos de segundos que separam, no experimento da flexão do punho, o aparecimento da vontade consciente (juízo *W*) da flexão propriamente dita (EMG), haveria a possibilidade de o indivíduo interromper o processo fisiológico que conduz ao movimento, impedindo a sua realização. Fundamentalmente, esse “poder de veto” do indivíduo não estaria associado a um novo potencial de prontidão (RP); em vez disso, conformaria uma atividade estritamente mental (não física). Tratar-se-ia, pois, de um veto da mente sobre o corpo.

Essa foi a forma que Libet (1999) encontrou para explicar o fato de alguns participantes não terem flexionado o punho, a despeito da detecção inicial do potencial de prontidão (RP) e do aparecimento da vontade consciente.

Para Libet, uma decisão consciente seria finalizada com o uso (ou não) do poder de veto pela mente, vale dizer, com o uso (ou não) do livre-arbítrio pelo indivíduo. Aqui, a noção de livre-arbítrio é claramente ressignificada. Ela não é mais entendida como a possibilidade de agir a partir de uma vontade espontânea ou livre de condicionamentos. Ela passa a ser entendida como o poder de interromper um processo fisiológico que conduziria à ação.

Os experimentos de Isham et al.

Há neurocientistas que questionam a existência de um “poder de veto” da mente. Eve Isham e parceiras (2017), através de uma análise comparada dos intervalos de tempo de que disporia a mente para exercer o seu “poder de veto” em decisões fáceis e difíceis, se surpreenderam com o fato de que o tempo disponível para realizar o veto é maior nas decisões fáceis do que nas decisões difíceis.

Segundo as autoras, era de se esperar o contrário, na medida em que uma escolha fácil necessita, em tese, de menos deliberação. Daí concluírem que a mente não pode ser a causa direta de um veto – ou melhor, que o fenômeno do veto provavelmente tem uma origem fisiológica prévia ao aparecimento da vontade consciente.

Os experimentos liderados por Isham foram consideravelmente mais complexos do que o comandado por Libet, englobando uma série de variáveis novas. Num desses experimentos, os participantes foram expostos a gravações de áudio, em que eram lidas sentenças afirmativas (e.g., “eu prefiro a cor vermelha à cor azul”, “para salvar um povoado, seria legítimo sacrificar uma criança” etc.). À disposição dos participantes, havia os botões “Concordo” e “Não Concordo”, e um relógio.

Ao ouvirem as frases, os participantes deveriam apertar o botão que julgassem apropriado, atentando-se para a posição de um ponto vermelho no relógio no instante em que surgiu a primeira “vaga ideia” de qual seria sua resposta. Ou seja: antes de apertarem o botão, os participantes deveriam observar, no relógio, quando “começaram a chegar” a sua resposta. Adicionalmente, os participantes deveriam medir o grau de dificuldade de sua decisão, numa escala de 1 (mais fácil) a 5 (mais difícil).

Computados os dados, o que constaram as pesquisadoras? As decisões julgadas “mais fáceis” demoraram mais tempo para se afigurarem aos participantes do que as julgadas “mais difíceis”. Em outros termos, a primeira “vaga ideia” de qual seria a resposta dos participantes veio antes para as frases mais difíceis e depois para as frases mais fáceis.

No sentido inverso, a mente precisaria de mais tempo para exercer seu “poder de veto” nas decisões mais difíceis (e de menos, nas decisões mais fáceis). Por essa razão, na visão das autoras, a mente não pode ter sido a causa imediata dos vetos no seu experimento.

O experimento de Filevich et al.

Elisa Filevich e parceiros (2013) tampouco endossam a concepção de Libet sobre o livre-arbítrio. O grupo desenvolveu um experimento em que os participantes eram solicitados a, em resposta a um estímulo visual, decidir entre apertar rapidamente uma tecla ou inibir (vetar) a ação, atrasando o movimento.

Os pesquisadores então calcularam a amplitude média das respostas eletrofisiológicas ao estímulo visual, conhecidas como “potenciais relacionados a eventos” (ERPs, na sigla em inglês), tomando o cuidado de distinguir as ações “não livres” das “livres”, pois cada ensaio exigia, além de respostas “comportamentais” do indivíduo (e.g., “aperte imediatamente a tecla” ou “aguarde um pouco antes de apertar a tecla”), respostas “genuínas” [nomenclatura nossa] do indivíduo (e.g., “faça o que você quiser: aperte imediatamente a tecla ou aguarde um pouco antes de apertar a tecla”).

Os ERPs foram medidos em três intervalos subsequentes de 50ms. Em dois desses intervalos (-150ms a -100ms, e -100ms a -50ms em relação ao início da ação), constataram-se, para as respostas “genuínas”, diferenças nas amplitudes médias dos ERPs: as maiores amplitudes vieram associadas à decisão de agir imediatamente, e as menores, à escolha por inibir (vetar) a resposta imediata.

Ou seja: nos ensaios que exigiram respostas livres dos indivíduos, a atividade cerebral preparatória foi mais forte quando os participantes escolheram apertar imediatamente a tecla. No entanto, quando eles escolheram atrasar o movimento, sua decisão também foi precedida por ERPs, o que sugere uma origem fisiológica para o veto.

Filevich e parceiros concluíram que vetos de última hora provavelmente dependem de atividades preparatórias não conscientes, que são menos intensas do que as atividades necessárias para a ação.

O experimento de Haggard e Eimer

Dentre os estudiosos que se propuseram a avaliar e dar prosseguimento ao trabalho de Libet, inclusive por meio de variações de seu experimento, distingue-se a figura do britânico Patrick Haggard, professor de neurociência cognitiva da Universidade de Londres.

Em 1999, Haggard e seu colega Martin Eimer publicaram artigo que acrescentaria novas variáveis ao problema do livre-arbítrio: no experimento da dupla, os participantes foram solicitados a mover, sempre que desejassem, um dos dedos indicadores. A principal diferença com o experimento de Libet estava justamente na possibilidade de os participantes escolherem um dos dedos para mover, avalizada pelo fato de os autores não terem constatado diferença entre a “condição de movimento fixo” [*fixed movement condition*], em que apenas um dedo indicador poderia ser movido, e a “condição de movimento livre” [*free movement condition*], em que qualquer um dos dedos indicadores poderia ser movido. Com Libet, lembre-se, os participantes podiam mover apenas o punho direito.

Para fins de análise, uma vez coletados os dados, Haggard e Eimer dividiram os juízos *W* (isto é, os instantes reportados pelos participantes como sendo o da “primeira vontade” de mexer um dos dedos) em dois grupos: os “precoces” [*early*] (aqueles que antecederam em maior medida o movimento de um dos dedos), e os “tardios” [*late*] (aqueles que antecederam em menor medida o movimento de um dos dedos).

Se o juízo *W* fosse de fato causado pelo potencial de prontidão (RP), como indicou Libet, um juízo *W* “precoce” deveria ser causado por um potencial de prontidão “precoce”, e um juízo *W* “tardio”, por um potencial de prontidão “tardio”. Observou-se, entretanto, que não houve uma correlação entre juízos *W* “precoces” e potenciais de prontidão “precoces”, ou entre juízos *W* “tardios” e potenciais de prontidão “tardios”.

Os autores levantaram, então, os dados dos potenciais de prontidão lateralizados [*lateralized readiness potentials, LRPs*], que se manifestam no hemisfério contralateral à ação motora, com o intuito de comparar o sinal contralateral ao sinal ipsilateral, medindo a diferença entre eles, num “método de subtração dupla”.

Constatou-se que os potenciais de prontidão lateralizados (LRPs) poderiam ser a causa do aparecimento da vontade consciente, pois LRPs “precoces” vieram associados a juízos *W* “precoces”, e LRPs “tardios”, a juízos *W* “tardios”. Convém reforçar que o potencial de prontidão lateralizado (LRP) é aquele diagnosticado no hemisfério oposto ao lado da mão flexionada (se o dedo indicador flexionado é o direito, o potencial de prontidão lateralizado é o do córtex motor do hemisfério esquerdo, e vice-versa).

Libet e parceiros fizeram a maior parte das medições no “vértice” (i.e., no topo) do crânio, onde o sinal do potencial de prontidão (RP) era mais intenso. A prevalência de medições no córtex motor direito e esquerdo é outra novidade do experimento de Haggard e Eimer.

Libet versus Haggard e Eimer

Pode-se considerar o resultado de Haggard e Eimer um avanço em relação ao resultado de Libet? É preciso enfatizar que Haggard e Eimer não fizeram uma réplica do experimento de Libet. Eles adaptaram o experimento de Libet, incluindo tarefas, medições e métodos de análise novos. Por consequência, sua conclusão não necessariamente invalida a conclusão de Libet.

O próprio Libet (2001) não julgou a conclusão de Haggard e Eimer superior a sua: “suas descobertas até o momento precisam ser tratadas em importantes aspectos técnicos antes que suas conclusões possam ser consideradas válidas” (Libet e Haggard, p.61, tradução nossa).

Entre outras críticas, Libet sugeriu que os potenciais de prontidão (RPs) coletados por Haggard e Eimer estariam contaminados por um componente de pré-planejamento. Esses potenciais de prontidão (RPs) corresponderiam, na verdade, ao que, no seu experimento de 1983, Libet chamou de RPI. Na visão de Libet, Haggard e Eimer deveriam, ao contrário, ter coletado os dados do RPII, que aparece cerca de meio segundo depois do RPI.

Beatriz Sorrentino Marques e os diferentes tipos de decisões

Beatriz Sorrentino Marques (2015) distingue quatro tipos de decisões: as ativas e reflexivas [*active and reflective decisions*]; as ativas e não reflexivas [*active and unreflective decisions*]; as não ativas e reflexivas [*non-active and reflective decisions*]; e as não ativas e não reflexivas [*non-active and unreflective decisions*].

Apenas as últimas seriam contempladas por um experimento de neurociência que conduz (ou poderia conduzir) à conclusão de que o livre-arbítrio não existe (Bode et al., 2011). Sorrentino Marques reconhece que atos “não ativos e não reflexivos”, como os aferidos no experimento comandado por Stefan Bode, podem carecer de liberdade. Por outro lado, a autora deixa em aberto a possibilidade de o livre-arbítrio existir nas decisões ativas e reflexivas.

Experimentos de Trevena e Miller

Judy Trevena e Jeff Miller (2010) idealizaram dois experimentos para avaliar se as ações voluntárias são precedidas por atividade estritamente cerebral. A dupla procurava evidência para a tese de Libet et al. sobre a iniciação inconsciente de movimentos voluntários.

Recorde-se que, para Libet, sinais eletrofisiológicos [os “potenciais de prontidão” (RPs)] indicariam a preparação cortical para decisões conscientes de movimentos corpóreos. Trevena e Miller não se revelaram dispostos a endossar, a priori, tal interpretação, reputando-a como controversa. À dupla, pareceu desejável formular um teste em que se comparassem os potenciais de prontidão associados a decisões conscientes para mover [*decision to move*] com potenciais de prontidão associados a decisões conscientes para não mover [*decision not to move*].

No experimento 1, foi empregado o método da eletroencefalografia (EEG). Os participantes foram instruídos a decidir entre apertar uma tecla rapidamente ou não fazer nenhum movimento, ao ouvirem um tom temporalmente imprevisível. A utilização do tom (não empregado no experimento original de Libet) tinha o propósito de fixar o instante da medição do potencial de prontidão (RP). Para que a interpretação de Libet et al. fosse endossada, seria necessário que a magnitude do RP fosse maior nos participantes que decidissem apertar a tecla.

No experimento 2, em que também foi empregado o método da eletroencefalografia (EEG), Trevena e Miller centraram esforços em medir o potencial de prontidão lateralizado (LRP). Aos participantes, foi-lhes dada a oportunidade de escolher qual mão mover, em resposta a um tom temporalmente imprevisível. O LRP mede a diferença entre os potenciais de prontidão nos dois hemisférios do cérebro. Ao menos desde o final da década de 1980, entendia-se que um sinal eletrofisiológico mais intenso no hemisfério esquerdo estaria associado a uma preparação de movimento da mão direita, e vice-versa.

Em seus experimentos, Trevena e Miller não encontraram sinais eletrofisiológicos mais fortes em decisões para mover do que em decisões para não mover (experimento 1). Tampouco constataram que a prevalência desses sinais em um dos hemisférios do cérebro estaria associada ao movimento de uma mão específica (experimento 2). Concluíram que “estes sinais claramente não são específicos para a preparação de movimento” (Trevena e Miller, p. 447, tradução nossa), e que as conclusões de Libet carecem de evidência.

Os autores mencionaram, por fim, uma interpretação alternativa para o papel desempenhado pelo potencial de prontidão (RP): este, em vez de atividade preparatória para um movimento voluntário, poderia simplesmente indicar a preparação para o envolvimento contínuo em uma tarefa.

Crítica de Gilberto Gomes a Trevena e Miller

Em artigo publicado na revista *Consciousness and Cognition* (2010), Gilberto Gomes questiona as conclusões de Trevena e Miller (2010) em duas frentes. Em primeiro lugar, o autor indaga se as decisões dos participantes nos experimentos de Trevena e Miller são comparáveis às decisões dos participantes no experimento de Libet et al. (1983). Em segundo lugar, Gomes pergunta se os potenciais neurais medidos por Trevena e Miller são similares aos potenciais medidos por Libet et al. (1983).

No que toca ao primeiro ponto, o autor brasileiro considera as decisões dos experimentos de Trevena e Miller “menos espontâneas” do que as do experimento de Libet et al. Com efeito, ao contrário de Libet et al., Trevena e Miller fixaram o momento em que os participantes deveriam tomar a decisão, apelando a um estímulo sonoro (tom). Para Gomes, “dificilmente pode ser visto como os resultados obtidos aqui [experimentos de Trevena e Miller] podem ser relevantes para a questão da iniciação espontânea de um movimento” (Gomes, 2010, p. 458, tradução nossa, colchetes nossos).

No que toca ao segundo ponto, Gomes constata que a negatividade medida por Trevena e Miller é menor e evolui de forma mais lenta do que aquela aferida por Libet et al. Há diferenças significativas nas amplitudes e nos cursos de tempo dos potenciais. Em Trevena e Miller, os potenciais mal atingem $-2 \mu\text{V}$, e seu início ocorre, em média, 2500ms antes do movimento. Em Libet et al., são registrados potenciais que atingem facilmente $-10 \mu\text{V}$, que começam a ocorrer, em média, 550ms antes do início do movimento.

Na visão de Gilberto Gomes, muito provavelmente o índice medido por Trevena e Miller não foi o potencial de prontidão (RP), mas a “variação negativa contingente” [*contingent negative variation* (CNV)], descoberta no início dos anos 1960 por William Grey Walter e outros pesquisadores. *Grosso modo*, a CNV indica a antecipação de um estímulo sensorial, e relaciona-se à expectativa de realização de determinada tarefa. O RP, por sua vez, desde sua descoberta por

Kornhuber e Deecke em meados da década de 1960, é associado à preparação cerebral de um movimento voluntário.

Em seu artigo, Trevena e Miller chegaram a discutir a “variação negativa contingente” (CNV), sugerindo que ela, além de refletir a antecipação para um estímulo sensorial, poderia também indicar uma preparação cerebral para o movimento. Mas, para Gomes, há uma razão fundamental para CNV e RP designarem coisas distintas: “a CNV pode também estar presente na ausência de qualquer movimento, real ou intencional, portanto não é apenas uma negatividade precedente de movimento” (Gomes, 2010, p. 458, tradução nossa).

Algo que realmente se parece com o potencial de prontidão (RP) no experimento de Trevena e Miller, na visão de Gomes, é a negatividade medida após o estímulo sonoro (tom). De fato, o formato das ondas registradas por Trevena e Miller após o tom em muito se assemelha ao das ondas encontradas nos experimentos de veto de Libet et al. Gomes conjectura que, em Trevena e Miller, os participantes teriam preparado sua decisão, de forma inconsciente, antes mesmo de ouvirem o tom. De todo modo, para o autor brasileiro, o fenômeno do veto (preparação cerebral para o movimento, seguida de decisão para não mover) mereceria novos estudos.

Comentários finais (Bloco VI)

Em primeiro lugar, gostaria de fazer um comentário acerca da seção “Os experimentos de Isham et al.”. As pesquisadoras apelam a uma problemática distinção entre “decisões fáceis” e “decisões difíceis” para concluir que a mente não tem poder de veto, como queria Libet.

Ora, a suposição de que uma decisão fácil é aquela que necessita de um menor tempo para deliberação, ao passo que uma decisão difícil necessita de um maior tempo, parece-me muito mal fundada. Decerto, há decisões difíceis que nos ocupam longamente, e há decisões fáceis que resolvemos de forma célere.

Mas é perfeitamente possível que a decisão mais difícil que tenhamos que tomar em nossa vida ocorra numa fração de segundos, a depender das circunstâncias. Se, numa praia, avistarmos duas pessoas se afogando e tivermos uma única boia à disposição, teremos que tomar, num tempo curtíssimo, uma decisão de grande envergadura moral: qual das duas pessoas salvar? Assim como esse, há vários outros exemplos que põem em questão a suposição das autoras.

Não seria sequer necessário apelar a exemplos, contudo. Dado o modelo da Câmara de Representantes que tem sido defendido, entre outros, por Osvaldo Pessoa Júnior, é simples perceber que uma “escolha fácil” nem sempre é sinônimo de uma “escolha rápida”, assim como uma “escolha difícil” nem sempre é sinônimo de uma “escolha demorada”.

No diálogo com Pessoa Jr., apresentado no bloco temático anterior, foi observado que há dois sentidos de “dificuldade”: um primeiro que, de fato, se relaciona ao tempo para a deliberação; e um segundo que nada tem a ver com o tempo para a deliberação, mas com o grau de certeza da decisão (i.e., quando todos os “eus” votam de uma mesma maneira). Portanto, não parece legítimo tirar qualquer conclusão de uma suposição que, a meu juízo, é flagrantemente equivocada.

Em segundo lugar, gostaria de escrever algumas palavras sobre o que chamei de “A Neurociência do Livre-Arbítrio”. Por um lado, parece-me formidável que neurocientistas se interessem por temas de longa tradição filosófica, como o livre-arbítrio. É também salutar que filósofos se deem ao trabalho de ler artigos da neurociência sobre temas com algum impacto na Filosofia. O tema do livre-arbítrio é, possivelmente, um dos mais interdisciplinares do momento (considerando, ao menos, os temas filosóficos), dada a diversidade de pessoas que escrevem a respeito.

Esse processo não acontece sem riscos, entretanto. Em particular, preocupa-me a tendência de neurocientistas (talvez menos acostumados a cultivar a dúvida do que filósofos) quererem dar, muitas vezes de forma açodada, uma “palavra final” sobre o assunto.

De Libet para cá, muitos se arvoraram a dizer que o livre-arbítrio é uma ilusão (Libet, faça-se justiça, não o fez). Há uma certa ousadia nisso. Mas será que um problema de dois milênios, que ocupou mentes verdadeiramente brilhantes (como lembra Leão Tolstói em *Guerra e Paz*) e que, não obstante, permaneceu insolúvel, foi mesmo resolvido num curto espaço de tempo com o advento da neurociência?

Se for esse o caso, os cientistas responsáveis pela “descoberta” de que o livre-arbítrio é uma ilusão são, de longe, os mais injustiçados da história. Em primeiro lugar, porque não se tem registro de alguém que tenha recebido um prêmio Nobel por essa “descoberta”. E, em segundo, porque esse “fato” não foi sequer noticiado pelos principais jornais.

Ora, pode ser que o livre-arbítrio seja mesmo uma ilusão – mas, definitivamente, não estaríamos falando de uma ilusão qualquer, mas de uma “ilusão incorrigível” para a maior parte das pessoas.

Acreditar que a mente é redutível ao corpo, que a consciência não tem nenhum controle sobre as decisões, que existir, em suma, não é mais do que assistir passivamente a um incontável número de eventos físicos que se sucedem até a morte, quando a consciência desaparece para sempre, não é algo tão trivial quanto alguns neurocientistas fazem crer.

BLOCO VII: UMA CIÊNCIA PARA A SENSAÇÃO DE LIVRE-ARBÍTRIO

Filosofia da neurociência

John Bickle, Peter Mandik e Anthony Landreth são especialistas no tema “filosofia da neurociência”, e é deles o verbete homônimo da *Stanford Encyclopedia of Philosophy*. No verbete, é observado que a filosofia da neurociência se insere no campo mais amplo da filosofia da ciência, e responde a uma tendência, já presente há quatro décadas, de especialização dessa disciplina.

Em que pese “filosofia da neurociência” e “neurofilosofia” não serem termos sinônimos, muitas vezes eles foram empregados indistintamente. Na virada do século XX para o XXI, começou a tornar-se mais nítida a diferença entre eles.

A “filosofia da neurociência” tem como ponto de partida a Filosofia e como ponto de chegada a neurociência. Ela adapta as questões tradicionais da filosofia da ciência à área da neurociência, para tratar de suas questões fundacionais. A “neurofilosofia”, por outro lado, tem como ponto de partida a neurociência e como ponto de chegada a Filosofia. Ela aplica os conceitos neurocientíficos às questões filosóficas tradicionais.

Seguindo os exemplos dos autores do verbete, pode-se dizer que, ao explorar os vários conceitos de “representação” empregados nas teorias neurocientíficas, o investigador pratica “filosofia da neurociência”; mas, ao examinar as implicações das síndromes neurológicas para o conceito de “eu”, é “neurofilosofia” a sua ocupação. Ressalte-se que, na prática, muitos “filósofos da neurociência” são também “neurofilósofos”.

Na medida em que a neurociência é uma ciência em plena evolução, um dos aspectos mais excitantes da filosofia da neurociência, nos dizeres de Bickle et al., é seu “contínuo elemento de surpresa”. Aquilo que, há pouco, era uma mera suposição, pode tornar-se agora a posição majoritária dos neurocientistas, e os filósofos devem estar atentos para acompanhar essas rápidas transformações.

A filosofia da neurociência nem sempre é um retrato fiel da neurociência, entretanto. Os autores do verbete mencionam que, não obstante a maior parte dos membros da Sociedade de Neurociência [*Society for Neuroscience*] classificar sua pesquisa como “neurobiologia”, com foco nos mecanismos moleculares e celulares da atividade neuronal, a maior parte da atenção filosófica

está voltada para a “neurociência de alto nível”, vale dizer, para as redes e os sistemas cognitivos, e a neurociência computacional.

Além disso, nem mesmo as filosofias da mente de teor materialista foram, historicamente, muito influenciadas pelas descobertas neurocientíficas. As teorias da identidade mente-cérebro de Ullin Place e John Smart nos anos 1950, por exemplo, estavam fundadas mais em intuições filosóficas do que em detalhes da emergente neurociência, na visão de Bickle et al.

Não escapa a essa tendência o materialismo eliminativista, que tem em Paul e Patrícia Churchland seus defensores mais emblemáticos. Os autores do verbete apresentam o materialismo eliminativista como uma conjunção de duas teses: a tese de que é falsa a “psicologia do senso comum” sobre processos e eventos mentais; e a tese de que essa “psicologia do senso comum” será inteiramente substituída por uma neurociência futura.

Os Churchland caracterizam a “psicologia do senso comum” como um conjunto de máximas a que as pessoas apelam para explicar seu comportamento. Via de regra, essas máximas pressupõem causalção do mental sobre o físico. É o que sucede quando uma pessoa alega ter agido de tal ou qual modo porque seguiu os ditames de sua “alma”. Para os Churchland, o mecanismo real de causalção será, sempre, puramente físico.

Mas há uma área em que as descobertas neurocientíficas causam grande impacto no debate filosófico, que lhes acompanha de perto: o estudo da consciência e a sua relação com os mecanismos físicos. Os autores mencionam a comissurotomia (cirurgia que desconecta os dois hemisférios cerebrais, cortando as comissuras do cérebro), e também aludem às experiências subjetivas de dor e de visão de cores. Todas essas questões têm sido objeto de análise de filósofos.

Destaque-se, ainda, a menção de Bickle et al. a uma das primeiras questões a emergir na neurologia: a localização, no cérebro, de funções cognitivas específicas. Nesse aspecto, os estudos sobre afasia de Paul Broca, em meados do século XIX, foram seminais. O anatomista francês pôde comprovar que lesões no hemisfério cortical esquerdo eram predominantes nos casos de afasia, e que eram necessárias lesões na segunda e na terceira convoluções frontais para que déficits de produção de fala ocorressem.

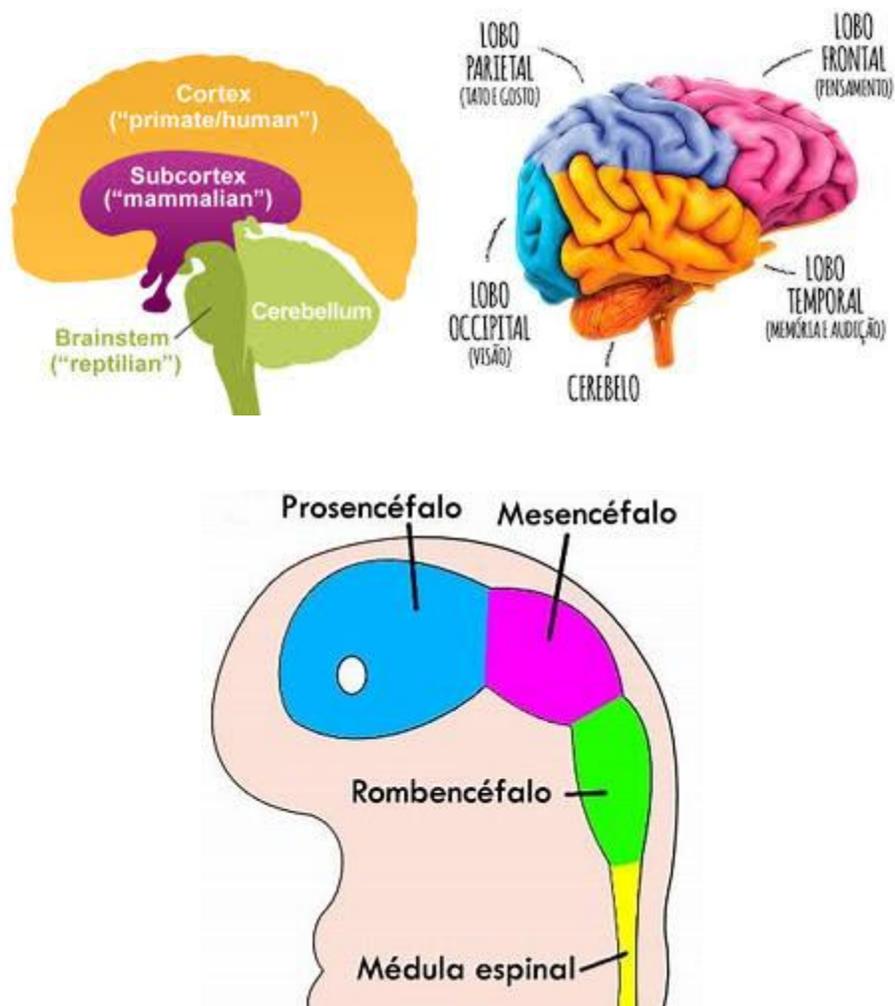
A neurociência da consciência

Wayne Wu assina o verbete “A neurociência da consciência”, da *Stanford Encyclopedia of Philosophy*. No verbete, é observado que a experiência consciente nos seres humanos depende de atividade cerebral, razão pela qual a neurociência pode contribuir para explicar a consciência. A neurociência é entendida como o conjunto de campos científicos cujas explicações “aludem a propriedades de neurônios, populações de neurônios e a uma grande parte do sistema nervoso”.

A questão da consciência pode ser abordada de duas maneiras, para Wayne Wu: a genérica e a específica. A questão genérica procura simplesmente entender por que um determinado estado mental é consciente, em vez de inconsciente – em outras palavras, em que condições o sistema nervoso gera um estado mental consciente. A questão específica, por sua vez, entra no mérito do conteúdo do estado mental. Ela procura entender quais são as condições necessárias e suficientes para que determinado estado consciente possua um certo conteúdo.

Um desafio que se coloca à toda “ciência da consciência” é o fato de a consciência ser um fenômeno essencialmente subjetivo. O acesso à consciência depende, em larga medida, da introspecção e dos relatos subjetivos, que se tornam, assim, o “método de ouro” das ciências da consciência. Não obstante, esse método é geralmente visto como pouco fiável.

Aos interessados na neurociência da consciência, é desejável um mínimo de familiaridade com o cérebro. Wayne Wu oferece um mapa rudimentar do cérebro, enfatizando a diferença entre o córtex cerebral e o subcórtex. O córtex cerebral é dividido em dois hemisférios, esquerdo e direito, cada qual possuindo quatro lobos: o frontal, o parietal, o temporal e o occipital. O subcórtex divide-se em cérebro anterior (prosencefalo), cérebro médio (mesencefalo) e cérebro posterior (rombencefalo), e localiza-se abaixo do córtex cerebral.



Figuras 7, 8 e 9: Acima e à esquerda, uma representação do córtex humano (herdado dos primatas) e do subcórtex humano (herdado dos mamíferos). O tronco cerebral (herdado dos répteis) e o cerebelo complementam o encéfalo. Acima e à direita, uma representação dos lobos cerebrais e das funções cognitivas a que se relacionam: o lobo frontal (associado ao pensamento), o lobo parietal (associado ao tato e ao gosto), o lobo occipital (associado à visão) e o lobo temporal (associado à memória e à audição). Abaixo, uma representação do subcórtex de um embrião de 4 meses, com suas áreas principais: prosencéfalo, mesencéfalo e rombencéfalo. Imagens retiradas da Internet.

São presumivelmente muitas as propriedades neurais capazes de explicar o fenômeno mental. Os *spikes*, “uma grande mudança no potencial elétrico através da membrana celular de um neurônio, que pode ser transmitida entre os neurônios que formam um circuito neural”, nos dizeres de Wu, têm recebido especial atenção dos neurocientistas da consciência. *Spikes* carregam informações sobre estímulos, que são, em última instância, variáveis randômicas.

No verbete, Wayne Wu procura distinguir duas consciências: a “consciência de acesso” [*access consciousness*] e a “consciência fenomênica” [*phenomenal consciousness*]. A “consciência fenomênica” diz respeito à experiência subjetiva do indivíduo – para alcançá-la, é

apenas necessário ter uma experiência relevante. A “consciência fenomênica” da cor vermelha, por exemplo, é a experiência subjetiva do indivíduo, proporcionada por uma experiência visual relevante (no caso, ver a cor vermelha).

A mera experiência subjetiva do indivíduo não implica que este tenha aptidão para relatá-la. No entanto, se a experiência subjetiva do indivíduo puder ser utilizada no raciocínio, para um relato ou mesmo para o controle de uma ação, ela já será uma “consciência de acesso”, na distinção proposta por Wu.

Uma questão central colocada pelos filósofos (David Chalmers à frente) aos neurocientistas da consciência é o “problema difícil” [*hard problem*] da consciência. Wayne Wu elabora esse problema da seguinte forma: há uma limitação inerente às explicações empíricas dos fenômenos mentais. Ou bem as explicações empíricas são estruturais, ou bem elas são funcionais. Sucede que os fenômenos da consciência não são redutíveis nem a explicações estruturais nem a explicações funcionais. Logo, há algo que é deixado de fora pelas explicações empíricas da consciência. Queda, em seu lugar, uma “lacuna explicativa” [*explanatory gap*], na célebre expressão de Joseph Levine.

Outro tópico de destaque da “neurociência da consciência” é a busca pelos correlatos neurais da consciência. Wayne Wu observa, com correção, que os correlatos neurais não necessariamente explicam a consciência – descobrir um correlato neural é apenas o primeiro passo que o neurocientista deve dar na busca pelas explicações dos fenômenos mentais. Fundamentalmente, o neurocientista precisa mostrar que o correlato neural encontrado desempenha papel relevante para o evento mental em questão, o que não é tarefa banal.

Onde estariam localizados esses correlatos? A resposta depende da teoria adotada, segundo Wu. As “teorias de processamento recorrente” [*recurrent processing theories*] enfatizam as áreas sensoriais – no caso da visão, a parte de “trás” do cérebro. Adeptos da “teoria da integração da informação” [*information integration theory*] também se referem a uma “zona quente posterior” do cérebro, abarcando as áreas parietal e occipital. Por outro lado, para alguns teóricos do “pensamento de ordem superior” [*higher-order thought theorists*], a parte frontal do cérebro seria suficiente para a consciência. Por fim, para os defensores do “espaço de trabalho neuronal global” [*global neuronal workspace*], a consciência emerge globalmente no cérebro – nem a parte dianteira nem a parte de “trás” do cérebro seriam, sozinhas, suficientes para a sua ocorrência.

Dois defensores do status científico da consciência

O neurofisiologista australiano John Eccles (1903-97), agraciado com o Nobel de Medicina em 1963, frisava que os eventos mentais são “fatos da experiência” (1951) e que, como tais, são perfeitamente adequados para a investigação científica (1953).

O filósofo estadunidense John Searle, celebrado teórico dos atos de fala, reagia, já no final do século passado, à tese de que não poderia haver uma “ciência da consciência” – sob hipótese de que a ciência, por definição, seria objetiva, e a consciência, por definição, seria subjetiva – observando que (i) “nos escritos de Descartes (...) ‘objetivo’ tinha algo próximo do oposto do sentido atual. Em algum momento (...) a distinção ‘objetivo-subjetivo’ virou ao avesso” (Searle, 1998, p.381, tradução nossa), e que (ii) “o fato de a ciência perseguir objetividade no sentido epistêmico não nos deveria cegar para o fato de que há entidades ontologicamente subjetivas que devem ser matéria de investigação científica tanto quanto qualquer outro fenômeno biológico” (Searle, 1998, p. 382, tradução nossa).

Pessoa Jr. e a consciência como observação do cérebro

Em artigo de 2015, Osvaldo Pessoa Júnior define “observação” como “uma percepção, na qual se salienta um foco de atenção e que envolva a aquisição de conhecimento”. Tomada a palavra nesse sentido, o autor brasileiro defende que a *consciência* pode ser entendida como uma *observação do cérebro sobre si mesmo*.

Entre os exemplos mais evidentes de auto-observação cerebral, estariam o fenômeno do “fosfeno” (sensação de ver manchas vermelhas, a partir de uma estimulação da retina ou do córtex visual), as ilusões de ótica e as alucinações próprias de um indivíduo esquizofrênico. Em todos esses casos, os fenômenos observados estariam no interior do cérebro do indivíduo.

Extrapolando esses exemplos, Pessoa Jr. defende que toda observação é uma auto-observação do cérebro, pois nos processos finais (proximais) da cadeia de observação, esta se enrola “em um novelo de retroalimentação”. Em outras palavras, a cadeia causal da percepção termina, sempre, em “áreas de ordem superior” do cérebro, onde há robusta integração de informação, contemplando inúmeros circuitos de retroalimentação [*feedback*].

A consequência de “o cérebro observar a si mesmo” é a impossibilidade de se separar o sujeito do objeto. Pessoa Jr. o reconhece e, a fim de ilustrar esse fenômeno, apela a uma pinça fictícia que, se cutucasse uma região milimétrica de um subconjunto adequado do cérebro, provocaria, ao mesmo tempo, um estímulo desse subconjunto e uma alteração de consciência.

John Locke, ao caracterizar a “consciência introspectiva”, já afirmara que “a consciência é a percepção do que se passa na própria mente de um homem”. A diferença é que, para Pessoa Jr., que funda sua posição numa meta-teoria causal pluralista, o que é percebido é o cérebro, e não a mente.

Em seu artigo, o filósofo e físico paulistano relembra o lema de Marvin Minsky, “a mente é aquilo que o cérebro faz”, mas procura se afastar de concepção funcionalista, negadora da importância do substrato material do cérebro, porventura escondida no lema. O autor também rejeita concepções solipsistas, enfatizando que a sua postura é realista quanto à existência do mundo externo.

Lembrando a conhecida influência dos campos elétricos cerebrais sobre as atividades neurais, Pessoa Jr. acredita ser plausível que a consciência fenomênica tenha sua origem “na movimentação eletroquímica dos neurônios e do meio circundante”, de modo que os *qualia* correspondam a configurações espaço-temporais desses campos elétricos. O autor atesta a semelhança entre essa interpretação e as leis psicofísicas concebidas por David Chalmers, enfatizando, porém, que os campos elétricos são físicos, pelo que seriam parte das cadeias causais do mundo físico.

Sem se comprometer com um modelo específico, Pessoa Jr. fala em um “processo sigma-ômega” ($\Sigma\omega$) que ocorreria na região do cérebro “mais proximal para a consciência”, possivelmente nas áreas corticais superiores, onde há forte integração e *feedback* (nos últimos anos, o autor reviu essa posição, passando a adotar o localizacionismo subcortical).

No par sigma-ômega, ω se refere ao substrato material dos *qualia*, cujo mecanismo de funcionamento ainda é desconhecido, em função, muito provavelmente, de seguir princípios físico-químico-biológicos ainda não descritos pela ciência. Σ , por seu turno, se refere ao elemento organizacional (funcional) dos neurônios.

Para Pessoa Jr., a vivência de um *quale* corresponderia a um processo $\Sigma\omega$ específico, e esse processo seria estatisticamente determinista.

Metaestabilidade no cérebro

O dicionário online de química da *Oxford* define “estado metaestável” como a “condição de um sistema que possui uma estabilidade precária que pode ser facilmente perturbada” (*Metastable State, A Dictionary of Chemistry, Oxford*, tradução nossa). O verbete informa, na sequência, que o “estado metaestável” se diferencia do “estado mais estável” justamente pelo fato de poder ser reduzido a um nível de energia mais baixo quando submetido a uma pequeníssima perturbação.

No verbete, são oferecidos alguns exemplos dessa distinção (estado metaestável x estado estável), dos quais destaco o seguinte (com adaptações): um livro pesado “deitado” no centro de uma mesa estaria em estado de equilíbrio estável; um livro leve “em pé” na extremidade de uma mesa estaria em estado de equilíbrio metaestável.

Faça-se agora uma pequena digressão: na história das ciências do cérebro, prevaleceram, *grosso modo*, duas grandes correntes ou escolas: a que entende que há um conjunto de regiões distintas localizadas no córtex cerebral, responsáveis por tarefas exclusivas (vale dizer, cada região responderia por uma função específica); e a que entende que o cérebro é um órgão essencialmente integrado, não cabendo essas compartimentações funcionais.

A primeira escola pode ser chamada de *localizacionista*; a segunda, de *holista*. A oposição entre elas, patente, é capaz de separar caminhos de pesquisa desde pontos muito remotos. Mas há quem queira suplantá-la por uma sofisticada complementaridade (Scott Kelso e Tognoli, 2007).

Se, por um lado, as regiões cerebrais expressam a capacidade de autonomia e de execução de funções especializadas dos conjuntos neurais, submetidos a processos de segregação ou modularidade, por outro lado, também se verificam, no cérebro, tendências de coordenação global e acoplamento desses conjuntos. A coexistência de segregação e integração no cérebro pode ser explicada pelo mecanismo da metaestabilidade, postulado pela “teoria da dinâmica de coordenação” [*theory of coordination dynamics*].

Emmanuelle Tognoli e Scott Kelso (2014) são referências dessa abordagem. Não é meu objetivo aqui apresentar detalhadamente a teoria da dupla, de resto bastante complexa. Sinalizo, entretanto, seus aspectos gerais: (i) a metaestabilidade é entendida como o resultado de uma quebra de simetria na equação de fase relativa que representa a coordenação entre osciladores (não

lineares) não linearmente acoplados; (ii) em termos não matemáticos, metaestabilidade é a mistura da integração e da segregação de conjuntos neurais observados em várias escalas do cérebro; (iii) os processos de escolha podem estar relacionados com transições de regimes metaestáveis para regimes estáveis.

As condições para que ocorra metaestabilidade no cérebro são claramente preenchidas (presença de conjuntos neurais exibindo comportamento de oscilação espontânea e quebra de simetria). Isso não significa, entretanto, que o cérebro de fato opere nesse regime.

O problema da percepção

A dupla Tim Crane e Craig French assina o verbete “O problema da percepção” [*The problem of perception*], da *Stanford Encyclopedia of Philosophy*. No verbete, é notado que os fenômenos da ilusão e da alucinação dão ensejo a um tradicional problema filosófico: como a percepção pode constituir, a um só tempo, “uma abertura para o mundo e uma consciência do mundo”, se ela se sujeita a erros? Se as ilusões e as alucinações são possíveis, então a percepção, enquanto guia seguro para a compreensão do mundo, deve ser impossível. Esse é o problema da percepção.

Do ponto de vista filosófico, “ilusão” e “alucinação” designam coisas distintas. Uma ilusão é uma situação perceptual na qual um objeto físico (ou um evento físico) não é percebido da forma como realmente é. É o caso de uma parede branca vista sob projeção de luz amarela, que parecerá amarela, ou de um corpo celeste gigantesco visto da Terra, que parecerá pequeno.

Uma alucinação, por seu turno, é uma percepção de um objeto (ou evento), sem que haja, na realidade, esse objeto (ou evento) para ser percebido. É o que ocorre com as alucinações auditivas, que costumam apresentar-se sob a forma de sons inespecíficos (chiados, zumbidos, assoviros e ruídos).

No campo da “experiência perceptual”, vigoram dois conceitos-chave, de acordo com Crane e French: “abertura” [*openness*] e “consciência” [*consciousness*]. O conceito de “abertura” alude às circunstâncias em que o indivíduo experimenta como presentes alguns objetos independentes de sua mente. Por “abertura”, entende-se também que a natureza da experiência do indivíduo é imediatamente responsiva à natureza de objetos independentes de sua mente.

O conceito de “consciência” expressa as situações em que o indivíduo possui consciência perceptual dos objetos independentes de sua mente que ele experimenta como presentes, ou aos quais sua experiência responde. Em princípio, pode haver “abertura” sem “consciência”, mas não o contrário. Se uma experiência perceptual for “consciente”, certamente ela também será “aberta”, na medida em que aquele conceito engloba este.

Há uma gama de respostas possíveis para o problema da percepção. Crane e French abordam quatro delas. Em primeiro lugar, é apresentada a “teoria dos dados dos sentidos” [*the sense-datum theory*]. Para essa teoria, “a experiência [perceptual] é fundamentalmente uma relação com um objeto não ordinário; um dado do sentido”. A real presença dos dados dos sentidos e de suas qualidades na experiência explicariam a natureza da experiência.

Os autores mencionam, em seguida, a “teoria adverbial” [*the adverbial theory*], para a qual “a experiência [perceptual] não é relacional; ela é fundamentalmente uma questão de se representar objetos ordinários de certas maneiras”. As qualidades intrínsecas da experiência e os modos nos quais ela é modificada é que explicariam a natureza da experiência.

Em terceiro lugar, é apresentada a “teoria intencionalista” [*the intentionalist theory*]. Para essa teoria, a experiência também é “não relacional e fundamentalmente uma questão de se representar objetos ordinários de certas maneiras”. Mas o que explicaria o caráter da experiência seria a sua natureza representacional.

Por fim, Crane e French tratam da “teoria realista ingênua” [*the naïve realist theory*], segundo a qual “a experiência [perceptual] é fundamentalmente uma relação com aspectos ordinários da realidade independente da mente”. A real presença de aspectos ordinários da realidade independente da mente na experiência é que explicaria a natureza da experiência.

Aspecto singular da “teoria realista ingênua” (também chamada de disjuntivista) é a rejeição da “tese do tipo comum” [*common kind claim*], segundo a qual “experiências verídicas, ilusórias e alucinatórias são fundamentalmente as mesmas, formando um tipo comum”.

O verbete é encerrado com a constatação de que há um vívido debate na filosofia da percepção presentemente, opondo os intencionalistas, de um lado, aos “realistas ingênuos” (disjuntivistas), de outro.

Daniel Wegner: a experiência de livre-arbítrio como resultado de uma ilusão causal

Em célebre artigo (2003), Daniel Wegner defende que a experiência da vontade livre resulta de uma ilusão causal: a ilusão de que nossas ações são causadas por nossos pensamentos. Isso ocorreria sempre que convergissem três princípios: (i) o princípio da prioridade do pensamento em relação à ação; (ii) o princípio da consistência do pensamento em relação à ação; e (iii) o princípio da exclusividade do pensamento em relação à ação.

Em outras palavras, as pessoas experimentaríamos uma vontade livre sempre que (i) agissem depois de um pensamento relevante (princípio da prioridade), (ii) agissem de maneira compatível com o pensamento relevante (princípio da consistência), e (iii) agissem sem nenhuma outra causa aparente além do pensamento relevante (princípio da exclusividade). Fundamentalmente, de acordo com Wegner, esses princípios não seriam guias seguros para a conclusão de que os pensamentos são a causa das ações – casos de esquizofrenia, projeção e hipnose exibiriam as falhas desses princípios.

Em trabalho anterior (2002), Wegner comparara a vontade consciente à bússola de um navio (p. 317): da mesma forma que seria ilusório atribuir os movimentos de um navio a sua bússola, seria ilusório atribuir as ações humanas à vontade consciente. O autor sugerira a alcunha “epifenômeno” (i.e., fenômeno sem poder causal) para designar ambas as coisas – vontade consciente e bússola.

A posição de Wegner nega, em síntese, a crença ordinária de que as ações humanas são causadas pela vontade consciente.

Crítica de Beatriz Sorrentino Marques a Daniel Wegner

Em artigo de 2017, Beatriz Sorrentino Marques coloca um problema para a teoria da causalção mental aparente de Wegner: o potencial de prontidão (RP), assumido por Wegner como atividade preparatória para a ação, não o é de forma conclusiva. A teoria de Wegner teria como um de seus pilares a interpretação de Libet et al. (1983) sobre o papel desempenhado pelo potencial de prontidão (RP), mas haveria interpretações alternativas igualmente possíveis. Sorrentino Marques cita, entre outras, a interpretação de Trevena e Miller (2010).

A teoria da causação mental aparente de Wegner, recorde-se, afirma que a vontade consciente não desempenha nenhum papel na produção causal das ações humanas, embora os seres humanos tenham a ilusão de que seus pensamentos geram suas ações. É patente a separação operada por Wegner entre “experiência” e “causa”: o neurocientista estadunidense reconhece que temos a experiência da vontade consciente, mas nega que a vontade consciente seja a causa das nossas ações.

Sorrentino Marques aponta, de forma oportuna, para a vagueza do conceito “experiência da vontade consciente”, utilizado por Wegner. Em seu lugar, a filósofa propõe o conceito “experiência de origem da ação” [*experience of source of action* (ESA)], lançado por Haggard e Clarke (2003). Ter a experiência de origem da ação é ter a experiência de que a ação foi causada por nós mesmos, de forma proposital.

A última palavra sobre o papel desempenhado pelo potencial de prontidão (RP) ainda não foi dada. Trata-se de questão empírica aberta. Por ora, as evidências não nos permitiriam concluir que a experiência de origem da ação (ESA) não cumpre nenhum papel causal. Em outras palavras, não se pode descartar que a ESA seja a experiência consciente de um mecanismo real, relevante para a ação.

Outrossim, se a experiência de origem da ação fosse de fato ilusória, como quer Wegner, seria difícil justificar sua existência de uma perspectiva da teoria da evolução: “(...) seria estranho considerar que os seres humanos possuem experiências conscientes se elas não desempenham nenhuma função relevante” (Sorrentino Marques, 2017, p. 873, tradução nossa).

Sorrentino Marques reconhece que os seres humanos podem agir de forma automática, isto é, sem a experiência de origem da ação (ESA). Mas disso não se segue, para a autora, que a ESA, quando presente, seja ilusória.

Agência

Markus Schlosser é o autor do verbete “Agência” [*Agency*], da *Stanford Encyclopedia of Philosophy*. No verbete, o professor de Filosofia da *University College Dublin* afirma, talvez de forma redundante, que “‘agente’ é o ser que tem a capacidade de agir, e ‘agência’ é o exercício ou a manifestação dessa capacidade”.

No campo da filosofia da ação, vigoram uma concepção padrão de “agência” e uma teoria padrão de “ação”. “Agência”, via de regra, designa performance de ações intencionais. A intencionalidade da ação, por seu turno, é explicada em termos de causação mental, vale dizer, apelando-se aos estados e aos eventos mentais do agente.

Reforce-se: tanto na tradição filosófica (em Aristóteles e em David Hume, por exemplo) quanto na filosofia analítica contemporânea (mais notoriamente, em Gertrude Anscombe e em Donald Davidson), as ações costumam ser definidas em termos de intencionalidade. Em sentido inverso, a noção de “intencionalidade” parece fundamentar a noção de “ação”. Na obra de Davidson, é ainda postulada uma conexão próxima – pode-se dizer, de quase equivalência – entre “agir intencionalmente” e “agir por uma razão”.

No verbete, Schlosser nota que o conceito “agência” também pressupõe, no mais das vezes, “iniciação da ação pelo agente”. A concepção padrão explica essa iniciação aludindo aos putativos poderes causais dos estados e dos eventos mentais do agente. Por óbvio, não são todos os filósofos que adotam essa estratégia. Há, na literatura filosófica – ainda que de forma minoritária –, resistência à definição de “agência” como “ação intencional”.

Os que se opõem à concepção padrão costumam apontar para a “natureza inteiramente espontânea da agência”, que não demandaria razões ou intenções primárias. Para eles, o poder do agente para iniciar uma ação não pode ser reduzido a uma capacidade para agir intencionalmente e por razões.

A teoria padrão, ainda de acordo com as concepções críticas, demandaria muito dos agentes. Há, na natureza, seres capazes de ações genuínas que não possuem estados mentais representacionais. A “agência intencional” deveria, desse modo, ser entendida apenas como um tipo de agência (entre outros). Da mesma forma que há tipos de agência ainda mais refinados e elevados do que a agência preconizada pela teoria padrão – por exemplo, a agência que, além de intencionalidade, pressupõe autocontrole, autonomia e livre-arbítrio –, há também tipos mais básicos de agência, que não requerem representação de estados mentais.

De forma paralela a esse debate, viceja, na metafísica da agência, o tema da individuação da ação. Para abordá-lo, Schlosser parte do seguinte exemplo: um indivíduo aperta o interruptor, acende a luz, ilumina a sala e surpreende o ladrão. Quantas ações ele praticou? Para as “visões grosseiras (ou minimizadoras) sobre a individuação de ações” [*coarse-grained (or minimizing views)*], apenas uma ação foi praticada, e essa ação comporta diferentes descrições.

É outra a posição das “visões refinadas (ou maximizadoras) sobre a individuação de ações” [*fine-grained (or maximizing) views*], para as quais o número de ações depende do número de propriedades instanciadas (no caso, quatro).

Last but not least, mencione-se, no verbete de Schlosser, a temática da sensação de agência [*sense of agency*]. Na visão de muitos filósofos, e em particular na visão de Gertrude Anscombe, o conhecimento que um indivíduo tem sobre suas próprias ações não depende de observação ou inferência – em uma palavra, é um conhecimento “direto”. Não raro, ao agir, o indivíduo possui a sensação de controle e agência sobre sua ação. Essa percepção também é direta.

Sucedem que essa percepção do senso comum tem sido desafiada por neurocientistas. Na obra de Daniel Wegner, por exemplo, sobressai o “modelo de causação mental aparente”, que define a sensação de agência como o resultado de uma interpretação equivocada do indivíduo. Para o indivíduo, seus pensamentos geram suas ações. Essa percepção se deve, em grande medida, à antecedência dos pensamentos em relação às ações, e a eventuais compatibilidades entre ambos. As reais causas das ações seriam, no entanto, inconscientes e fisiológicas, e não acessadas diretamente pelo indivíduo, segundo o neurocientista estadunidense.

Beatriz Sorrentino Marques, a sensação de agência e o controle regulador

Em artigo recente (2021), Beatriz Sorrentino Marques defende que a sensação de agência (i.e., a “experiência pré-reflexiva da ação como causada por si mesmo”) não é evidência para o controle regulador [*regulative control*] (i.e., o controle que pressupõe a “capacidade para decidir de outro modo”).

No debate filosófico entre compatibilistas e libertarianistas, estes costumam apelar à sensação de agência para defender sua crença no livre-arbítrio (no sentido de possibilidades alternativas de escolha), lançando mão do argumento da introspecção.

Para Sorrentino Marques, argumentos mais robustos, que abordam os mecanismos que estão por trás da sensação de agência – em particular, o “modelo comparador” (um “mecanismo de controle que conta com uma cópia eferente e um *feedback* sensorial”) e a “percepção da relação entre a ação e eventos no ambiente” – indicariam, ao contrário, sua compatibilidade com o determinismo causal.

A análise dos mecanismos da sensação de agência, mais do que permitir uma interpretação compatibilista, denuncia, para Sorrentino Marques, vulnerabilidades das visões incompatibilistas. À guisa de ilustração, a autora observa que eventos posteriores à ação podem contribuir para a sensação de agência, o que parece entrar em conflito com a tese (libertarianista) de que a capacidade para decidir de outro modo é decorrente daquela sensação.

As razões que levariam alguns a associar a sensação de agência ao indeterminismo causal seriam: o fato de eles interpretarem a sensação de agência à luz de crenças prévias à ação; o fato de eles considerarem apenas as crenças das quais estão conscientes; e o fato de eles entenderem a “experiência de si mesmo” [*experience of oneself*] como causa da ação.

Ciosa do bom uso dos termos filosóficos, Sorrentino Marques distingue a “sensação de agência” [*the sense of agency*], acima definida, da “experiência de escolha” [*the experience of choice*], que constitui uma ação mental. A “experiência de escolha” pode ocorrer muito antes da “sensação de agência”, i.e., sem que nenhuma ação seja efetivamente realizada (é o que sucede, no exemplo de Sorrentino Marques, quando uma pessoa escolhe a comida que irá preparar para um jantar na semana seguinte).

No que toca à “experiência de escolha”, a filósofa brasileira endossa a abordagem compatibilista de Oisín Deery. No que se refere à “sensação de agência”, as considerações de Sorrentino Marques (de teor compatibilista) estão na vanguarda do debate.

Por fim, frise-se que, para a filósofa brasileira, a sensação de agência relaciona-se não ao controle regulador [*regulative control*], mas ao controle de orientação [*guidance control*], que não necessariamente pressupõe capacidade para decidir de outro modo.

A ciência da mente e do encéfalo

O neologismo *mindbrain* (mente-encéfalo) foi provavelmente cunhado pelo psicólogo neozelandês John Money, que o empregou no livro *Evolutionary sexology: the hypothesis of song and sex* (1997). O estudioso dos sonhos John Allan Hobson já utilizara a expressão inversa *brainmind* (encéfalo-mente) no seu célebre “The dreaming brain...” (1988). Mais recentemente, o psicólogo e psicanalista americano Mark Blechner adotou o termo de Money no título de seu livro (2018), e o filósofo e físico Osvaldo Pessoa Júnior introduziu o vocábulo “mentencéfalo” na língua portuguesa, em disciplina de pós-graduação na Universidade de São Paulo (2019).

Na página de apresentação da disciplina, Pessoa Júnior oferece uma rápida descrição da ciência do mentencéfalo. Trata-se de ciência interdisciplinar, que abarca elementos da neurociência, da ciência cognitiva e da filosofia da mente, e que tem como objeto de estudo o fenômeno da consciência. Para o filósofo e físico brasileiro, a próxima revolução científica de larga escala poderá ocorrer no seio da ciência do mentencéfalo. Questões empíricas como a localização dos correlatos neurais imediatos da consciência – estão eles localizados numa região cortical ou subcortical específica, ou a consciência emerge globalmente no encéfalo? –, uma vez respondidas, suscitariam essa revolução.

A expressão “mentencéfalo”, empregada por Pessoa Junior, é deveras próxima de “metencéfalo”, parte do encéfalo embrionário que, a partir do terceiro mês, passa a se desenvolver na ponte e no cerebelo. Para evitar confusões, prefiro a forma não contraída “ciência da mente e do encéfalo”.

Comentários finais (Bloco VII)

No que se refere ao tema do livre-arbítrio, acredito que a ciência da mente e do encéfalo possa ser de grande valia. Para mim, sua maior contribuição seria fugir de questões suficientemente antigas (do tipo “o livre-arbítrio é compatível com o determinismo causal?”) e encetar um novo debate. Indagações como “o que acontece com uma pessoa quando ela diz que está experimentando o livre-arbítrio? Há algo que muda em sua mente? Há algo que muda em seu encéfalo?” parecem-me bastante promissoras, no sentido de permitirem respostas a uma ciência que não descure nem os aspectos mentais nem os aspectos fisiológicos que caracterizam a experiência humana.

Em suma, é à ciência da mente e do encéfalo que me refiro quando menciono, no título deste bloco temático, “uma ciência para a sensação de livre-arbítrio”.

A seguir, enuncio algumas proposições sobre a sensação de livre-arbítrio, resultantes do processo reflexivo que empreendi nos últimos anos. Na hipótese de a ciência da mente e do encéfalo tornar-se paradigmática (ainda que sob um novo nome), cuja probabilidade não sou capaz de mensurar, e na hipótese de esta pesquisa (ou sua versão em inglês, a ser elaborada) chegar ao conhecimento de algum cientista da mente e do encéfalo, cuja probabilidade é seguramente ínfima, têm-se aqui, creio eu, ideias que podem ser parcialmente testadas.

A sensação de livre-arbítrio não precisa ser confirmada por movimentos

Por “sensação de livre-arbítrio”, entendo, a princípio, apenas a “experiência de escolha”, e não a “sensação de agência”, para usar a distinção proposta por Beatriz Sorrentino Marques (2021).

Afinal, em algumas ocasiões, pode-se escolher uma opção e, nos instantes finais, vetar essa escolha, impedindo a ação – e é razoável que esse fenômeno do “veto” (em que há escolha mental/encefálica, mas não há agência corporal) constitua igualmente uma experiência subjetiva de livre-arbítrio.

Ademais, há aquilo que os gregos chamavam de *akrasia* (“fraqueza da vontade”), situação em que o indivíduo age contra o seu melhor juízo. É o que sucede com um fumante que sabe que o melhor a fazer é largar o cigarro, mas que, ainda assim, continua fumando. Nesse caso, a “sensação de agência” está em franca oposição com a “sensação de livre-arbítrio”: a ação mina o livre-arbítrio do indivíduo, que tem consciência de que deliberou previamente no sentido contrário (ele decidiu não mais fumar, mas continua fumando).

Quando não há o fenômeno do veto (e isso acontece, a meu ver, na maior parte das vezes), os seres humanos deliberam e agem de maneira compatível com essa deliberação. Aqui, é ainda mais evidente a *dispensabilidade* da “sensação de agência” (que pressupõe, reforça-se, movimento corporal) para a definição da “sensação de livre-arbítrio” (que é a mera resolução de um processo decisório em seus níveis mental e encefálico), pois a ação motora apenas confirma o que o processo decisório já deliberou.

Se isso está certo, experimentos científicos sobre o livre-arbítrio não precisam envolver ações motoras (e muito menos medi-las). Não é preciso, por exemplo, que se peça ao indivíduo que flexione o seu punho, ou que aperte um botão, para aferir a sua sensação de livre-arbítrio. Seria mais interessante, penso eu, colocar o indivíduo em uma situação de dilema (e.g., diante de um problema lógico de difícil resolução), medindo, a todo instante, o nível de sua atividade encefálica, e tentar “adivinhar”, apenas a partir desse nível, o momento em que ele resolveu o problema. Esse será, para mim, o momento em que ele teve a “sensação de livre-arbítrio” – que é uma sensação de *autoria e controle* sobre a decisão.

Autoria e agência são noções distintas

Autoria é uma noção relacionada ao que chamei de “segundo sentido filosófico de livre-arbítrio” (o meu preferido), que é o de “origem ou determinação das decisões pela vontade consciente” (o primeiro sentido é o de possibilidades alternativas, ou “capacidade para decidir de outro modo”).

Se a minha vontade consciente é a *autora* das minhas decisões, isso significa que ela (minha vontade consciente) é a *origem* delas (das minhas decisões) e que as *determina* (determina as minhas decisões), o que é o mesmo que dizer que ela (minha vontade consciente) tem *controle* sobre elas (sobre as minhas decisões).

Outro seria o sentido da frase: “a minha vontade consciente é o *agente* (por trás) das minhas ações”. Nesse caso, pressupõe-se *agência* da vontade consciente sobre o corpo. Pressupõem-se, ademais, “movimento corporal” (causado pela vontade consciente) e “causação descendente” (causação da mente sobre o corpo).

É claro que uma decisão livre (no sentido de exemplificar o livre-arbítrio), se existir, pode culminar em movimentos corporais. Em outras palavras, a noção de agência não é incompatível com a noção de livre-arbítrio (exceto na situação de *akrasia*), pois, depois de tomada uma deliberação, normalmente o que ocorre é a sua efetivação. Mas – e isto é importante frisar – a noção de agência tampouco é *necessária* para a noção de livre-arbítrio. Para que haja livre-arbítrio, basta haver deliberação (nas condições expostas nas proposições seguintes), com ou sem ação motora subsequente, vale dizer, com ou sem *agência*.

No plano das sensações (ou da experiência subjetiva), que é o que me interessa, o raciocínio é o mesmo. Para haver “sensação de livre-arbítrio”, basta haver “sensação de *autoria e controle* sobre a decisão”. Nesse caso, a “sensação de agência”, embora não seja incompatível com a “sensação de livre-arbítrio” (na situação de *akrasia*, reforça-se, há incompatibilidade), é claramente dispensável para as análises.

Em uma palavra, penso que a noção de *autoria* está para as *decisões* assim como a noção de *agência* está para as *ações*. Ambas designam que a *origem* de alguma coisa (da decisão, num caso; da ação, no outro) está na vontade consciente. Para a sensação de livre-arbítrio, a noção mais relevante é a de *autoria*, pois o livre-arbítrio sempre diz respeito a uma *decisão*

(mental/encefálica), mas nem sempre diz respeito a uma *ação* (que envolva movimentos corporais macroscópicos).

A sensação de livre-arbítrio é resultante de um processo em duas etapas

A sensação de livre-arbítrio é resultante de um processo decisório em duas etapas: a etapa da ordenação de preferências, que pode ser inconsciente; e a etapa da deliberação, que é necessariamente consciente.

A etapa da ordenação de preferências antecede a etapa da deliberação – em primeiro lugar, as opções precisam ser produzidas e ordenadas; em segundo lugar, é preciso que o indivíduo delibere pela opção preferida.

A sensação de livre-arbítrio demanda algum grau de reflexão racional

A sensação de livre-arbítrio é uma propriedade das decisões refletidas ou conscientes. Uma decisão irrefletida ou inconsciente não pode ser subjetivamente experimentada como livre; talvez ela sequer possa ser experimentada como uma “decisão”, na medida em que “decisão” significa “resolução tomada após juízo” (dicionário online Google, ênfase nossa).

Negar essa proposição implicaria considerar que atos inconscientes (ou irrefletidos) podem ser subjetivamente experimentados como livres, vale dizer, que esses atos também resultariam de um processo em duas etapas, que tem na ordenação de preferências sua etapa inicial, e na deliberação sua etapa final.

Independentemente de como se dê, no caso humano, o processo de formação de gostos e preferências (primeira etapa), “deliberar” é “decidir(-se), após reflexão e/ou consultas” (dicionário online Google, ênfase nossa). Resulta claro que a sensação de livre-arbítrio, por pressupor “deliberação” (segunda etapa), demanda algum grau de reflexão racional.

A sensação de livre-arbítrio ocorre em graus

Com efeito, nem todas as decisões subjetivamente experimentadas como livres o são da mesma forma – i.e., há decisões que são experimentadas como mais livres do que outras, e há

decisões que são experimentadas como menos livres do que outras. Particularmente, decisões que envolvem maior reflexão racional são experimentadas como mais livres do que decisões que envolvem menor reflexão racional, tudo o mais constante.

Negar essa proposição implicaria considerar que os graus de experiência de livre-arbítrio são sempre os mesmos, a despeito do que nos informa a introspecção, ou que a reflexão racional é irrelevante para a experiência de livre-arbítrio, a despeito da proposição anterior.

Exemplo

Numa questão de múltipla escolha de um concurso público, um candidato pode conscientemente escolher uma opção depois de uma maior ou de uma menor reflexão racional (“maior” ou “menor” reflexão racional significam, aqui, nível de atividade encefálica, e não “maior” ou “menor” tempo para deliberação – embora, eventualmente, essas coisas possam coincidir). Se a questão lhe parecer fácil, o grau de reflexão racional demandado será menor, e sua experiência de livre-arbítrio será menos intensa (nesse caso, o candidato atribuirá sua decisão à obviedade da resposta – ele estará “preso” à resposta evidente).

Se a questão lhe parecer difícil, o grau de reflexão racional demandado será maior, e sua experiência de livre-arbítrio será mais intensa (nesse caso, o candidato atribuirá sua decisão ao processo reflexivo que empreendeu – ele não estará “preso” a nenhuma resposta evidente).

No limite, se a questão lhe parecer muito difícil, a ponto de ser necessário “chutar” a resposta, sua experiência de livre-arbítrio será mínima (se houver alguma). Nesse caso, ou o grau de reflexão racional será muito baixo (ou nenhum) (ao ler a questão e dar-se conta de que ignora o tema, o candidato simplesmente desiste de pensar a respeito), ou o grau de reflexão racional será muito alto (heroicamente, o candidato vasculha sua razão para descobrir alguma resposta, mas acaba não encontrando nada).

A possibilidade de haver uma experiência de livre-arbítrio mínima com um alto grau de reflexão racional é reveladora de que a reflexão racional é “condição necessária, mas não suficiente” para a experiência de livre-arbítrio – i.e., além da reflexão racional, a sensação de livre-arbítrio deve estar ligada a outras variáveis (que serão exibidas nas proposições seguintes).

Primeira leitura conjunta das proposições anteriores

Quando as proposições anteriores são lidas conjuntamente, tem-se o seguinte cenário: com tudo o mais constante, (i) há um espectro de decisões; (ii) algumas decisões não são subjetivamente experimentadas como livres (é o caso das “decisões” irrefletidas); (iii) algumas decisões são subjetivamente experimentadas como pouco livres (é o caso das decisões pouco refletidas); (iv) algumas decisões são subjetivamente experimentadas como mais ou menos livres (é o caso das decisões mais ou menos refletidas); e (v) algumas decisões são subjetivamente experimentadas como muito livres (é o caso das decisões muito refletidas).

A sensação de livre-arbítrio corresponde a uma alteração mental

O livre-arbítrio é experimentado como uma alteração de estado mental. Mais precisamente, ele é experimentado como a passagem do estado mental da *ordenação de preferências* para o estado mental da *deliberação* (ou do estado mental da *dúvida* para o estado mental da *certeza*, ou do estado mental da *indefinição* para o estado mental da *definição*, ou de qualquer par análogo).

Por conseguinte, o livre-arbítrio jamais será experimentado como manifestação de um estado mental único (ordenação de preferências *ou* deliberação), mas sempre como um contraste entre dois estados subsequentes (ordenação de preferências *e* deliberação).

Afirmar o contrário seria aceitar a experiência do livre-arbítrio associada apenas ao estado mental da ordenação de preferências (ou seus análogos), ou associada apenas ao estado mental da deliberação (ou seus análogos).

Um indivíduo que esteja permanentemente em dúvida, sem nunca deliberar, não pode ter a experiência de livre-arbítrio, dado que não toma decisão nenhuma. Sucede o mesmo com o indivíduo que apenas delibera, sem levantar dúvida sobre nada. Nesse caso, o indivíduo age maquinalmente, e não possui a sensação de livre-arbítrio.

A sensação de livre-arbítrio corresponde a uma alteração encefálica

A sensação de livre-arbítrio corresponde a uma alteração de estado encefalo. Mais precisamente, um indivíduo experimenta o livre-arbítrio quando o seu encefalo deixa de funcionar

em regime metaestável, com grande integração dos conjuntos neurais, e passa a funcionar em regime mais estável, com segregação dos conjuntos neurais, a partir de uma “quebra de simetria”.

Pela definição de metaestabilidade, a passagem do regime metaestável para o regime mais estável decorre de uma pequeníssima perturbação no encéfalo, da qual o indivíduo, presumivelmente, sequer toma nota (sequer tem consciência).

Essa proposição é compatível com a situação de dilema, em que o indivíduo possui razões igualmente fortes para deliberar por mais de uma opção. Em situações desse tipo, qualquer fator pequeníssimo pode fazer o indivíduo pender para um lado ou outro.

Enquanto o indivíduo não delibera, há um alto nível de atividade em seu córtex, detectável pela eletroencefalografia. O seu encéfalo encontra-se em regime metaestável, com funcionamento integrado das redes neurais. Uma vez ocorrida a deliberação, o encéfalo passa a operar em regime mais estável: há uma diminuição do nível de atividade no córtex, com a desintegração das redes neurais.

A passagem do regime metaestável para o regime mais estável do encéfalo a partir de uma pequeníssima perturbação corresponde, pois, à sensação de livre-arbítrio.

A sensação de livre-arbítrio expressa os graus de dúvida e de certeza

A sensação de livre-arbítrio expressa os graus de dúvida e de certeza envolvidos no processo decisório. Levando em conta as etapas sucessivas que constituem a sensação de livre-arbítrio – *ordenação de preferências e deliberação* –, quanto maior for a distância entre a dúvida, própria da etapa inicial, e a certeza, própria da etapa final, mais intensa será a experiência de livre-arbítrio.

Em outras palavras, o processo decisório que assuma a forma “no início, muita dúvida; no final, muita certeza” é o que mais potencializa a sensação de livre-arbítrio.

Processos decisórios que assumam formas alternativas, como “no início, pouca dúvida; no final, muita certeza”, ou “no início, muita dúvida; no final, pouca certeza”, ainda que proporcionem alguma experiência de livre-arbítrio ao indivíduo, o fazem de maneira menos intensa.

Filósofos, via de regra, são treinados para questionarem aquilo que muitos acreditam ser “verdades incontestáveis”. A Filosofia, afinal, é a atividade que cultiva (e cultua) a dúvida, não raro partindo do que se considerava “a absoluta certeza”. Tem-se um caso curioso: o filósofo,

provavelmente, é o sujeito que, em sua profissão, menos experimenta o livre-arbítrio, pois, em vez de partir da dúvida e chegar à certeza, percorre o caminho inverso.

Não surpreende que uma das mais célebres frases filosóficas já proferidas, “sei que nada sei”, possa implicar um estado subjetivo de *desconforto* para a tomada de deliberações (se eu não sei de nada, qualquer decisão minha soará arbitrária). Trata-se, pois, do avesso da sensação de livre-arbítrio, que pressupõe *autoria e controle*.

Segunda leitura conjunta das proposições anteriores

Numa segunda leitura conjunta das proposições anteriores, chega-se ao seguinte cenário: as decisões que envolvem um maior grau de reflexão racional são experimentadas como mais livres do que as decisões que envolvem um menor grau de reflexão racional, tudo o mais constante.

As decisões que inicialmente geram muitas dúvidas são experimentadas como mais livres do que as decisões que inicialmente geram poucas dúvidas, tudo o mais constante; as decisões que inicialmente geram muitas dúvidas, e que acabam sendo tomadas com maior grau de certeza, são experimentadas como mais livres do que as decisões que inicialmente geram muitas dúvidas, e que acabam sendo tomadas com menor grau de certeza, tudo o mais constante.

Em resumo, uma decisão que envolva um alto grau de reflexão racional, mas que produza uma pequena certeza (e.g., um “chute”, depois de muita reflexão, numa prova de múltipla escolha) será experimentada como pouco livre; uma decisão tomada com muita certeza, mas que envolva um pequeno grau de reflexão racional (e.g., uma resposta óbvia numa prova de múltipla escolha) também será experimentada como pouco livre.

Uma decisão tomada com muita certeza e que envolva um alto grau de reflexão racional (e.g., uma resposta convicta, depois de muita reflexão, numa prova de múltipla escolha) será experimentada como muito livre. Com efeito, as decisões do último tipo (muita dúvida, seguida de muita certeza) são as que mais fortalecem a experiência de livre-arbítrio.

CONSIDERAÇÃO FINAL

A minha escolha por estudar o tema do livre-arbítrio foi livre?

As considerações desta tese não me permitiriam asserir que a minha escolha por estudar o tema do livre-arbítrio foi livre, pela simples razão de que eu não provei a existência do livre-arbítrio.

Mas é certo que, ao escolher estudar o tema do livre-arbítrio, eu tive a “sensação de livre-arbítrio”, e que essa sensação correspondeu à passagem de um estado mental da *dúvida* para um estado mental da *certeza*. Deliberei, e sobre essa deliberação senti-me *autor* e *controlador*. Ademais, acredito ter havido, simultaneamente, uma alteração de um *estado metaestável* para um *estado mais estável* do meu encéfalo.

Ao fim e ao cabo, continuo sem saber se possuo livre-arbítrio. Mas já se ganha, um tanto, em compreender melhor o que não sabemos.

REFERÊNCIAS BIBLIOGRÁFICAS

ANSELM. The Philosophical Dialogues: On Truth, On Freedom of Choice, On the Fall of the Devil. Translated by Thomas Williams. Hackett Publishing Company, 2002.

ARAUJO, Saulo. O Eterno Retorno do Materialismo: Padrões Recorrentes de Explicações Materialistas dos Fenômenos Mentais. Revista de Psiquiatria Clínica, 40 (3), São Paulo, 2013.

BICKLE, John et al. The Philosophy of Neuroscience. Stanford Encyclopedia of Philosophy. Acesso em 05/03/2019.

BLECHNER, Mark. The Mindbrain and Dreams: An Exploration of Dreaming, Thinking, and Artistic Creation. New York: Routledge, 2018.

BODE, Stefan et al. Tracking the Unconscious Generation of Free Decisions Using Ultra-HighField fMRI. PLOS ONE, 6 (6), pp. 1-13, 2011.

BRODY, Nathan; OPPENHEIM, Paul. Application of Bohr's Principle of Complementarity to the Mind-Body Problem. The Journal of Philosophy, 66 (4), pp. 97-113, 1969.

BROWN, Theodore. Making Truth: Metaphor in Science. Urbana: University of Illinois Press, 2003.

CRANE, Tim; FRENCH, Craig. The Problem of Perception. Stanford Encyclopedia of Philosophy. Acesso em 07/08/2020.

DAWKINS, Richard. The God Delusion. Boston, Massachusetts: Houghton Mifflin, 2006.

ECCLES, John. Hypotheses Relating to the Brain-Mind Problem. Nature, 168, pp. 53-57, 1951.

ECCLES, John. The Neurophysiological Basis of Mind: The Principles of Neurophysiology. Oxford: Clarendon Press, 1953.

FILEVICH, Elisa et al. There Is No Free Won't: Antecedent Brain Activity Predicts Decisions to Inhibit. PLOS ONE, 8 (2), 2013.

FRITH, Chris. Making Up the Mind: How the Brain Creates our Mental World. Malden, Massachusetts: Blackwell Publication, 2007.

GAZZANIGA, Michael. Facts, Fictions and the Future of Neuroethics. In: J. Illes (Editor), *Neuroethics: Defining the Issues in Theory, Practice, and Policy*. Oxford: Oxford University Press, pp. 141-148, 2006.

GINGRICH, Paul. Class Notes: Social Facts and Suicide. University of Regina. Department of Sociology and Social Studies, 1999.

GOMES, Gilberto. Free Will, the Self, and the Brain. Behavioral Sciences and the Law, 25, pp. 221-234, 2007.

GOMES, Gilberto. Preparing to Move and Deciding not to Move. Consciousness and Cognition, 19, pp. 457–459, 2010.

GOMES, Gilberto. Que Tipo de Determinação é Compatível com que Tipo de Liberdade? Unisinos Journal of Philosophy, 20 (2), pp. 113-127, 2019.

GREENE, Joshua. Social Neuroscience and the Soul's Last Stand. In: A. Todorov, S. Fiske, D. Prentice (Editors), *Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind*. Oxford: Oxford University Press, pp. 263–273, 2006.

HAGGARD, Patrick; EIMER, Martin. On the Relation Between Brain Potentials and the Awareness of Voluntary Movements. Experimental Brain Research, 126, pp. 128-133, 1999.

HAGGARD, Patrick; LIBET, Benjamin. Conscious Intention and Brain Activity. Journal of Consciousness Studies, 8 (11), pp. 47-63, 2001.

HOBBS, Thomas. The Questions Concerning Liberty, Necessity, and Chance. In: *The English Works of Thomas Hobbes of Malmesbury*. London: John Bohn, 1841.

HOBSON, John Allan. The Dreaming Brain: How the Brain Creates Both the Sense and the Nonsense of Dreams. New York: Basic Books, 1988.

HUME, David. Tratado de la Naturaleza Humana: Libro II, Parte Tercera, Sección II [1739]. Edición Electrónica: Diputación de Albacete, Servicio de Publicaciones, Gabinete Técnico, 2001.

ISHAM, Eve et al. Deliberation Period During Easy and Difficult Decisions: Re-examining Libet's "Veto" Window in a More Ecologically Valid Framework. Neuroscience of Consciousness, 2017.

KANE, Robert. Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism. The Journal of Philosophy, 96 (5), pp. 217-240, 1999.

LIBET, Benjamin et al. Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness Potential): The Unconscious Initiation of a Freely Voluntary Act. Brain, 106, pp. 623-642, 1983.

LIBET, Benjamin. Do We Have Free Will?. Journal of Consciousness Studies, 8, pp. 47-57, 1999.

LIBRE ARBITRE. In: *Wikipedia*. Acesso em 09/12/2020.

LIVRE-ARBÍTRIO. In: *Dicionário Priberam da Língua Portuguesa*. Acesso em 09/12/2020.

LIVRE-ARBÍTRIO. In: *Dicionário Online de Português*. Acesso em 09/12/2020.

LIVRE-ARBÍTRIO. In: *Michaelis Online*. Acesso em 09/12/2020.

MCKENNA, Michael; COATES, D. Justin. Compatibilism. Stanford Encyclopedia of Philosophy. Acesso em 01/05/2019.

MONEY, John. Evolutionary Sexology: The Hypothesis of Song and Sex. Medical Hypotheses, 48, pp. 399-402, 1997.

MUDRIK, Liad; MAOZ, Uri. Me and My Brain: Exposing Neuroscience's Closet Dualism. Journal of Cognitive Neuroscience, 27 (2), pp. 211–221, 2014.

MURCHO, Desidério. Livre-Arbítrio. Estado da Arte. São Paulo: Jornal O Estado de São Paulo. Publicado em 16/09/2020.

O'CONNOR, Timothy. Persons and Causes: The Metaphysics of Free Will. Oxford: Oxford University Press, 2000.

O'CONNOR, Timothy; FRANKLIN; Christopher. Free Will. Stanford Encyclopedia of Philosophy. Acesso em 10/11/2018.

ORTEGA Y GASSET, José. O Que é Filosofia? Campinas (São Paulo): Vide Editorial, 2016.

PEREBOOM, Derk. Determinism Al Dente. Noûs, 29, pp. 21-45, 1995.

PESSOA JUNIOR, Osvaldo. A Consciência Enquanto Observação do Cérebro. In: *História e Filosofia da Neurociência*, São Paulo: Editora Liberars, pp. 201-19, 2015.

PESSOA JUNIOR., Osvaldo; MELO, Luma. O Dualismo Interacionista Não Precisa Violar Leis De Conservação da Física. In: *Filosofia da Mente, Ciência Cognitiva e o Pós-Humano: Para Onde Vamos?* São Paulo: Filoczar, 2015.

PESSOA JUNIOR, Osvaldo. Philosophy and History of MindBrain Science. Curso de Pós-Graduação, Departamento de Filosofia, Universidade de São Paulo (USP), 2019.

POPPER, Karl. Materialism Criticized. In: K. Popper, J. Eccles: *The Self and its Brain. An Argument for Interactionism*. New York: Routledge, pp. 51-99, 1983.

PRIEST, Stephen. Mind-Brain Interaction. Encyclopedia of Science and Religion. Acesso em 06/01/2021.

ROBB, David. Moral Responsibility and the Principle of Alternative Possibilities. Stanford Encyclopedia of Philosophy, 2020.

SANTO AGOSTINHO. O Livre-Arbítrio. São Paulo: Paulus Editora, 1997.

SCHLOSSER, Markus. Agency. Stanford Encyclopedia of Philosophy. Acesso em 01/07/2020.

SCOTT KELSO, J.A.; TOGNOLI, Emmanuelle. Toward a Complementary Neuroscience: Metastable Coordination Dynamics of the Brain. *Neurodynamics of Cognition and Consciousness*, pp. 39-59, 2007.

SEARLE, John. How to Study Consciousness Scientifically. *Brain Research Reviews*, 26, pp. 379-87, 1998.

SHALEV, Baruch Aba. 100 Years of Nobel Prizes. Americas Group, 2002.

SORRENTINO MARQUES, Beatriz. Different Kinds of Decisions and an Experiment on Unconscious Generation of Free Decisions: a Conceptual Analysis. *Filosofia Unisinos*, 16 (1), pp. 44-57, 2015.

SORRENTINO MARQUES, Beatriz. An Issue for Wegner's Theory about the Conscious Will: the Readiness Potential does not Conclusively Represent Preparation for an Action. *Veritas*, 62 (3), pp. 860-876, 2017.

SWOYER, Chris; ELLIS, Stephen. Rational Choice. New Dictionary of the History of Ideas. Acesso em 13/01/2021.

TALBERT, Matthew. Moral Responsibility. Stanford Encyclopedia of Philosophy, 2019.

TOGNOLI, Emmanuelle; SCOTT KELSO, J.A. The Metastable Brain. Neuron, 81 (1), pp. 35-48, 2014.

TREVENA, Judy Arnel; MILLER, Jeff. Brain Preparation Before a Voluntary Action: Evidence Against Unconscious Movement Initiation. Consciousness and Cognition, 19 (1), pp. 447-456, 2010.

VAN INWAGEN, Peter. An Essay on Free Will. Oxford: Oxford University Press, 1983.

VIHVELIN, Kadri. Arguments for Incompatibilism. Stanford Encyclopedia of Philosophy, 2017

VOGT, Carl. Physiologische Briefe für Gebildete aller Stände. In: D. Wittich (Editor): *Schriften zum kleinbürgerlichen Materialismus in Deutschland*. Berlin: Akademie-Verlag, pp.1-24, 1971.

WEGNER, Daniel. The Illusion of Conscious Will. Cambridge: MIT Press, 2002.

WEGNER, Daniel. The Mind's Best Trick: How We Experience Conscious Will. Trends in Cognitive Sciences, 7 (2), pp. 65-69, 2003.

WIKIPEDIA. Necessity and Sufficiency. Acesso em 30/06/2021.

WOLF, Susan. Freedom Within Reason. Oxford: Oxford University Press, 1993.

WU, Wayne. The Neuroscience of Consciousness. Stanford Encyclopedia of Philosophy. Acesso em 05/03/2020.