LAMPIÃO

REVISTA DE FILOSOFIA

DOSSIÊ FILOSOFIA DA MENTE & CIÊNCIAS COGNITIVAS

PARTE 1

VOLUME 04
NÚMERO 1
2023
ISSN: 2675-9659


PPG
FIL
PROGRAMA DE
PÓS-GRADUAÇÃO
EM FILOSOFIA DA UFAL

LAMPIÃO

Revista do Programa de Pós-graduação em Filosofia da UFAL
ISSN: 2675-9659
Publicação anual

Endereço
Programa de Pós-graduação em Filosofia
Universidade federal de Alagoas – Campus A. C. Simões.
Instituto de Ciências Humanas, Comunicação e Artes
Av. Lourival Melo Mota, s/n. Tabuleiro dos Martins. Maceió, Al.
CEP: 57.072-970

E-mail
lampiaorevista@gmail.com

Sumário

Página

O ACENDEDOR DE LAMPIÕES

Jorge de Lima

Lá vem o acendedor de lampiões da rua!
Este mesmo que vem infatigavelmente,
Parodiar o sol e associar-se à lua
Quando a sombra da noite enegrece o poente!

Um, dois, três lampiões, acende e continua
Outros mais a acender imperturbavelmente,
À medida que a noite aos poucos se acentua
E a palidez da lua apenas se pressente.

Triste ironia atroz que o senso humano irrita: —
Ele que doira a noite e ilumina a cidade,
Talvez não tenha luz na choupana em que habita.

Tanta gente também nos outros insinua
Crenças, religiões, amor, felicidade,
Como este acendedor de lampiões da rua!

COUTINHO, Afrânio (org.). Jorge de Lima. Obra Completa. Rio de Janeiro: Editora José Aguilar, 1958, vol. I. p.208

APRESENTAÇÃO

O homem é visivelmente feito para pensar.
Aí reside toda a sua dignidade e todo o seu mérito,
e todo o seu dever é pensar com acerto.

(Pascal, Pensamentos Escolhidos, 60)

É com grande alegria e satisfação que publicamos estas duas partes do **Dossiê Filosofia da Mente & Ciências Cognitivas** pela *Lampião – Revista de Filosofia*, vinculada ao Programa de Pós-Graduação em Filosofia (PPGFil) da Universidade Federal de Alagoas (UFAL).

Publicar este Dossiê expressa o trabalho exitoso de dezenas de pesquisadores(as) de Filosofia do Brasil e de outros países em nosso jovem periódico. Expressa também a constância na investigação de tópicos específicos e importantes da linha de pesquisa Linguagem e Cognição do Programa, através do Grupo de Pesquisa Linguagem e Cognição. Este número especial mantém nosso trabalho de integrar estudiosos(as) de Filosofia publicando temáticas ligadas à filosofia da mente, às ciências cognitivas, à filosofia da linguagem, à lógica, à epistemologia etc.

As várias iniciativas do Grupo caminham na direção do significado da epígrafe de Pascal, trabalhar para ser digno do "pensamento com acerto", especialmente em áreas com tanta heterogeneidade teórica como a filosofia da mente e as ciências cognitivas. Assim, promover encontros, manter grupos, ensinar e, sobretudo, divul gar o pensamento filosófico pela publicação acessível de temáticas ligadas à mente, à linguagem, à lógica e ao conhecimento são modos de ação de cumprir o "dever" de "pensar com acerto".

Desde 2012, promovemos grupos de estudo, realizamos 10 encontros de pesquisadores(as), publicamos 6 livros de coletâneas, organizamos um número temático para *Prometheus – Journal of Philosophy* (UFS) e, agora, apresentamos estas duas partes da *Lampião – Revista de Filosofia*.

Ilustrado pelo artista cearense André Dias, este número especial conta com mais de 30 textos – entre artigos, resenhas e traduções. Neles, o(a) leitor(a) terá a oportunidade de tomar contato e refletir sobre uma gama bastante variada de temas, problemas, autores(as) e teorias da filosofia da mente e das ciências cognitivas. Lendo e refletindo criticamente sobre estes trabalhos, pensamos que os(as) leitores(as) podem nos ajudar a cumprir o compromisso de "pensar com acerto", pois somente a publicidade e a avaliação crítica compartilhada podem indicar o que esta expressão significa.

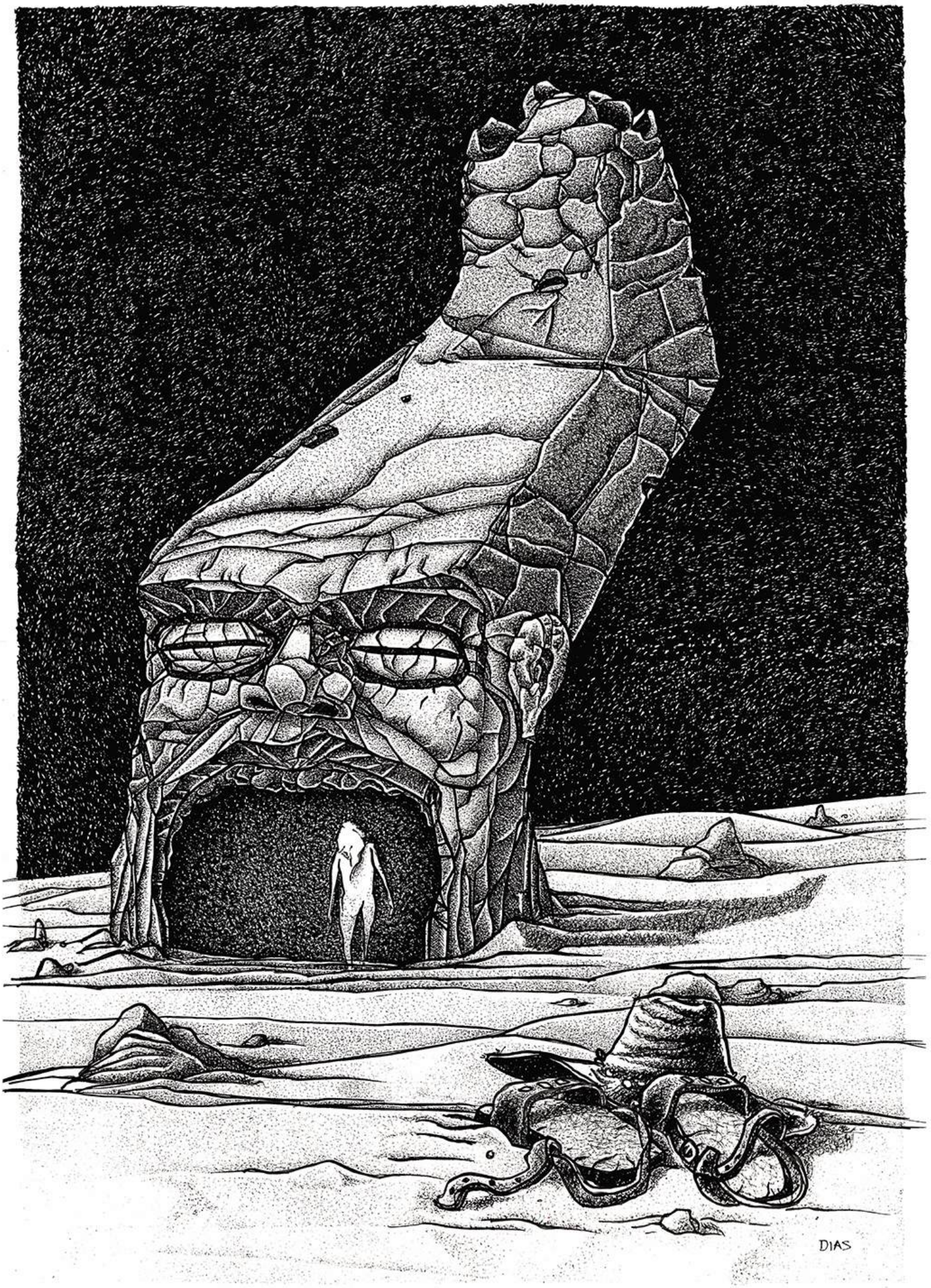
Por fim, agradecemos a todos(as) colaboradores(as), saibam que sua atividade fomenta diretamente o pensamento filosófico em Alagoas, no Nordeste, no Brasil e em muitos lugares do mundo. Em nome do Grupo de Pesquisa Linguagem e Cognição, agradecemos ao PPGFil-UFAL e aos Professores Fernando Monegalha e Rodrigo Barros Gewehr (Editores-Chefes da *Lampião*) pela confiança no trabalho desta publicação.

Dedicamos nosso número especial ao Jorginho e ao Rudá e desejamos a todos(as) uma ótima e proveitosa leitura!

Marcus José Alves de Souza
(Professor do PPGFil-UFAL)

Argus Romero Abreu de Morais
(Professor Visitante da FAALC/UFMS)

Maxwell Morais de Lima Filho
(Professor da UFCA)


DIAS

MONISMO NEUTRO E A INTEGRAÇÃO DE CONSCIÊNCIA E NATUREZA

Daniel Borgoni[1]

Resumo:

Monismo neutro é a teoria segundo a qual o nível mais fundamental da realidade é constituído por elementos cuja natureza é neutra em relação à mente e à matéria. Esse artigo tem como objetivo apresentar o monismo neutro em linhas gerais, expondo e esclarecendo suas teses principais, bem como apresentar as vantagens explicativas do monismo neutro frente ao fisicalismo e ao dualismo. Primeiramente, apresentarei seu advento e como seus precursores pretenderam conciliar o domínio da psicologia com o domínio da física, respondendo ao problema mente-corpo. Como doutrina contemporânea em filosofia da mente, argumentarei que o monismo neutro supera os principais problemas do fisicalismo e do dualismo, integrando consciência e natureza.

Palavras-chave:

Monismo neutro; dualismo; fisicalismo; consciência.

Abstract:

Neutral monism is the theory according to which the most fundamental level of reality is constituted of elements whose nature is neutral in relation to mind and matter. This article aims to present neutral monism in general terms, exposing and clarifying its main theses, as well as to present the explanatory advantages of neutral monism over physicalism and dualism. Firstly, I will present its advent and how its precursors intended to reconcile the domain of psychology with the domain of physics, responding to the mind-body problem. As a contemporary doctrine in philosophy of mind, I will argue that neutral monism overcomes the main problems of physicalism and dualism, integrating consciousness and nature.

Keywords:

Neutral monism; dualism; physicalism; consciousness.

1 Doutor em Filosofia pela Universidade Federal de São Paulo (Unifesp). Foi pesquisador visitante na Université du Québec à Montréal (UQÀM) e pesquisador de pós-doutorado na Universidade de São Paulo (USP). Atualmente, faz doutorado na Université du Québec à Trois-Rivières (UQTR). E-mail: dborgoni@hotmail.com

1. Introdução

Nossa mente é provavelmente o que temos de mais íntimo, e a ela vinculamos a razão, a intuição, a cognição, a imaginação, entre outras faculdades, bem como diferentes estados mentais, como as emoções, as sensações, as crenças, os desejos etc. Entre os aspectos do mental, um dos mais impressionantes é a consciência, entendida como experiência consciente. Nesse sentido, seres humanos e outros animais não-humanos, como penso, têm consciência, ou seja, têm estados mentais que apresentam uma fenomenologia, comumente caracterizada pelo termo *qualia*, ou qualidades fenomênicas da experiência, as quais me referirei em termos neutros, isto é, sem especificar sua natureza. Os *qualia* caracterizariam as experiências a que pertencem as nossas dores, os sentimentos de angústia, as sensações de cores e assim por diante. Por isso, os *qualia* são associados ao que a literatura denomina "consciência fenomênica", e é esse o sentido que o termo "consciência" assume neste artigo.

Embora "ter consciência" seja uma das coisas das quais mais temos certeza, nós não sabemos o que a consciência é. Esse problema tem sido abordado por diferentes linhas de trabalho, dentre as quais destaco as neurociências e a filosofia da mente. No entanto, nenhuma delas conseguiu desvelar a natureza da consciência. É uma substância, uma função, trata-se somente de informação integrada, ou não é nada além de ilusão? A consciência é um fenômeno físico ou algo não-redutível ao domínio material?[2] Como a consciência encaixa-se em nosso universo é uma questão de longa data.

Na perspectiva filosófica, as explicações sobre a natureza do mental, portanto, sobre a natureza da consciência, são geralmente oferecidas dentro de duas concepções de mundo: dualismo e monismo. O dualismo é a tese segundo a qual mente e matéria são fundamentalmente diferentes tipos de coisa, na medida em que essa distinção ocorre em nível ontológico. Ao corpo é atribuída uma natureza física e à mente, total ou parcialmente, é atribuída uma natureza não-física.

As visões de mundo dualistas dividem-se principalmente em dualismo substancial e dualismo de propriedades. O dualismo de substâncias, cujo precursor foi Descartes (1641), é a tese na qual mente e matéria são substâncias distintas, independentes e que

2 Embora possam existir diferenças entre o que é material e o que é físico, dependendo da acepção que se dê a esses termos, no debate contemporâneo, no qual este artigo se insere, é usual atribuir o mesmo sentido a ambos os termos, entendendo-se por "material" tudo aquilo que é constituído por entidades físicas e regido pelas leis da física. Consequentemente, fisicalismo, cuja definição veremos no decorrer do texto, e materialismo são termos intercambiáveis aqui.

podem existir separadamente. O dualismo de propriedades propõe que uma mesma realidade ou substância compreende propriedades fundamentalmente distintas. De um lado, corpos e demais objetos materiais são constituídos por propriedades físicas. De outro, a constituição do mental, especialmente a consciência, envolveria propriedades mentais irredutíveis a propriedades físicas.

Por sua vez, o monismo é a tese metafísica de que tudo o que existe está circunscrito a uma única realidade, ou seja, tudo é constituído por um substrato fundamental. Portanto, embora pareçam ser distintos, mente e corpo teriam a mesma natureza. Nesse viés, a história da filosofia nos mostra que teorias monistas são geralmente idealistas ou materialistas. O monismo idealista defende que a estrutura metafísica última do mundo é fundamentalmente mental. Idealistas "negam que a realidade (...) possa ser adequadamente caracterizada em termos puramente independentes da mente, isto é, em maneiras que não reconhecem uma contribuição fundamental e constitutiva do mental [à realidade]" (HUTTO, 2009, p. 357)[3]. Não existiria, portanto, uma realidade logicamente independente da consciência.

Segundo o monismo materialista, tudo o que existe é fundamentalmente constituído por algo material, de modo que, contemporaneamente, sua principal variante é o fisicalismo, segundo o qual tudo o que existe é constituído por entidades físicas e pode ser descrito por teorias científicas. Dentro da visão de mundo fisicalista, encontramos teorias filosóficas que têm como objetivo último a naturalização da mente, entendendo-se por "naturalização" estar circunscrito às categorias ontológicas das ciências da natureza. O behaviorismo, a teoria da identidade e o funcionalismo são algumas destas teorias que, *grosso modo*, defendem que podemos descrever o mental em estados ou disposições comportamentais, estados neurais e estados funcionais, respectivamente.

Nesse complexo e intricado debate entre dualistas e monistas, houve recentemente um crescente interesse pelo monismo neutro, uma concepção metafísica de mundo e da natureza do mental usualmente associada a Bertrand Russell (1921, 1927), e essa associação se justifica se consideramos que foi ele quem deu ao monismo neutro sua forma mais sistemática e abrangente[4]. De forma geral, podemos entender o monismo neutro como a teoria na qual não existe uma diferença ontológica fundamental entre os elementos que constituem o mental e os elementos que constituem o físico, na medida em que mente e

3 Todas as traduções são minhas.

4 Cf. Tully (2003, p. 335).

matéria seriam compostas por elementos neutros. Em outras palavras, nem propriedades mentais nem propriedades físicas constituiriam o nível de realidade mais básico, pois este seria constituído por entidades cuja natureza é neutra em relação à mente e ao físico. Isso daria ao monismo neutro vantagens explicativas em relação aos dualismos e monismos tradicionais, tal como endereçar uma solução para o problema mente-corpo, como veremos.

Este artigo tem como objetivo apresentar o monismo neutro em linhas gerais, expondo e esclarecendo suas teses principais, bem como argumentar que o monismo neutro tem vantagens explicativas quando comparado às principais teorias filosóficas que atualmente disputam com ele a concepção de mundo e de consciência: o fisicalismo e o dualismo. Para isso, dividirei o artigo do seguinte modo. Na próxima seção, abordaremos o advento do monismo neutro e como seus precursores pretenderam conciliar o domínio da psicologia com o domínio da física, respondendo ao problema mente-corpo. Então, tratarei do monismo neutro contemporâneo, abordando as teses que o sustentam e mostrando como os principais problemas que o fisicalismo e o dualismo têm são por ele superados. Por fim, tratarei de alguns problemas com os quais monistas neutros têm de lidar. Na última seção, farei minhas conclusões.

2. Monismo neutro

Monismo neutro é um tipo de monismo ontológico, segundo o qual existe algo neutro e mais primitivo que mente e matéria, de modo que "as coisas comumente consideradas como mentais e as coisas comumente consideradas como físicas não diferem em relação a qualquer propriedade intrínseca possuída por um conjunto e não pelo outro, mas diferem apenas em relação ao arranjo e contexto" (RUSSELL, 1913, p. 15). Mentes e corpos teriam suas próprias propriedades fundamentais, mas nenhuma delas seria o constituinte último da realidade, visto que propriedades físicas e propriedades mentais seriam diferentes ordenações das mesmas propriedades primordiais – os elementos neutros.

Monistas neutros, então, devem responder às seguintes questões: 1) qual a natureza dos elementos neutros? 2) Considerando que o domínio físico seria composto por entidades mais primitivas, como estas entidades se relacionam com a matéria? 3) Se o domínio mental é constituído por elementos neutros, qual o tipo de relação entre estes e a mente?

2.1 Precursores

O monismo neutro foi uma nova concepção de mundo cujo advento se deu por volta do início do século XX e teve como precursores Ernest Mach (1890), William James (1904) e Bertrand Russell (1921, 1927). Diferentemente dos monismos idealista e materialista, que vinham disputando as concepções de mundo e de mente, eles inovam ao propor e argumentar a favor de um nível de realidade mais profundo e constituído por entidades básicas que não seriam nem materiais nem mentais. Os pioneiros do monismo neutro estavam preocupados com a conciliação da física com a psicologia, uma vez que as ciências da natureza não conseguiam acomodar os fenômenos mentais em suas categorias explicativas. Negando a primazia ontológica da mente ou do corpo, eles propuseram a tese de que deveriam existir entidades neutras que vinculariam os fenômenos mentais aos fenômenos físicos. Nesse sentido conciliatório, declara Russell (1921, p. 244), no monismo "mente e matéria são vistas como sendo construídas a partir de um substrato neutro, cujas leis causais não têm a dualidade como as da psicologia, mas formam uma base sobre as quais a física e a psicologia são construídas". Desse modo, os conceitos da psicologia e os da física teriam uma base ontológica em comum, convergindo com o objetivo daqueles, qual seja, de que o monismo neutro fosse uma teoria metafisica sobre a realidade integrada com as ciências.

Embora os monismos neutros de Mach (1890), James (1904) e Russell (1921, 1927) tenham divergências, Banks afirma que suas posições compartilham as seguintes teses:

> 1. Monismo: os domínios mental e físico são parte de um domínio natural mais abrangente de elementos e suas variações funcionais.
> 2. Neutralismo: elementos não são nem mentais nem físicos; mentes e corpos físicos são complexos de elementos funcionalmente relacionados. Certas variações funcionais de elementos são chamadas mentais e outras são chamadas físicas, mas não há qualquer dualidade de variações a eles subjacentes.
> 3. Identidade psicofísica: cada sensação, tal como "verde", é também um elemento físico, uma energia neural no cérebro. Nem todo elemento é uma sensação, ou mesmo uma possível sensação.
> 4. Poderes: elementos são poderes com força causal. Eles são qualidades concretas e modos disposicionais de afetar as coisas em seus vários papéis causais ou funcionais. A qualidade concreta instancia o papel disposicional, relacional. Todo elemento é naturalmente embebido em seu papel funcional (BANKS, 2010, p. 175).

A primeira tese é a adesão ao monismo ontológico, segundo o qual a natureza última da realidade são elementos dos quais mentes e objetos físicos são derivados. Ain-

da que mentes e corpos sejam constituídos pelos mesmos elementos, a sua diferenciação ocorreria por meio da ordem e das suas variações funcionais. A segunda tese especifica a natureza dos elementos como neutros em relação ao mental e ao material, diferenciando o monismo neutro dos monismos tradicionais. Se, por um lado, o monismo neutro se aproxima do dualismo ao manter que propriedades mentais e propriedades físicas são diferentes, por outro, ele se diferencia do dualismo ao declarar que não existe uma distinção ontológica fundamental entre mente e objetos físicos, uma vez que ambos seriam ordenações funcionais dos mesmos elementos primordiais. A primeira e a segunda teses, em conjunto, mostram que o monismo neutro foi uma nova posição em metafísica e em filosofia da mente.

A terceira tese fala em identidade psicofísica entre o que a pessoa sente e o que ocorre no cérebro na sensação de verde, o que pode dar a entender que o monismo neutro defende um tipo de teoria da identidade mente-matéria, mas essa interpretação seria um equívoco. A sensação de verde pode ser interpretada como sendo uma qualidade do estado mental quando uma pessoa vê a cor verde de um abacate (*greenness* ou o *quale* do verde) ou como um evento físico quando um observador externo analisa a rede neural associada à sensação de verde. Contudo, a sensação de verde em si pode ser considerada em ambos os modos, na medida em que sua natureza seria neutra em relação à mente e ao corpo. Assim, na passagem de Banks (2010, p. 175), a expressão "identidade psicofísica" diverge do sentido daquela empregada na teoria da identidade.

Por fim, a quarta tese declara que os elementos neutros têm poderes causais e se unem em relações funcionais uns com os outros. Ela também trata os elementos neutros como qualidades concretas que, por serem ativas, se diferenciariam dos *qualia* e das qualidades secundárias. Em certa medida, esta última tese associa o monismo neutro ao estruturalismo sobre a física, uma vez que podemos interpretar que elementos neutros instanciam as características disposicionais da matéria. E, de fato, nós encontramos a adesão ao estruturalismo na seguinte passagem de Russell (1927, p. 10): "a física, em si mesma, é extremamente abstrata e revela somente certas características matemáticas da matéria com a qual ela lida. Ela não nos informa nada sobre a característica intrínseca da matéria". Por enquanto, não me estenderei mais sobre o estruturalismo, pois tratarei dele posteriormente.

Para tipificar as entidades primordiais propostas pelos pioneiros do monismo neutro, vimos que Banks (2010, p. 175) utiliza as sensações, provavelmente porque eles, ainda que pudessem divergir sobre o que tais entidades seriam, caracterizaram sua natureza

com base na experiência. Nesse sentido, Mach (1890) denominou "elementos/sensações" aos primitivos que constituiriam tudo o que existe. Para James (1904), seria a "experiência pura" as entidades neutras que, enquanto tais, podiam vir a ser objeto ou sujeito. Por sua vez, Russell (1921) propõe que as "sensações" seriam as entidades neutras que constituiriam os domínio mental e material:

> As coisas do mundo, até onde as experimentamos, consistem, na concepção que eu estou advogando, de inúmeros particulares transitórios, tal como ocorrem na visão, audição etc. (...) Sensações são o que é comum aos mundos mental e físico; elas podem ser definidas como a intersecção de mente e matéria (RUSSELL, 1921, p.118-9).

As sensações seriam particulares intrinsecamente nem mentais nem físicos, mas uma categoria metafisicamente neutra, de modo que mente e corpo seriam constituídos conforme as relações que tais particulares estabeleceriam uns com os outros. A eleição das sensações como elementos neutros é justificada por Russell (1921, p. 15): "Me parece, *prima facie*, que existem diferentes tipos de leis causais, umas pertencendo à física e outras à psicologia (...) Sensações são sujeitas a ambos os tipos de leis e, por isso, são verdadeiramente neutras". Ou seja, as sensações seriam a intersecção entre o mental e o material, na medida em que responderiam às leis físicas e às leis psicológicas. Desse modo, seriam as sensações o que vincularia o ponto de vista de primeira pessoa com o ponto de vista de terceira pessoa. Isso explicaria porque as sensações parecem ser materiais quando consideradas estritamente sob o ponto de vista objetivo e parecem ter uma natureza mental quando consideradas do ponto de vista de quem sente.

A escolha das sensações como as entidades neutras respondia ao anseio de Russell (1921) de não perpetuar a dicotomia tradicional entre mente e matéria, ao mesmo tempo em que as sensações, como fenômenos observáveis, seriam passíveis de afirmações empíricas, permitindo construir uma metafisica sem uma orientação exclusivamente subjetiva. A filosofia, assim, uniria os discursos da física e da psicologia por meio de entidades metafisicas consistentes com o conhecimento científico[5].

5 Note que Russell (1921) foi bastante perspicaz ao propor que as sensações seriam os elementos neutros entre os domínios mental e físico. As sensações são, entre os estados mentais que têm um caráter fenomênico, os paradigmas de estados mentais com *qualia*, visto que suas qualidades – o que é sentido – parecem constituir (se não toda) parte da sua essência. Ao mesmo tempo, as sensações estão intimamente associadas ao corpo, tal como a sensação de dor, cuja ocorrência está associada ao estímulo das fibras-C do sistema nervoso.

Em obra posterior, que consolida sua visão madura sobre o monismo neutro, Russell (1927, p. 107) começa a se referir aos elementos neutros como "perceptos", mas as sensações continuam com seu lugar privilegiado, pois "as sensações são o núcleo teórico na experiência real". Russell (1927, p. 402) declara que somente por meio dos perceptos teríamos acesso a propriedades intrínsecas, tendo em vista que "os perceptos são a única parte do mundo físico que nós conhecemos não abstratamente". Desse modo, diferentemente do conhecimento que adquirimos por meio das ciências naturais, para Russell (1921, 1927), as sensações/perceptos nos dão um conhecimento privilegiado em relação à estrutura metafisica última da realidade, sendo uma das razões para que ele elegesse as sensações/perceptos como os elementos neutros que constituiriam as ocorrências que denominamos "mentais" e as ocorrências que chamamos "físicas".

A adoção de um monismo cuja natureza das entidades básicas é neutra em relação aos domínios físico e mental responderia ao problema mente-corpo, como explica Hatfield:

> Conforme desenvolvido por James e Russell, o monismo neutro evitava o problema mente-corpo postulando apenas uma "entidade", a alegada "entidade" neutra de particulares momentâneos, experiências puras, ou elementos machianos. Mach, James e Russell poderiam então apontar para dois conjuntos de leis a serem encontrados empiricamente nos estados sucessivos desse substrato: leis psicológicas governando sucessões de percepções e outros estados mentais considerados como mentais, e leis físicas governando sucessões de percepções e *sensibilia*, ou experiências puras não vividas, consideradas como físicas. A relação mente-corpo tornou-se então uma questão de traçar conexões entre sequências físicas e sequências psicológicas entrecruzadas de particulares momentâneos (HATFIELD, 2002, p. 222).

Embora os domínios mental e material tenham suas próprias propriedades, mente e matéria não seriam intrinsecamente incompatíveis, pois seriam maneiras diferentes de como as entidades neutras se manifestam. Assim, tal como os monismos idealista e materialista, o monismo neutro evitava o problema enfrentado por dualistas de substância, qual seja, o de ter de explicar como interagiam duas substâncias cujas naturezas eram incompatíveis. Ao mesmo tempo, evitava o idealismo, considerado uma doutrina errônea pelos pioneiros do monismo neutro, e evitava o materialismo, considerado uma doutrina que tinha bons *insights*, mas incapaz de oferecer uma explicação suficiente para os fenômenos mentais[6].

6 A título de exemplo, no primeiro capítulo de *The Analysis of Mind*, Russell (1921) declara que o beha-

2.2 Monismo neutro contemporâneo

Após o advento do monismo neutro e algumas versões terem sido propostas ao longo do século XX, a discussão sobre a doutrina dos elementos neutros ganhou força recentemente. Eric Banks (2014), David Chalmers (2010) e Thomas Nagel (2012) são alguns filósofos que argumentam em prol da existência de entidades num nível de realidade mais fundamental do que os níveis mental e físico. Embora suas concepções do monismo neutro empreguem conceitos contemporâneos, elas compartilham diversas teses com os monismos neutros clássicos. Nesse sentido, Alter & Nagasawa (2015) afirmam que o desenvolvimento da concepção monista de Russell (1927) deu origem a um conjunto de teorias que se enquadram no denominado "Monismo Russelliano", pois compartilham as seguintes teses:

> *Estruturalismo sobre a física*: a física descreve suas propriedades básicas somente em termos estruturais/disposicionais.
> *Realismo sobre as propriedades intrínsecas relevantes*: existem propriedades intrínsecas que constituem a consciência e servem como bases não-estruturais/categoriais para as propriedades estruturais/disposicionais descritas pela física.
> *Fundacionismo (proto)fenomênico*: ao menos algumas destas propriedades intrínsecas são propriedades fenomênicas ou propriedades protofenomênicas – propriedades não-fenomênicas que juntas (talvez em combinação com propriedades estruturais/disposicionais) constituem a consciência (ALTER & NAGASAWA, 2015, p.3).

Antes de abordarmos as três teses acima, é preciso esclarecer que os monismos que se enquadram no monismo russelliano não precisam sustentar que os elementos fundacionais da realidade são neutros em relação à mente e ao corpo. Assim, como este artigo trata do monismo neutro, acrescento às teses acima uma quarta tese: a *neutralidade das propriedades intrínsecas* em relação às propriedades físicas e às propriedades mentais. Por consequência, podemos considerar o monismo neutro como um subgrupo do monismo russelliano. Posto isso, esclareçamos as outras teses.

De acordo com a primeira tese, a física lida somente com as características relacionais ou disposicionais da matéria, isto é, a física caracteriza os objetos físicos pelas relações que uns estabelecem com os outros. Não revelaria, portanto, a natureza intrínseca das entidades físicas básicas, ou os *relata* destas relações. Desse modo, a primeira afirma-

---

viorismo tinha importantes elementos de verdade, mas o behaviorismo, como uma variante do monismo materialista, era incapaz de oferecer uma explicação satisfatória para a natureza da mente.

ção é a adesão do monismo neutro ao estruturalismo, ou realismo estrutural, segundo o qual o conhecimento que adquirimos por meio de nossas teorias científicas é estrutural, visto que a ciência teria acesso somente à estrutura do mundo. Por "estrutura" podemos entender o conjunto de relações que envolvem um objeto, como as relações causais, por exemplo. A ciência, assim, teria acesso somente às propriedades relacionais das coisas. Por exemplo, a física caracteriza os "quarks" por suas interações com outras entidades físicas, e a "massa" é caracterizada por meio dos seus papeis relacionais, tais como resistir à aceleração e atrair outras massas (quando estas atingem dimensões consideráveis). Mas se nós temos disposições, é razoável perguntar o que as instancia?

A segunda tese responde à questão acima propondo a existência de propriedades intrínsecas que constituiriam um nível de realidade mais profundo do que o domínio mental e o domínio físico. Propriedades intrínsecas instanciariam as propriedades disposicionais exibidas pelas entidades físicas e constituiriam a consciência. Monistas neutros, então, ao considerar que existe uma realidade intrínseca às estruturas descritas pela física (realidade extrínseca), endossam o estruturalismo epistemológico. Por consequência, nega-se o estruturalismo ontológico, que sustenta uma ontologia de estruturas na qual não haveria uma realidade intrínseca às mesmas, isto é, não haveria o *relata* das relações.

Mediante o exposto, a natureza dos *qualia* vivenciados em nossas experiências e a natureza dos objetos físicos seriam fundadas nas mesmas propriedades primordiais. Os *qualia* seriam irredutíveis ao físico justamente porque mente e matéria envolveriam ordenações e sequências diferentes de elementos neutros. Essa irredutibilidade dos *qualia* daria a impressão errônea que existe uma diferença ontológica fundamental entre o mental e o material, como defendem os dualistas. Ao mesmo tempo, explicaria porque as tentativas de reduzir a consciência às categorias ontológicas fisicalistas não têm tido sucesso.

A terceira tese caracteriza a natureza das propriedades que constituiriam a estrutura metafisica última da realidade como propriedades fenomênicas ou propriedades protofenomênicas[7]. As primeiras são propriedades com as quais entraríamos em contato em nossas experiências e que dariam aos estados mentais conscientes seu caráter fenomênico. Ao mesmo tempo, as propriedades fenomênicas seriam as propriedades intrínsecas das entidades descritas pela física. Daí seu caráter neutro. Por sua vez, propriedades protofenomênicas constituiriam as propriedades fenomênicas e as propriedades físicas e "seriam

7 Note que essa não é uma caracterização exaustiva do que seriam as propriedades intrínsecas. Por exemplo, podemos caracterizá-las também como propriedades físicas de um tipo especial (cf. ALTER & NAGASAWA, 2015, p. 434).

relacionadas com a experiência da mesma maneira que propriedades físicas são relacionadas com propriedades menos básicas, como a temperatura" (CHALMERS, 1996, p. 126-7). Consideradas como elementos neutros, as propriedades protofenomênicas constituiriam o domínio físico por meio de relações e constituiriam o domínio fenomênico por sua natureza intrínseca coletiva[8].

Tomadas em conjunto, as teses acima mostram que o monismo neutro retém a concepção física da realidade, suplementando-a com uma base categorial, uma vez que a física descreveria suas entidades básicas em termos puramente estruturais. Neste sentido, os "quarks" seriam partículas subatômicas que, associadas com a força nuclear forte, constituiriam os prótons e nêutrons, mas sua natureza intrínseca seria descrita como ordenações de propriedades (proto)fenomênicas. Em relação à "massa", ela pode ser fisicamente caracterizada pela sua resistência à aceleração, mas seriam elementos neutros que descreveriam sua natureza última.

Como posição metafisica sobre a realidade e sobre a consciência, o monismo neutro tem vantagens explicativas se o compararmos ao dualismo e ao fisicalismo. Retomando o que foi dito, dualistas afirmam que a consciência é ontologicamente irredutível ao cérebro, seja porque envolvem duas substâncias distintas, seja porque envolvem propriedades fundamentalmente distintas. Fisicalistas, por seu turno, defendem que a consciência é um fenômeno natural, isto é, que pode ser descrita pela ontologia da biologia, química, física ou das ciências delas derivadas. Entretanto, fisicalistas e dualistas enfrentam problemas para sustentar suas respectivas concepções de consciência, que são superadas pelo monismo neutro. Se não, vejamos.

Começando pelo dualismo, uma crítica usualmente feita à sua versão interacionista é que, ao defender o papel causal do mental sobre o cerebral, isto é, a causação mental, tem-se de contestar o fechamento causal do mundo físico:

> Selecione um evento físico (...) e trace suas causas anteriores ou posteriores tanto quanto queira; o princípio do fechamento causal diz que isso nunca levará você para fora do domínio físico. Portanto, nenhuma cadeia causal envolvendo um evento físico cruza os limites do físico para o não-físico: se x é um evento físico e y é a causa ou o efeito de x, então y também deve ser um evento físico (KIM, 2011, p. 214).

8 Cf. Chalmers (2010, p. 134).

Se o mundo físico é causalmente fechado, para cada estado ou efeito físico existe uma causa física suficiente, não havendo, portanto, lugar para causas não-físicas, de modo que o funcionamento de sistemas físicos, como um cérebro, seria explicado fisicamente. Ao examinarmos o cérebro do lado de fora, ou seja, a partir de uma perspectiva de terceira pessoa, "podemos, em princípio, traçar os efeitos dos estímulos de entrada sobre o sistema nervoso central por todo o caminho, do *input* ao *output*, sem encontrar qualquer hiato na cadeia de causação que pudesse ser preenchido pela consciência" (VELMANS, 2002, p. 5). Em resposta, dualistas interacionistas têm questionado o sentido de "mundo físico causalmente fechado", uma vez que isso parece fazer sentido somente em um universo determinista, mas a questão de se o universo é ou não determinista está em aberto. A causalidade num universo determinista parece tornar o dualismo interacionista inviável, o que não ocorreria num universo indeterminista[9].

Uma maneira que dualistas têm de se esquivar do problema do fechamento causal do mundo físico é adotar o epifenomenalismo, tese segundo a qual estados mentais podem ser causados por estados neurais, mas estados mentais não têm qualquer papel causal no cérebro. Em outras palavras, epifenomenalistas aceitam a causação corpo-mente, mas não a causação mente-corpo. Assim, a causação mental não seria um problema porque não seria endossada. O epifenomenalismo, porém, é contraintuitivo. Não parece correto afirmar que nossas sensações, emoções e outros tipos de estados mentais não afetam causalmente nosso corpo. Além de ser contraintuitivo, existem evidências da causação mental e algumas das melhores, segundo Velmans (2002, p. 2), são as seguintes: "a pressão sanguínea, a atividade vasomotora, os níveis de glicose sanguíneo, a dilatação da pupila, a atividade eletrodérmica e o funcionamento do sistema imunológico podem ser influenciados por estados conscientes". Em contraponto, o fato de uma teoria contradizer nossa intuição parece não ser um problema sério, e o exemplo mais notório de que a nossa intuição pode falhar parecem ser os fenômenos quânticos. Até que ponto o epifenomenalismo é sustentável frente às evidências da causação mental é uma discussão em aberto.

Uma solução para os problemas enfrentados por dualistas é defender um ponto de vista fisicalista sobre a consciência. Ora, se a natureza do mental puder ser descrita fisicamente, podemos aceitar o argumento do fechamento causal do mundo físico e não precisamos adotar o epifenomenalismo. Mas o fisicalismo também não tem oferecido uma explicação satisfatória para o fenômeno da consciência e, portanto, para a natureza

9 A defesa da tese de que o mundo é indeterminista tem se apoiado na mecânica quântica.

do mental. A principal dificuldade enfrentada pelas linhas de trabalho que visam reduzir a consciência ao físico parece ser a de explicar a qualidade subjetiva que aparece à consciência de uma pessoa quando ela tem uma experiência, os *qualia*. Nas palavras de Crick & Koch (2003, p. 119), "ninguém produziu qualquer explicação plausível de como a experiência da vermelhidão do vermelho [o *quale* do vermelho] poderia surgir das atividades do cérebro". Vejamos esse problema com mais vagar.

Essa dificuldade foi claramente exposta por Nagel (1974, p. 436) ao argumentar que o caráter subjetivo da experiência consciente não estaria incluso em nenhuma teoria que visa circunscrever o mental ao físico, pois "todas elas são logicamente compatíveis com a sua ausência". Ou seja, nenhum fato físico implica necessariamente a existência de fatos sobre a consciência. Por exemplo, nós sabemos que experiências conscientes estão associadas ao corpo, como, por exemplo, a associação entre a sensação de dor e o estímulo das fibras-C do sistema nervoso. Contudo, não conseguimos deduzir a sensação de dor somente com base nos mecanismos neurais associados à dor. Em outras palavras, não conseguimos explicar porque alguém está sentindo dor, e não cócegas, baseados somente nas estruturas, funções e atividades cerebrais.

Nesse sentido, Jackson (1982) propôs o argumento do conhecimento, que parte da insuficiência do conhecimento físico em descrever o conhecimento fenomênico, ou seja, conhecimento de como as qualidades sentidas dos estados mentais conscientes aparecem para nós, para concluir que os *qualia* são não físicos. Por consequência, existiria uma diferença ontológica fundamental entre consciência e matéria[10].

O obstáculo a ser ultrapassado por fisicalistas é o de descrever a experiência consciente sob um ponto de vista de terceira pessoa, ou perspectiva objetiva. "Como é sentir determinada dor" parece acessível somente pela pessoa que está sentindo dor, isto é, o dolorido da dor (o *quale* da dor) seria apreendido sob um ponto de vista em primeira pessoa, ou perspectiva subjetiva. Desse modo, ainda que fatos físicos (como o estímulo das fibras-C) e a sensação de dor estejam associados, é sensato dizer que a descrição física da dor não explica o que é a dor, pois deixa de fora sua fenomenologia. Embora o "sentir dor" dependa das fibras-C, as propriedades físicas não seriam suficientes para explicar a natureza da sensação de dor. Em outros termos, se a consciência tem características qualitativas que só podem ser apreendidas subjetivamente, como conciliar a objetividade das

10 Ver Borgoni (2016) para uma análise do argumento do conhecimento frente às diversas objeções que lhe foram opostas.

explicações científicas com a subjetividade da consciência, se "essa objetividade resulta, justamente, da tentativa de eliminar tudo que seja relativo a um determinado ponto de vista?" (ABRANTES, 2005, p. 225). Como um sistema físico como o cérebro pode instanciar a experiência de dor de alguém ou a experiência de vermelho que tenho quando vejo um tomate maduro é um mistério.

Dado o exposto, é racional supor que propriedades não-físicas são responsáveis pelas características qualitativas exibidas nas experiências conscientes. Contudo, se adotarmos uma explicação dualista para a natureza da consciência, voltamos a enfrentar os problemas que incidem sobre o dualismo, conforme expostos. Em contrapartida, o monismo neutro é capaz de acomodar a tese de que os *qualia* são não-físicos e ao mesmo tempo integrar consciência e natureza ao propor a existência de entidades neutras e subjacentes aos domínios mental e físico.

Diferentemente do dualismo interacionista, que parece ser incompatível com a tese do fechamento causal do mundo físico, o monismo neutro pode acomodar a causação mental. A consciência teria um papel causal sobre o físico, mas, diferentemente da causação na qual o mental interfere diretamente no cerebral, o monismo neutro mantém que a causação mental passaria por elementos neutros, preservando a tese de que o mundo físico é causalmente fechado. Em outras palavras, as relações causais entre propriedades mentais e propriedades físicas ocorreriam entre elementos neutros, não afetando a natureza das relações causais entre entidades físicas, pois nestas os elos causais seriam todos físicos.

O monismo neutro também supera a dificuldade do fisicalismo em integrar consciência e natureza. Ao tentar eliminar as assimetrias entre consciência e matéria, as explicações fisicalistas parecem deixar de fora o fenômeno da experiência consciente, justamente porque não capturam a perspectiva subjetiva. Ao contrário, o monismo neutro não precisa desfazer as assimetrias entre mente e corpo, conciliando a perspectiva de primeira pessoa com a perspectiva de terceira pessoa, ao mesmo tempo que acomoda os *qualia*, descritos como resultado de certa ordem e combinação de elementos neutros. Desse modo, nós podemos acomodar a intuição de que a consciência não é redutível ao físico sem aderir ao dualismo e manter a parcimônia ontológica, cara aos pioneiros do monismo neutro.

2.3 Problemas

Como vimos, monistas neutros caracterizam a realidade última por meio de entidades que são neutras em relação à mente e aos objetos físicos. O nível de realidade mais profundo seria constituído por propriedades intrínsecas que instanciariam as propriedades disposicionais exibidas pelas entidades físicas e constituiriam o domínio experiencial. Contudo, o monismo neutro enfrenta problemas, e alguns deles incidem sobre o modo como a natureza dos elementos neutros é caracterizada, isto é, como propriedades fenomênicas ou como propriedades protofenomênicas.

Entre filósofas e filósofos que partilham a tese de que as entidades fundacionais do mundo são propriedades fenomênicas, parece razoável afirmar que Russell (1921, 1927) estaria nesse grupo, uma vez que ele atribui esse papel às sensações/perceptos. Retomemos a afirmação de Russell (1927, p. 402) de que "os perceptos são a única parte do mundo físico que nós conhecemos não abstratamente". Como as sensações, os perceptos têm um caráter fenomênico e parecem nos colocar em contato com propriedades fenomênicas que, no contexto do monismo neutro, seriam as propriedades intrínsecas que constituiriam o mental e o físico.

Para esclarecer melhor esse ponto, vejamos um exemplo. Quando nós temos uma experiência consciente, como uma sensação de dor, nós sabemos como é sentir dor. Essa experiência aparece à consciência de uma maneira especifica, isto é, com um caráter qualitativo distintivo que é apreendido subjetivamente por quem experiencia a dor. Assim, "sentir dor" é ter experiência direta e imediata da qualidade sentida. Isso sugere, como argumenta quem afirma que os elementos neutros são propriedades fenomênicas, que nós conhecemos as propriedades fenomênicas pela sua natureza intrínseca. Desse modo, assim como Russell (1921, 1927) declara que as sensações e os perceptos nos revelam sua natureza intrínseca, o mesmo ocorreria com as propriedades fenomênicas.

No entanto, ao afirmar que propriedades fenomênicas também constituem propriedades físicas, o monismo neutro parece nos conduzir ao pampsiquismo, a tese de que o mental, em menor ou maior grau, existe em todas as coisas, de átomos a seres humanos. Ressalto que não se trata de sustentar que tudo tem uma mente, mas de que ao menos estados mentais ou aspectos do mental permeiam todas as coisas. De qualquer forma, o pampsiquismo é uma concepção bastante controversa e pouco aceita em filosofia da mente, mas que tem suscitado interessantes debates sobre sua razoabilidade. Se o pampsiquismo é um problema para a doutrina dos elementos neutros, é uma outra questão em aberto.

Monistas neutros podem evitar a possibilidade de terem de lidar com o pampsiquismo se mantiverem que as entidades neutras são propriedades protofenomênicas, como Chalmers (2010) propõe. Como vimos, propriedades protofenomênicas seriam propriedades intrínsecas que instanciariam propriedades físicas e que "coletivamente constituiriam as propriedades fenomênicas quando organizadas de determinada maneira" (CHALMERS, 2010, p. 151). Mas se, por um lado, as propriedades protofenomênicas desvinculam o monismo neutro do pampsiquismo, por outro, nós não temos qualquer critério por meio do qual poderíamos delimitar o que é uma propriedade protofenomênica. Se existe uma racionalidade no conceito de "propriedades fenomênicas", visto que podemos sustentar sua existência nos baseando em nossas experiências conscientes, essa racionalidade parece ser enfraquecida (ou perdida) no caso do conceito de "propriedades protofenomênicas". Como propriedades protofenomênicas fundariam propriedades fenomênicas e propriedades físicas? Como nós teríamos acesso a elas? Como justificar sua existência?

Às dificuldades que monistas neutros encontram em fundar a realidade em propriedades (proto)fenomênicas, soma-se o problema da combinação:

> Experiências familiares se apresentam como regulares, contínuas e unificadas. E elas parecem pertencer a um único sujeito. Certamente, elas têm vários aspectos. Mas esses aspectos têm uma homogeneidade subjacente. Em resumo, nossa experiência parece ter um caráter específico e homogêneo (ALTER & NAGASAWA, 2015, p. 446).

Nossas experiências apresentam uma unidade que seria contraditória à tese de que experiências seriam o resultado da combinação de propriedades mais primitivas em relação às propriedades mentais e propriedades físicas. Se este for o caso, o monismo neutro falharia na descrição das nossas experiências, uma vez que experiências conscientes não poderiam ser decompostas em elementos mais básicos. Ora, se a condição *sine qua non* para o monismo neutro é a existência de um nível de realidade mais profundo que os níveis mental e material, como conciliá-lo com o caráter homogêneo das nossas experiências? Se os elementos neutros forem não-experienciais, como sustentar a tese de que a experiência é construída a partir de elementos não-experienciais?

3. Conclusão

Compreender a natureza da mente tem sido um grande desafio para os seres humanos e parte desse desafio é fornecer uma explicação racional para o fenômeno da consciência. Dualistas e monistas tradicionais têm disputado a concepção de mundo e a melhor explicação para a natureza da consciência, mas ela resiste a ser explicada por ambas as abordagens. Uma forma teórica alternativa de abordar o problema da consciência é o monismo neutro, no qual as entidades fundacionais da realidade seriam neutras no que se refere aos domínios mental e material. Mente e corpo difeririam um do outro conforme o arranjo e a ordenação destas entidades primordiais.

Embora enfrente problemas e tenha uma aura de mistério ao propor um nível mais profundo de realidade, o monismo neutro tem vantagens explicativas quando comparado às teorias contemporâneas dele rivais, o dualismo e o fisicalismo: 1) responde ao problema mente-corpo ao propor elementos cuja natureza seria neutra em relação à mente e à matéria; 2) tem espaço teórico para abrigar as qualidades fenomênicas da experiência sem defender que a natureza da consciência envolve propriedades fundamentalmente diferentes das propriedades físicas, o que evita os problemas que dualistas enfrentam; 3) o monismo neutro retém a concepção física da realidade, suplementando-a com uma base categorial; e 4) o monismo neutro acomoda a intuição de que a consciência não é redutível ao físico, mantendo a parcimônia ontológica. Por isso, o monismo neutro parece ser capaz de integrar consciência e natureza.

Referências

ABRANTES, P. Thomas Nagel e os limites de um reducionismo fisicalista. **Cadernos de História e Filosofia das Ciências**, v.15, p. 223-244, 2005.

ALTER, T. & NAGASAWA, Y. What is Russellian Monism? In: Alter, T. & Nagasawa, Y. (Eds.). **Consciousness in the physical world**. Oxford University Press, 2015.

BANKS, C. Neutral Monism Reconsidered. **Philosophical Psychology**, v. 23 (2), p. 173-187, 2010.

BORGONI, D. As qualidades fenomênicas da experiência e o argumento do conhecimento. **Principia**, v. 20 (3), p. 393-416, 2016.

CHALMERS, D. **The conscious mind.** New York: Oxford University Press, 1996.

CHALMERS, D. **The character of consciousness**. Oxford University Press, 2010.

CRICK, F. & KOCH, C. A framework for consciousness. **Nature neuroscience**, v.6 (2), p. 119–126, 2003.

DESCARTES, R. **Meditações metafísicas**. Trad. de J. Guinsburg e Bento Prado Jr. São Paulo: Abril Cultural, 1973 (Coleção Os Pensadores) [1641].

HATFIELD, G. Sense-data and the Philosophy of Mind: Russell, James and Mach. **Principia**, v. 6 (2), p. 203-30, 2002.

JACKSON, F. Epiphenomenal Qualia. **Philosophical Quartely**, v. 32, p.127-136, 1982.

JAMES, W. Does Consciousness Exist? **Journal of Philosophy, Psychology and Scientific Methods**, v. 1(18), 1904.

KIM, J. **Philosophy of mind**. Boulder. Westview Press. 2011

MACH, E. The Analysis of Sensations. **The Monist**, v. 1(1), p. 48-68, 1890.

NAGEL, T. **Mind and cosmos**. Oxford: Oxford University Press, 2012.

NAGEL, T. What is it like to be a Bat? **The Philosophical Review**, v. 82, p. 435–450, 1974.

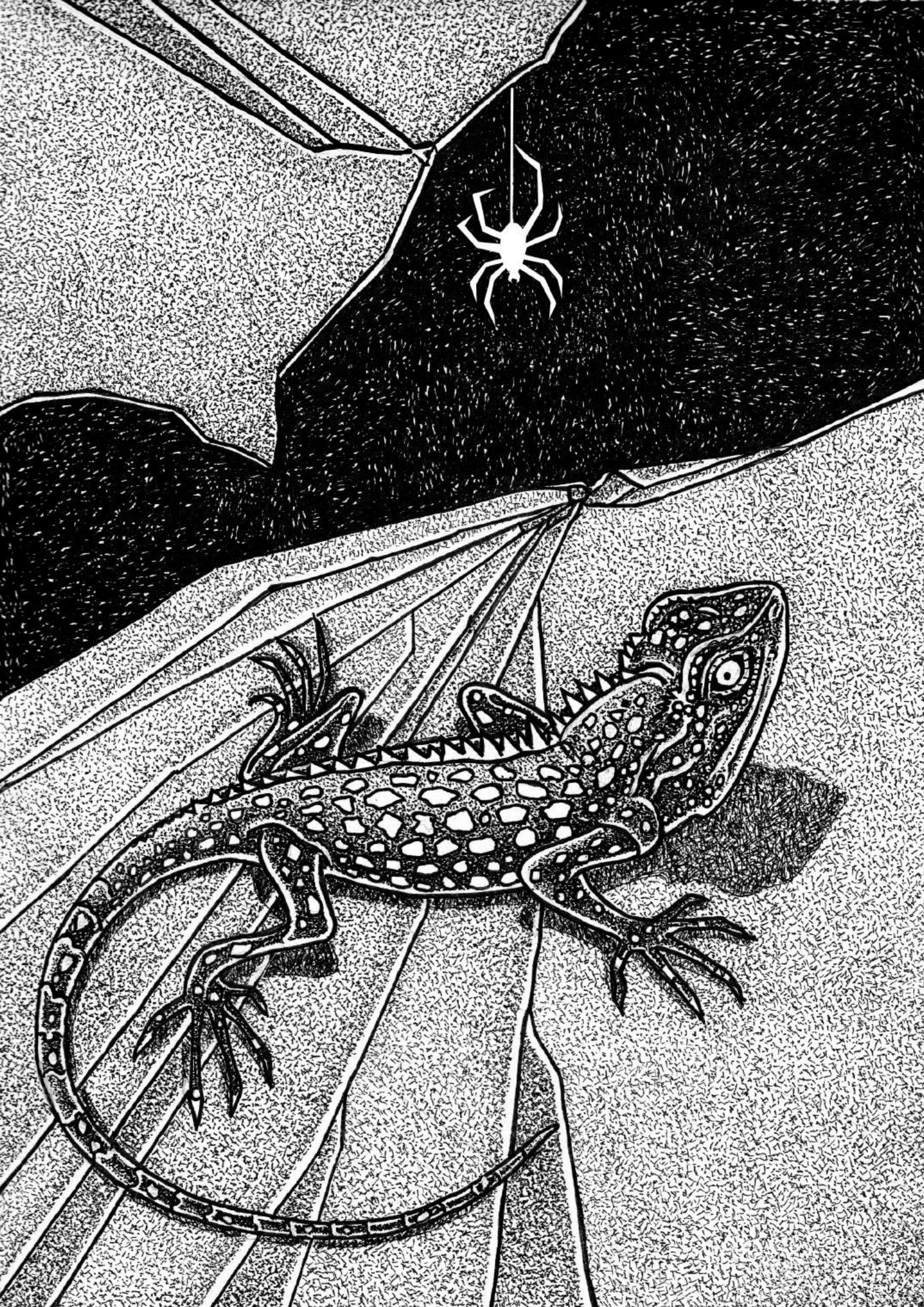
RUSSELL, B. **Theory of knowledge**. The 1913 Manuscript. London: Routledge, 1984 [1913].

RUSSELL, B. **The Analysis of matter**. London: Routledge, 1992 [1927].

RUSSELL, B. **The Analysis of mind**. New York: Taylor & Francis Group, 2005 [1921].

TULLY, R.E. Russell's Neutral Monism. In: GRIFFIN, Nicholas (Ed.). **Cambridge Companion to Bertrand Russell**. Cambridge University Press, 2003.

VELMANS, M. How could conscious experiences affect brains? **Journal of Consciousness Studies**, v. 9, p. 3-29, 2002.



A NEUROFILOSOFIA E O MATERIALISMO ELIMINATIVISTA

João de Fernandes Teixeira[1]

Resumo:

O artigo enfatiza a necessidade de distinguir entre neurofilosofia e materialismo eliminativo. Tais concepções são frequentemente confundidas, o que provoca confusões conceituais na filosofia da mente.

Palavras-chave:

Churchlands. Neurofilosofia. Materialismo eliminativo. Problema de Molyneux.

Abstract:

The paper emphasises the need of a distinction beteween neurophilosophy and eliminative materialism. Such conceptions are often conflated, leading to conceptual confusion in the philosophy of mind.

Keywords:

Churchlands. Neurophilosophy. Eliminative materialism. Molyneux problem.

Nos últimos anos, venho acompanhando, assiduamente, publicações sobre neurofilosofia e sobre materialismo eliminativo. Mas serão estas disciplinas distintas? A neurofilosofia é um novíssimo ramo da investigação filosófica. Ela surgiu na metade da década de 1980 e sua proposta é ajudar a resolver os problemas da filosofia da mente por meio da neurociência. Mas será isso o mesmo que propõe o materialismo eliminativo? Penso que a resposta é negativa.

Examinemos a década de 1980. Não havia muita aproximação entre neurociência e filosofia. Poucos filósofos sequer se referiam a problemas neurocientíficos. Thomas Nagel e Daniel Dennett foram exceções. Nagel escreveu um artigo para filósofos da mente, "Brain Bissection and the Unity of Consciousness", no qual aborda o problema da comissurotomia, ou, mais precisamente, das consequências cognitivas e filosóficas da separação cirúrgica dos hemisférios cerebrais. Dennett, por sua vez, escreveu um artigo sobre a im-

1 PhD pela Universidade de Essex, Inglaterra.

possibilidade de uma máquina sentir dor, no qual se utiliza de várias explicações neurocientíficas. E no final admite que a única abordagem não-dualista da dor é o materialismo eliminativo.

Em 1986, Patrícia Churchland publica o livro *Neurophilosophy*: Towards a unified science of mind/brain, um marco fundamental na aproximação entre neurociência e filosofia. O livro teve grande impacto, especialmente porque, até então, quem queria estudar a mente buscava dados e pesquisas na ciência da computação e na inteligência artificial. O computador era considerado o modelo da mente. Achava-se que um cérebro biológico igual ao nosso não seria necessário para produzir uma mente; ela poderia ser produzida, por exemplo, por um dispositivo de silício.

Com o advento da década do cérebro, os anos 1990, essa situação muda radicalmente. A invenção da neuroimagem e outros desenvolvimentos da neurociência impediam os filósofos da mente de ignorar a neurociência, especialmente no que diz respeito à discussão do problema mente-cérebro. Essas novas técnicas traziam mais força à proposta dos Churchlands, que passou a contar com mais adeptos. Nomes como Pete Mandik, Valerie Hardcastle e outros engrossaram as fileiras dos novos neurofilósofos. A neurofilosofia torna-se mais ambiciosa e se propõe a explicar outras questões filosóficas, além dos problemas da filosofia da mente. Mas será que isso é possível? O que têm feito os neurofilósofos nos últimos anos?

Um neurofilósofo destacado é Shaun Gallagher. No seu livro, *How the Body Shapes the Mind*, ele relata como a neurociência pôde resolver a questão de Molyneux, um problema filosófico que se arrastou por séculos. William Molyneux, numa célebre carta a John Locke, há 300 anos, fez uma pergunta que filósofos como Berkeley, Condillac e Diderot tentaram responder no século XVIII.

O que é o problema de Molyneux? Imagine uma pessoa cega de nascença que possa, através do tato, distinguir entre uma esfera e um cubo. Suponha agora que, através de uma operação, o cego subitamente recobre a visão. Será que ele poderia distinguir a esfera e o cubo somente pela visão, sem antes tocá-los?

Essa questão tornou-se particularmente séria para os empiristas nos séculos XVII e XVIII. Eles precisavam sustentar que o cego não pode fazer a distinção entre esfera e cubo e relacioná-la com a visão, sem que para isso fosse necessária a experiência. Ou seja, eles precisavam responder negativamente a essa questão. O cego precisa aprender a distinção; nunca se poderia presumir conhecimento prévio à experiência.

No século XX, Donald Hebb, um famoso psicólogo, respondeu negativamente ao problema de Molyneux. E Merleau-Ponty também abordou esse problema para tentar respondê-lo negativamente.

Todavia, só nos últimos anos se chegou a uma resposta definitiva ao problema, e graças à neurociência. Hebb, Merleau-Ponty e outros estavam certos ao respondê-la negativamente. Na verdade, o que a neurociência mostrou é que é impossível resolver o problema de Molyneux ao elucidar que existe um período crítico de 3 a 12 semanas no início da infância, no qual a experiência visual é necessária para que o córtex visual se desenvolva. Um cego de nascença, permanecendo nessa condição durante esse período, não pode mais recuperar a visão.

Locke, Hebb e Merleau-Ponty estão comprovadamente corretos na sua resposta negativa ao problema de Molyneux, mas por razões diferentes daquelas que a neurociência aponta. Não é o empirismo que permite sustentar uma resposta negativa ao problema de Molyneux, mas o fato bruto de que ele não faz sentido. O paciente de Molyneux nunca poderia sequer enxergar, pois nunca poderia ter desenvolvido seu córtex visual. Mas, será que com isso, a neurociência teria dissolvido um problema filosófico clássico?

Certamente, afirmar isso não passa de um exagero. Ela é, antes de tudo, mais uma tentativa de seu refinamento em termos de análise conceitual. E aqui podemos detectar que neurofilosofia e materialismo eliminativo são posições filosóficas distintas, apesar de, muitas vezes, entrelaçarem-se.

Como toda nossa percepção e raciocínio passam pelo cérebro, podemos até supor que a explicação de sua natureza deve ser parcialmente obtida pela neurociência. Mas isso não implica que possamos reduzir todos esses fenômenos – a percepção e o raciocínio – a neurônios. Fazer isso consiste em adotar uma perspectiva reducionista, e é isso o que, infelizmente, os Churchlands fazem quando tentam eliminar inteiramente a psicologia e a filosofia e substituí-las pela neurociência. Mas esse tipo de reducionismo, que no caso dos Churchlands é chamado de *materialismo eliminativo*, leva a neurofilosofia a um compromisso que não parece ser dela. Não é preciso reduzir os problemas filosóficos a regiões do cérebro ou a neurônios para se praticar a neurofilosofia.

Em outras palavras, a neurofilosofia se contenta em colaborar com os problemas da filosofia da mente. O materialismo eliminativo deseja superar esses problemas. É possível que em algum momento a neurofilosofia precise adotar uma postura eliminativista para resolver algum problema. Mas dificilmente acharemos um manual completo de ma-

terialismo eliminativo, ou seja, um manual que explique todo funcionamento mental. O eliminativista retrucará que isso não ocorreu ainda, mas não é impossível.

Isso nos leva a um fato curioso: até hoje não encontrei nenhuma eliminação de algo mental por algo físico. Nada, desde que o materialismo eliminativo foi proposto há quase 50 anos. Talvez se adotasse esta postura mais modesta, os materialistas eliminativistas tivessem algo consensual para entregar.

O eliminativismo não parece ser um compromisso da neurofilosofia. Como nos disse Dennett no seu famoso dicionário filosófico on-line: "Os Churchlands são churchland". Ou seja, a terra que fica em volta da igreja, que antigamente era usada como cemitério. É lá que os Churchlands enterrarão as vítimas do seu materialismo eliminativo. Até que um dia a psicologia desapareça.

Referências

CHURCHLAND, P. **Neurophilosophy**: Towards a unified science of mind/brain. Cambridge: Bradford Books, 1989.

DENNETT, D. **Brainstorms**. Cambridge: The MIT Press, 1978.

GALLAGHER, S. **How the body shapes the mind**. Oxford: Clarendon Press, 2006.

NAGEL, T. Brain Bissection and the Unity of Consciousness. In: NAGEL, T. **Mortal Questions**. Cambridge: Cambridge University Press, 1979.


ANDRÉ DIAS

DEMONSTRAÇÕES GEOMÉTRICAS AUTOMÁTICAS, SIMPLES, LEGÍVEIS E INTERESSANTES♠

Pedro Quaresma

CISUC, Departmento de Matemática, Universidade de Coimbra, Portugal
pedro@mat.uc.pt
[0000−0001−7728−4935]

Pierluigi Graziani

Departmento de Ciências Puras e Aplicadas, Universidade de Urbino, Itália
pierluigi.graziani@uniurb.it
[0000−0002−8828−8920]

Resumo:

A demonstração automática de teoremas é uma área de pesquisa bem estabelecida em matemática com inúmeros métodos, programas computacionais e resultados, mas também com inúmeros problemas em aberto que realçam a sua vitalidade. Entre os problemas em aberto, os três seguintes problemas estão na agenda dos especialistas na área da dedução automática: a simplicidade de uma demonstração; a legibilidade de uma demonstração; o quão interessante um dado teorema/demonstração pode ser considerado.

Palavras-chave:

Matemática; Geometria; teorema; demonstração automática.

♠ P. Quaresma foi parcialmente supportado pela FCT – Fundação para a Ciência e Tecnologia, I.P., no âmbito do projecto CISUC – UID/CEC/00326/2020 e por Fundos Europeus, no âmbito do Programa Operacional Centro 2020. P. Graziani foi parcialmente suportado pelo Ministério da Educação de Itália, Universidade e Investigação, no âmbito do projecto PRIN 2017, "The Manifest Image and the Scientific Image", prot. 2017ZNWW7F_004.

1. Introdução

A Matemática é um campo de conhecimento formal que tem como objecto central a demonstração matemática. De axiomas, e de regras de inferência deduzir novo conhecimento, novos teoremas, suportados por demonstrações, em que a conclusão segue logicamente do conjunto inicial de premissas.

A área da demonstração automática de teoremas (*Automated theorem proving*, ATP) é uma área de pesquisa bem estabelecida em matemática/ciências da computação, com inúmeros métodos, programas computacionais e resultados, mas também com inúmeros problemas em aberto que realçam a sua vitalidade[1]. Entre os problemas em aberto, os três problemas a seguir estão na agenda dos especialistas da área da dedução automática: a simplicidade de uma demonstração; a legibilidade de uma demonstração; o quão interessante um dado teorema/demonstração pode ser considerado.

– O primeiro problema pede um critério para quantificar a simplicidade de uma demonstração.
– O segundo problema pede um critério para quantificar a legibilidade de uma demonstração.
– O terceiro problema pede um critério para quantificar o grau de interesse de um teorema e/ou uma demonstração.

A relevância destes problemas pode ser aferido pela sua menção pelo matemático David Hilbert, na sua palestra intitulada "Problemas Matemáticos", no *2º Congresso Internacional de Matemáticos* realizado em 1900. Na versão impressa desta palestra (Hilbert, 1901; 1902), Hilbert destaca a relevância da existência de critérios gerais para a classificação de um problema matemático como um bom problema matemático, ou dito de outra forma um problema matemático interessante. Destacou também a importância da clareza e facilidade de compreensão e legibilidade, como características de um bom problema. Hilbert também enfatizou a relevância da simplicidade na matemática, sua conexão com o conceito de rigor, e o facto de o mesmo esforço de rigor nos obrigar a descobrir métodos de demonstração mais simples.

Todos esses três problemas têm seu próprio valor na matemática, em geral, no entanto, o presente texto irá abordá-los com um forte viés à área da geometria. Este texto pode ser visto como um panfleto, texto introdutório de um projecto de longo prazo, que

[1] Veja: *Conference on Automated Deduction*, https://cadeinc.org/ e *Association for Automated Reasoning*, https://aarinc.org/

visa abordar estes problemas, não apenas com novas análises dos três problemas, mas também investigando as suas conexões, tanto usando estudos teóricos, como métodos empíricos, e avaliando sua implementação na demonstração e descoberta automática de teoremas em geometria.

2. Simplicidade de uma Demonstração Matemática

No texto da conferência acima citado, Hilbert analisa 23 problemas (muito mais do que os analisados durante a conferência). No entanto, Hilbert também refletiu sobre um vigésimo quarto problema dedicado à simplicidade. Esse problema foi então excluído da versão impressa, e só descoberto posteriormente em um de seus cadernos (Thiele, 2003; Thiele & Wos, 2002)[2]:

> O vigésimo quarto problema da minha palestra em Paris seria: *Critérios de simplicidade*, ou prova da maior simplicidade de certas demonstrações. Desenvolver uma teoria do método de demonstração em matemática, em geral. Sob um determinado conjunto de condições, só poderá existir uma demonstração considerada a mais simples. Em geral, se há duas demonstrações para um teorema, deverá ser possível derivar uma da outra, ou até que se torne bastante evidente quais são as variações (e auxiliares) a serem usadas nas duas demonstrações [...][3] (Thiele, 2003; Thiele & Wos, 2002).

Como observou Thiele (Thiele, 2003; Thiele & Wos, 2002), Hilbert não estava sozinho em seu desejo de máxima simplicidade em demonstrações matemáticas. O matemático francês, Emile Lemoine (1840-1912), mostrou um grande interesse na simplificação das construções geométricas propondo a, *Geometrografia*, uma forma de classificar uma construção geométrica em termos de complexidade e exactidão (Lemoine 1902). Em tempos mais recentes, é possível encontrar referências ao problema da simplicidade na pesquisa de R. Thiele e L. Wos (Thiele & Wos, 2002), ou na coletânea de artigos sobre o tema editada por I. Hipólito e R. Kahle (Hipólito & Kahle, 2019).

---

[2] "Mathematisches Notizbuch", preservado em, *Niedersaechsische Staats- und Universitaets-bibliothek Goettingen, Handschriftenabteilung* (Cod. ms. D. Hilbert 600).

[3] Notebook Cod. ms. Hilbert 600:3, p. 25-26; trans. Rudiger Thiele.

2.1. Simplicidade em Geometria

Conforme observado por Thiele, E. Lemoine abordou o problema de simplicidade através daquilo que designou por *Geometrografia*.

Geometrografia

Geometrografia, "aliás a arte das construções geométricas", visa fornecer uma ferramenta (Lemoine, 1902; Loria, 1908; Mackay, 1893; Merikoski & Tossavainen, 2010; Pinheiro, 1974; Quaresma et al., 2020; Santos, Baeta & Quaresma, 2019):

i) para designar cada construção geométrica por um número que manifesta a sua simplicidade e exactidão[4];
ii) ensinar a maneira mais simples de executar uma construção;
iii) discutir uma solução conhecida para um problema e eventualmente substituí-la por uma solução melhor;
iv) comparar diferentes soluções para um problema, decidindo qual é a mais exacta e a solução mais simples do ponto de vista da *Geometrografia*.

Em 1888, Lemoine reduziu todas as construções geométricas executadas recorrendo a uma régua e/ou compasso, a cinco operações básicas. Considerando as modificações propostas por Mackay (Mackay, 1893), as seguintes construções de régua e compasso e os coeficientes correspondentes podem ser analisados.

- Colocar a borda da régua coincidente com um ponto ..... R_1.
- Colocar a borda da régua em coincidência com dois pontos ..... $2R_1$.
- Desenhar uma linha reta ..... R_2.
- Colocar um ponto do compasso em um ponto determinado ..... C_1.
- Colocar os dois pontos do compasso em dois pontos determinados ..... $2C_1$.
- Descrever um círculo ..... C_2.

Então, uma determinada construção, é medida em relação ao número de utilizações daqueles passos elementares. Para uma determinada construção tem-se:

[4] Exactidão, ou a falta dela, neste contexto referem-se à possível imprecisão introduzida por ferramentas físicas, como uma régua ou um compasso.

$$l_1R_1 + l_2R_2 + m_1C_1 + m_2C_2$$

onde l_i e m_j são coeficientes que denotam o número de vezes que qualquer operação é executada. O número, $l_1 + l_2 + m_1 + m_2$, é designado por, *coeficiente de simplicidade (cs)* da construção, e denota o número total de operações realizadas. O número, $l_1 + m_1$, é designado por, *coeficiente de exactidão (ce)* da construção, e denota o número de operações preparatórias em que a exactidão da construção (feita com a ajuda de ferramentas físicas imprecisas) depende (Mackay, 1893; Merikoski & Tossavainen, 2010).

Geometrografia em Geometria Dinâmica

A geometrografia clássica aplica-se a construções geométricas feitas com a ajuda de uma régua e de um compasso. Sua modernização, proposta em (Quaresma et al., 2020; Santos, Baeta & Quaresma, 2019) utiliza as ferramentas dos sistemas de geometria dinâmica (DGS). Em (Quaresma et al., 2020) foi mostrado como modernizar a Geometrografia usando o *GCLC*[5] (Janičić & Quaresma, 2006), em (Santos, Baeta & Quaresma, 2019) é mostrada a generalidade da abordagem, usando o *GeoGebra*[6] (Hohenwarter, 2002). Considerando as operações: definir um ponto, em qualquer lugar do plano, D, e definir um determinado objecto, usando n pontos, C, temos os seguintes valores para as construções realizadas recorrendo ao *GCLC*:

- point – fixa um ponto no plano ……...D
- line – usa dois pontos ……... $2C$
- circle – usa dois pontos …… $2C$
- intersec – usa duas linhas …… $2C$
- intersec – usa quatro pontos …… $4C$
- intersec2 – usa um círculo e um círculo ou linha …… $2C$
- midpoint – usa dois pontos …….. $2C$
- med – usa dois pontos …….. $2C$
- bis – usa três pontos …… $3C$
- perp – usa um ponto e uma linha …….. $2C$
- foot – usa um ponto e uma linha …….. $2C$

[5] https://github.com/janicicpredrag/gclc
[6] https://www.geogebra.org/

- parallel – utiliza um ponto e uma reta ..... $2C$
- onsegment – usa dois pontos ..... $2C$
- online – usa dois pontos ..... $2C$
- oncircle – usa dois pontos ..... $2C$

Na modernização (extrapolação) da Geometrografia, considerando as "ferramentas" dos sistemas de geometria dinâmica, o coeficiente de exatidão perde seu significado. As construções serão executadas pelo DGS, portanto são precisas (exatas). Contudo, o coeficiente de simplicidade das construções ainda pode ser útil, pode ser usado para classificar construções por níveis de simplicidade. Uma nova dimensão também pode ser adicionado, o, dado pelo grau de liberdade que um determinado objeto geométrico tem, por exemplo "um ponto numa recta" tem um grau de liberdade, um ponto no plano tem dois graus de liberdade, etc. Este novo coeficiente dará um valor para o dinamismo da construção geométrica. Os graus de liberdade são medidos tendo em conta as definições dos pontos. A definição de um ponto define um ponto com dois graus da liberdade, as construções onsegment, online e oncircle, definem pontos com um grau de liberdade. Para as construções realizadas recorrendo ao *GCLC*, contidas no repositório de problemas geométricos para demonstradores automáticos de teoremas, *Thousand of Geometric problems for geometric Theorem Provers (TGTP)*[7], obteve-se um valor médio de simplicidade, CS_{gcl}, de $20,8$. Usando a análise de agrupamentos (*clustering*), k-means, função implementada no pacote estatístico do *Octave*[8], foram definidos três agrupamentos, que descrevem um nível crescente de complexidade: construções simples, $1 \leq CS_{gcl} \leq 18$; construções de complexidade média, $18 < CS_{gcl} \leq 28$; construções complexas, $CS_{gcl} > 28$. O *TGTP* contém 71 construções simples; 81 construções de complexidade média; e 28 construções complexas.

Por exemplo, o problema (GEO0369, *TGTP*): "No triângulo ΔABC, seja F o ponto médio do lado BC, e D e E as projecções ortogonais de C e B sobre AB e AC, respectivamente. FG é perpendicular a DE em G. Mostre que G é o ponto médio de DE", tem uma construção geométrica com coeficiente de simplicidade 19 (ver Fig. 1), isto é, uma construção de complexidade média. O valor de 6 para seu coeficiente de liberdade é dado pelo facto de que apenas os três pontos A, B e C estão livres no plano, enquanto todos os outros pontos estão completamente vinculados, por construção.

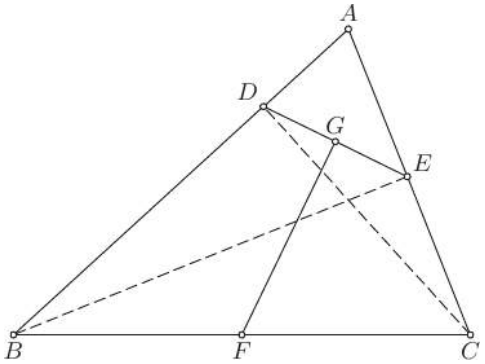
---

[7] http://hilbert.mat.uc.pt/TGTP/index.php

[8] GNU Octave, versão 6.1.1, pacote, octave-statistics, função kmeans, https://octave.sourceforge.io/statistics/function/kmeans.html

Geometrografia na Demonstração Automática de Teoremas

A mesma abordagem pode ser (novamente) extrapolada para levar em consideração as demonstrações geométricas, ou seja, demonstrações baseadas em uma teoria geométrica axiomática, usando regras de inferência geométricas.



$\mathrm{CS}_{\mathrm{gcl}} = 19 = 3\mathbf{D} = 16\mathbf{C}$
$\mathrm{CF}_{\mathrm{gcl}} = 6$

Figura 1. Coeficientes para a construção, GEO0369 *(TGTP)*

Considerando as demonstrações produzidas pelo GATP, *GCLC*, implementando o *método da área* (Janičić, Narboux & Quaresma, 2012), podemos calcular o coeficiente de simplicidade para todos os axiomas e lemas em que o método está baseado.

Além das construções geométricas nas quais as demonstrações se baseiam, (com coeficiente de simplicidade $n\mathbf{Cnst}$, existem outras etapas a serem consideradas.

- Simplificação Algébrica (Elementar) ……...…… (**AS**)
- Simplificação Geométrica (Elementar) …… (**GS**)
- Aplicação do Lema do Método da Área n ……... ($\mathbf{AML_n}$)

Uma determinada demonstração pode, portanto, ser medida em relação ao número dessas etapas. Por simplificação algébrica elementar entendem-se as operações algébricas básicas: adição, subtração, multiplicação, divisão e suas propriedades de comutatividade, associatividade e distributividade. Por simplificação geométrica elementar entende-se a aplicação directa das definições das grandezas geométricas em que o método de área está baseado. Para uma dada demonstração, expressa pela equação:

$$n_1\mathbf{Cnst} + n_2 \times \mathbf{AS} + n_3 \times \mathbf{GS} + \sum_{j=l_1}^{l_k} \mathbf{AML_j}$$

onde n_1 é o coeficiente de simplicidade da construção geométrica, n_2 é o número de simplificações algébricas e n_3 é o número de simplificações geométricas, e $\sum_{j=l_1}^{l_k} \mathbf{AML_j}$, o somatório dos coeficientes de simplicidade dos diferentes lemas usados na demonstração.

O coeficiente de simplicidade para a demonstração seria:

$$\mathrm{CS}_{\mathrm{proof}} = n_1 + n_2 + n_3 + \sum_{j=l_1}^{l_k} \mathrm{CS}_{\mathrm{proof}}\left(\mathbf{AML_j}\right)$$

O coeficiente de liberdade não tem significado neste cenário. Cada lema do método da área, $\mathbf{AML_j}$, tem um coeficiente de simplicidade correspondente, o termo, $\sum_{j=l_1}^{l_k} \mathrm{CS}_{\mathrm{proof}}\left(\mathbf{AML_j}\right)$, é a soma de todos esses valores, para todos os lemas usados na demonstração. Para conseguir isso para cada lema do método da área, calcularam-se os correspondentes coeficientes de simplicidade (Quaresma & Graziani, 2021).

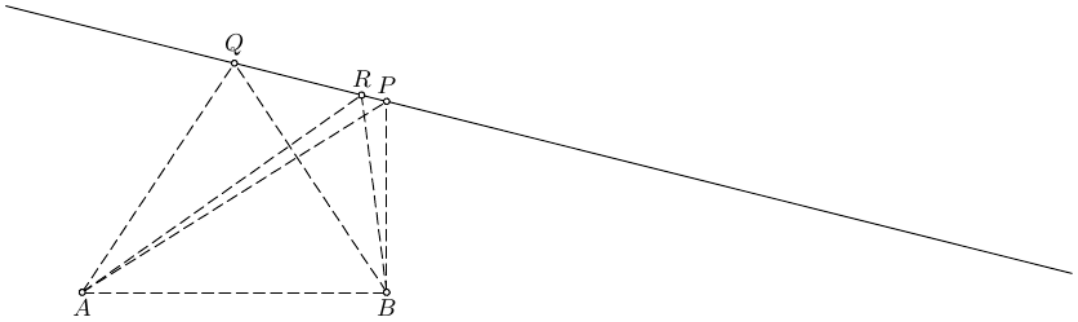
Por exemplo, a demonstração do Lema $\mathbf{AML_9}$ terá o seguinte coeficiente de simplicidade, $\mathrm{CS}_{\mathrm{proof}}(\mathbf{AML_9}) = 74$[9].

Lema ($\mathbf{AML_9}$). *Seja R um ponto na reta PQ. Então, para quaisquer dois pontos A e B tem-se que* $\mathcal{S}_{RAB} = \frac{\overline{PR}}{\overline{PQ}}\mathcal{S}_{QAB} + \frac{\overline{RQ}}{\overline{PQ}}\mathcal{S}_{PAB}$.

O que se segue é uma versão condensada, a versão completa pode ser consultada em (Quaresma & Graziani, 2021).

[9] O método da área está baseado em três quantidades geométricas e suas propriedades. A saber: a *área com sinal*, para o triângulo ΔABC, $\mathcal{S}_{ABC}$; a *razão entre segmentos paralelos* AB e CD, $\frac{\overline{AB}}{\overline{CD}}$; e a *diferença pitagórica*, para o triângulo ΔABC, $\mathcal{P}_{ABC}$.

Geometrography of Lemma 9 ($\mathbf{AML_9}$)



$$\begin{aligned} \mathrm{CS}_{\mathrm{gcl}} &= 22 = 4\mathbf{D} + 18\mathbf{C} \\ \mathrm{CF}_{\mathrm{gcl}} &= 8 \end{aligned}$$

- $s = \mathcal{S}_{ABPQ}$, construção inicial;
- $1 \times$ **GS**, áreas de triângulos com a mesma orientação, $\mathcal{S}_{RAB} = s - \mathcal{S}_{ARQ} - \mathcal{S}_{BPR}$;
- $1 \times \mathbf{AML_{14}}$, lema 14, $\frac{\overline{PR}}{\overline{PQ}} = r$ $\left(\mathrm{CS}_{\mathrm{proof}}(\mathbf{AML_{14}}) = 8\right)$;
- $1 \times \mathbf{AML_5}$, lema 5, $\frac{\mathcal{S}_{ARQ}}{\mathcal{S}_{APQ}} = \frac{\overline{RQ}}{\overline{PQ}}$ $\left(\mathrm{CS}_{\mathrm{proof}}(\mathbf{AML_5}) = 18\right)$;
- $1 \times$ **GS**, segmentos com a mesma orientação, $\frac{\overline{RQ}}{\overline{PQ}} = \frac{\overline{PQ} - \overline{PR}}{\overline{PQ}}$;
- $2 \times$ **AS**, simplificações algébricas, $\frac{\overline{PQ} - \overline{PR}}{\overline{PQ}} = (1 - r)$ e $\mathcal{S}_{ARQ} = (1 - r)\mathcal{S}_{APQ}$;
- $1 \times \mathbf{AML_5}$, lema 5, $\frac{\mathcal{S}_{BPR}}{\mathcal{S}_{BPQ}} = \frac{\overline{PR}}{\overline{PQ}}$ $\left(\mathrm{CS}_{\mathrm{proof}}(\mathbf{AML_5}) = 11\right)$;
- $2 \times$ **AS**, simplificações algébricas, $\mathcal{S}_{BPR} = r\mathcal{S}_{BPQ}$ e $\mathcal{S}_{RAB} = s - (1 - r)\mathcal{S}_{APQ} - r\mathcal{S}_{BPQ}$;
- $2 \times$ **GS**, áreas de triângulos com a mesma orientação, $\mathcal{S}_{RAB} = s - (1 - r)(s - \mathcal{S}_{PAB}) - r(s - \mathcal{S}_{QAB})$;
- $7 \times$ **AS**, simplificações algébricas, $\mathcal{S}_{RAB} = s - s + rs + \mathcal{S}_{PAB} - r\mathcal{S}_{PAB} - rs + r\mathcal{S}_{QAB}$, $\mathcal{S}_{RAB} = r\mathcal{S}_{QAB} + (1 - r)\mathcal{S}_{PAB}$ e $\mathcal{S}_{RAB} = \frac{\overline{PR}}{\overline{PQ}}\mathcal{S}_{QAB} + \frac{\overline{RQ}}{\overline{PQ}}\mathcal{S}_{PAB}$;

Geometrografia para demonstração: $4\mathbf{D} + 18\mathbf{C} + 4\mathbf{GS} + 11\mathbf{AS} + 1\mathbf{AML_{14}} + 2\mathbf{AML_5}$

$$\mathbf{AML_9}\begin{cases}\mathrm{CS_{proof}} &= 74 = 22 + 4 + 11 + 8 + (18 + 11)\\ \mathrm{CS_{gcl}} &= 22\end{cases}$$

com $\mathrm{CS_{proof}}(\mathbf{AML_{14}}) = 8$, $\mathrm{CS_{proof}}(\mathbf{AML_5}) = 18$ (primeira aplicação) e $\mathrm{CS_{proof}}(\mathbf{AML_5}) = 11$ (em segunda aplicação).

Considera-se que, a partir da segunda aplicação de um lema, a sua demonstração é aceite, portanto, é necessária apenas a sua adaptação à nova configuração, ou seja, a correspondência de padrões da configuração do lema com uma nova configuração. Por essa razão, em qualquer segunda, terceira, etc. aplicação de um lema, apenas os valores do coeficiente da construção, $\mathrm{CS_{gcl}}$, são considerados.

Dado que uma demonstração matemática é uma sequência de passos, além do coeficiente de simplicidade, seria útil ter outros coeficientes: por exemplo, o total número de passos da demonstração; o valor da passagem mais difícil da demonstração; o número de passos diferentes de alta dificuldade na demonstração; o número de diferentes tipos de passos (lemas) da demonstração; um texto, em língua natural, da demonstração; e um valor numérico assim com um gráfico de linhas da demonstração, do ponto de vista da geometrografia.

Portanto, para caracterizar completamente uma demonstração sintética formal produzida por um GATP, podemos definir e considerar os seguintes coeficientes:

- $\mathrm{CS_{proof}}$, o coeficiente de simplicidade (como acima), dá o coeficiente de simplicidade para a demonstração, de um ponto de vista global;
- $\mathrm{CT_{proof}}$, o número total de etapas da demonstração;
- $\mathrm{CS_{proofmax}}$, o maior coeficiente de simplicidade dos lemas/definições aplicáveis, fornece o coeficiente de simplicidade para a etapa mais difícil da demonstração;
- $\mathrm{CD_{typeproof}}$, o número de diferentes tipos de lemas utilizados na demonstração;
- $\mathrm{CD_{highproof}}$, o número de diferentes etapas de alta dificuldade na demonstração;
- Um texto, em língua natural, da demonstração;
- O gráfico de linhas correspondente da demonstração no formato *tikz*[10].

[10] https://ftp.eq.uc.pt/software/TeX/graphics/pgf/base/doc/pgfmanual.pdf

É importante observar que para obter o coeficiente $\mathrm{CD}_{\mathrm{highproof}}$ (hp), os lemas do método de área implementados no GATP, *GCLC*, foram analisados e, usando a função de agrupamento k-means implementada no pacote de estatísticas do *Octave*, divididos em três categorias: dificuldade baixa ($hp < 284$), dificuldade média ($284 \leq hp < 1848$) e dificuldade alta ($hp \geq 1848$).

Usando os coeficientes definidos acima, temos os seguintes valores para a demonstração do lema $\mathbf{AML_9}$:

$$\mathbf{AML_9}\begin{cases} \mathrm{CS}_{\mathrm{proof}} &= 74 = 22 + 4 + 11 + 8 + (18 + 11) \\ \mathrm{CS}_{\mathrm{gcl}} &= 22 \\ \mathrm{CT}_{\mathrm{proof}} &= 19 \\ \mathrm{CS}_{\mathrm{proofmax}} &= 18 \\ \mathrm{CD}_{\mathrm{typeproof}} &= 2 \\ \mathrm{CD}_{\mathrm{highproof}} &= 0 \end{cases}$$

Com esta extensão da Geometrografica às demonstrações produzidas pelo GATP, *GCLC*, temos a definição de um coeficiente de simplicidade para as demonstrações geométricas, produzidas automáticamente pelo GATP.

Como veremos a seguir alguns destes coeficientes vão ser importantes para a questão da legibilidade de uma demonstração geométrica.

3. Legibilidade de uma Demonstração Matemática

De importância em tudo semelhante ao problema da simplicidade, é o problema da legibilidade de uma demonstração, basta recordar obras como como as de Chou S.C., Gao X.S. e Zhang J.Z. (Chou, Gao & Zhang, 1996a, 1996b), ou aquelas de Freek Wiedijk (Wiedijk, 2000)[11].

[11] Veja também (Johnson, 1957; Kane, 1967; Li & Zhang, 1999; Noonan, 1990; Smith, 1969; Stojanović, Pavlović & Janičić, 2011; Wang & Su, 2015, 2017; Y. Wang et al., 2017; Zou & Zhang, 2011). Para uma visão geral interessante, veja também (Jiang & Zhang, 2012).

3.1. Critérios de Legibilidade

Até onde sabemos, existem duas propostas precisas para medir a legibilidade de uma demonstração (Quaresma & Graziani, 2023). A primeira é a proposta por Shin-Chun Chou et al. (Chou, Gao & Zhang, 1994, 452), o factor *Maxt-Lems*, enquanto a segunda é a proposta por Freek Wiedijk e é conhecida como de *Factor de de Bruijn* (de Bruijn, 1994; Wiedijk, 2000).

Critério Maxt-Lems

Chou et al. (Chou, Gao & Zhang, 1994, 452) propuseram uma forma de medir o quão difícil é ler o texto de uma demonstração, obtida usando um provador automatizado de teoremas para geometria (GATP) que implemente o método da área (Janičić, Narboux & Quaresma, 2012). O critério *Maxt-Lems (ML)* considera o seguinte par, (*maxt,lems*), onde:

- *maxt* é o número de termos do polinómio maximal que ocorre na demonstração automática. Assim, *maxt* mede o número de cálculos necessários na demonstração;
- *lems* é o número de lemas de eliminação usados para eliminar pontos de grandezas geométricas. Em outras palavras, *lems* indica o número de passos de inferência na demonstração.

Usando esses dois elementos e analisando todas as demonstrações feitas pelo seu GATP, eles conseguiram determinar um limite indicativo de legibilidade. De acordo com Chou et al. (Chou, Gao & Zhang, 1994, 452) uma demonstração formal, que emprega o método da área, é considerada legível se uma das seguintes condições é válida:

- o número de termos do polinómio maximal da demonstração é menor ou igual a 5;
- o número de passos de inferência na demonstração for menor ou igual a 10;
- o termo máximo da demonstração é menor ou igual a 10 e o número de passos de inferência é menor ou igual a 20.

É interessante notar que, de acordo com o seu corpus[12]: 66,9% das demonstrações têm $\text{maxt} \leq 5$, 42,6% têm $\text{lems} \leq 10$ e 73,2% têm $\text{lems} \leq 20$.

Consideremos, por exemplo, o repositório *TGTP*, especificamente, o problema GEO0001, o *Teorema de Ceva.*

Teorema (Teorema de Ceva). *Seja ΔABC um triângulo e P um qualquer ponto no plano. Seja $D = AP \cap CB$, $E = BP \cap AC$ e $F = CP \cap AB$. Mostre que: $\frac{\overline{AF}}{\overline{FB}} \times \frac{\overline{BD}}{\overline{DC}} \times \frac{\overline{CE}}{\overline{EA}} = 1$. P não deve estar nas rectas paralelas a AC, AB e BC e passando por B, C e A respectivamente (Janičić, Narboux & Quaresma, 2012).*

Com relação ao *critério ML*, considerando a demonstração feita pelo GATP, *GCLC* (Janičić & Quaresma, 2006), os valores são: $\text{maxt} = 1$ e $\text{lems} = 3$. Portanto, isso seria considerado uma demonstração legível.

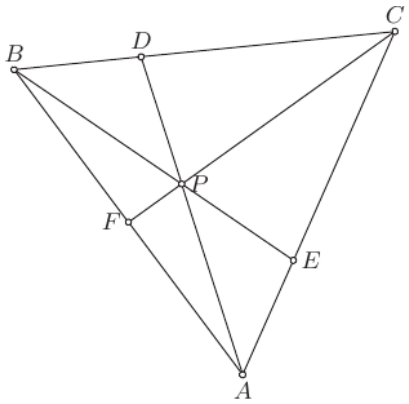


Figura 2. Teorema de Ceva, Construção Geométrica

Factor de de Bruijn

O projeto *Automath* teve como objetivo desenvolver um sistema que permitisse escrever teorias matemáticas de uma forma tão precisa que a verificação da correção dos teoremas em tais teorias poderia ser realizada por meios formais, operações

[12] Eles consideraram 478 problemas com demonstrações automáticas.

(automatizadas) aplicadas diretamente ao texto (de Bruijn, 1994). Este foi um primeiro esforço na direção da *Formalização da Matemática*, objectivo que actualmente é perseguida por pesquisadores trabalhando em sistemas como o *Coq*, *Isabelle* e *Mizar* entre outros[13]. Em "A Survey of Project Automath", de Bruijn introduziu o *factor de perda*, diferença entre o tamanho de um texto matemático comum e sua tradução formal realizada por um programa computacional. O *factor de perda* expressa o que alguém perde, em termos de síntese (laconismo), ao traduzir matemática informal para *Automath*. Wiedjk desenvolveu o conceito e o chamou de *factor de de Bruijn*. O factor de de Bruijn foi desenvolvido para um situação em que uma demonstração é criada/introduzida num computador com todos os detalhes, de tal forma que o computador pode verificar sua exactidão, por exemplo, quando para um dado teorema matemático informal existente, o texto é re-escrito, "traduzido", para uma representação formal, usando para tal um sistema como o *Automath*. Portanto, o factor de de Bruijn mede o quão eficiente é um tal sistema de conversão (Wiedijk, 2000). Wiedijk observou que questões não significativas sobre formatação poderiam afetar o cálculo do factor de perda, por exemplo: se a indentação for realizada empregando a tecla "tab", então tal indentação pode ser oito vezes menor em comparação com situações em que a indentação é feito através da tecla de "espaço"; também o nome da macro LaTeX para o símbolo '⇔' usa 15 caracteres, enquanto uma codificação como "<=>" usa apenas 3. Para que o factor de perda não fosse afectado por estas e outras opções de formatação, Wiedjk propôs compactar os arquivos antes de calcular a relação entre tamanhos. Wiedijk chama a relação entre textos não compactados, o *factor de de Bruijn aparente* e a relação entre textos compactados o *factor intrínseco de de Bruijn* (Wiedijk, 2000).

Consideramos que o *factor de de Bruijn* pode ser usado, num sentido mais lato, e.g. pode ser usado para medir o factor de perda das demonstrações matemáticas, quando produzidas por um dado GATP. Sempre que uma demonstração informal é conhecida para um determinado teorema, ela pode ser comparada com a demonstração formal produzida pelo provador automatizado de teoremas, usando uma determinada axiomatização. Isto é particularmente verdadeiro na geometria, onde uma determinada demonstração geométrica pode ser comparada com uma demonstração formal, também geométrica, produzida por um GATP.

[13] https://coq.inria.fr/, https://isabelle.in.tum.de/, http://mizar.org/

Usando novamente o Teorema de Ceva como exemplo, a legibilidade da sua demonstração formal, em relação ao factor de de Bruijn pode ser calculada[14] (ver Tabela 1).

	Informal	**Formal**	**Factor de de Bruijn**
Não comprimida	125KB	137KB	1,09
Comprimida	124KB	136KB	1,09

Tabela 1. Legibilidade da Demonstração do Teorema de Ceva – *GCLC*, Método da Área, *Factor de de Bruijn*

Wiedijk também introduziu o *limiar de Bruijn*, ou seja, um limite abaixo do qual "as pessoas começarão a usá-los (sistema tipo Automath) para trabalhos sérios". Seguindo o trabalho de Wiedijk pode-se considerar o valor de 2 como um limiar de legibilidade. Mais estudos são necessários com o fim de estabelecer um limite de legibilidade para demonstrações automatizadas, usando o factor de de Bruijn, é necessário uma comparação mais ampla entre demonstrações formais e demonstrações informais. Considerando o quociente do tamanho da demonstração formal (método da área) e o tamanho da demonstração informal, o factor de de Bruijn do Teorema de Ceva é 1,09. Seria, portanto, sensato considerar a demonstração formal, legível.

Limitações do Critério ML e Factor de de Bruijn

Uma limitação surge quando se considera a seguinte classificação (Quaresma et al., 2020) das demonstrações geométricas formais produzidas pelos GATP[15]:

1. nenhuma demonstração legível, apenas uma saída demonstrado/não demonstrado;

[14]Usou-se a demonstração que pode ser encontrada em https://artofproblemsolving.com/wiki/index.php/Ceva's_theorem como fonte para a demonstração informal.

[15] GATPs podem ser de dois tipos principais: algébricos, a demonstração, se existir, é feita recorrendo a um raciocínio algébrico (por exemplo, base de Gröbner); geométrica (sintética), a demonstração, se existir, é feita recorrendo a um conjunto de axiomas e regras de inferência da geometria, sem o uso de coordenadas. Métodos semi-sintéticos, por ex. o método da área, usa também os axiomas de um corpo de característica diferente de 2.

2. demonstração não sintética (ou seja, uma demonstração sem uma descrição geométrica correspondente, por ex. métodos algébricos);
3. demonstração semi-sintéctica com tradução correspondente na linguagem do provador;
4. demonstração (semi-)sintética com uma correspondente representação em linguagem natural;
5. demonstração (semi-)sintética com uma linguagem natural correspondente e com correspondência visual.

Relacionando o critério ML com esta classificação, podemos notar que tal critério só permite a definição de um limite para demonstrações semi-sintéticas que empregam o método da área (nível 3). A aplicabilidade direta do critério ML a outros métodos sintéticos, por ex. métodos de ângulo completo ou o método das bases de dados dedutivas (Chou, Gao & Zhang, 1996b; Ye, Chou & Gao, 2010), seria possível, considerando o número de etapas de dedução das demonstrações e adaptando a condição quanto ao número de termos do polinómio máximal nas demonstrações.

O factor de de Bruijn pode ser utilizado diretamente em todos os níveis acima de 1, embora é mais significativo em níveis maiores ou iguais a [sspl]. Considerando o (*GCLC*) e seus GATPs integrados baseados no *método de área*, *método de Wu* e *método das bases de Gröbner* (Janičić 2006), é possível calcular a legibilidade das demonstrações desenvolvidas usando os diferentes GATPs. É possível usar o factor de de Bruijn, para todos esses métodos, no entanto ter-se-iam de encontrar demonstrações informais correspondentes.[16]

Os dois critérios analisados são muito diferentes, o primeiro é muito específico enquanto o segundo é muito genérico.

Põe-se então a questão: será possível definir um novo critério que seja mais natural e expressivo do que o anteriores, e que possa ser generalizado para vários métodos de demonstração?

A proposta abaixo baseia-se na modernização da *Geometrografia* (Lemoine 1902; Quaresma et al., 2020; Santos, Baeta & Quaresma, 2019).

[16] O método de Wu e o método das bases de Gröbner são ambos métodos algébricos, do ponto de vista geométrico as suas demonstrações são ilegíveis (nível 2).

Um Critério Geometrográfico

É interessante notar como os coeficientes geometrográficos, acima descritos (§ 2.1), destacam muitos dos aspectos salientes de uma demonstração geométrica, aspectos que podem ser usados para analisar a legibilidade de tais demonstrações. Além disso, é interessante ressaltar como o gráfico da demonstração constitui uma espécie de eletroencefalograma da máquina enquanto demonstra o teorema. Assim como um eletroencefalograma pode ser útil para medir a actividade cerebral, o gráfico da demonstração ajuda a entender o esforço dispendido na compreensão da demonstração (Quaresma & Graziani, 2023).

Aplicando a *Geometrografia* às demonstrações do método de área contidas no repositório *TGTP*, usando o GATP *GCLC*, e usando os coeficientes geometrográficos podemos argumentar em favor do seguinte coeficiente de legibilidade:

Coeficiente de Demonstrações de Legibilidade Geometrográfica (GRCP)[17]

$$GRCP = \left(\left(\mathrm{CS}_{\mathrm{proof}} - \mathrm{CT}_{\mathrm{proof}}\right) \times \left(\mathrm{CD}_{\mathrm{highproof}} + \mathrm{CD}_{\mathrm{typeproof}}\right)\right)$$

Este coeficiente relaciona quatro quantidades: o coeficiente de simplicidade da demonstração, o número total de etapas da demonstração, o número de diferentes etapas com alta dificuldade na demonstração, o número de diferentes lemas usados na demonstração.

O primeiro factor, $\left(\mathrm{CS}_{\mathrm{proof}} - \mathrm{CT}_{\mathrm{proof}}\right)$, dá uma aproximação para o coeficiente geral de simplicidade das etapas não triviais na demonstração. É possível observar que $\mathrm{CT}_{\mathrm{proof}}$ conta o número de passos em vez do coeficiente de simplicidade de cada etapa. Por outro lado, em $\mathrm{CS}_{\mathrm{proof}}$, é o coeficiente de simplicidade que conta. Cada passo trivial tem um coeficiente de simplicidade igual a um, e os coeficientes de simplicidade para etapas não triviais, como a construção e os lemas, são muito maiores que um. À luz disto, pode-se concluir que a diferença entre $\mathrm{CS}_{\mathrm{proof}}$ e $\mathrm{CT}_{\mathrm{proof}}$ enfatiza a complexidade da demonstração, desconsiderando a sua extensão.

O segundo factor, $\left(\mathrm{CD}_{\mathrm{highproof}} + \mathrm{CD}_{\mathrm{typeproof}}\right)$, dá conta do número de passos difíceis. Passos que, potencialmente, tornam a demonstração muito mais difícil de seguir, etapas onde o fluxo normal da demonstração seria interrompido para passar para a demonstração do lema, retomando após o completar da demonstração do lema. A adição

[17] *Geometrographic Readability Coefficient of Proofs (GRCP).*

do número de etapas de alta dificuldade com o número de diferentes lemas usados no demonstração, providencia um factor multiplicador para a complexidade geral da demonstração. Uma nota final sobre este segundo factor: um passo de alta dificuldade é, com certeza, uma aplicação de um lema, no entanto sentimos que a natureza de alta dificuldade do lema em questão, é uma razão suficiente para que esta dupla contagem não seja significativa no computo geral.

Multiplicando estes factores, a aproximação para o coeficiente de simplicidade e o número de etapas difíceis — ambos os elementos que acreditamos caracterizar a legibilidade de uma demonstração — obtemos um coeficiente de legibilidade para as demonstrações geométrica produzidas por GATP.

Considerando os 71 teoremas e suas demonstrações do método de área, contidos no repositório *TGTP* e usando, novamente, a análise de agrupamentos, k-means, função do *Octave*, as demonstrações podem ser divididas nas seguintes classes de legibilidade geometrográfica:[18] legível (alta legibilidade), $\mathrm{GRCP} \leq 48000$; legibilidade média, $48000 < \mathrm{GRCP} \leq 135000$; baixa legibilidade, $\mathrm{GRCP} > 135000$.

O valor de *GRCP* para o exemplo GEO0001, é: $\mathrm{GRCP}_{\mathrm{GEO0001}} = (220 - 32) \times (0 + 3) = 564 \leq 48000$, portanto uma demonstração legível (alta legibilidade).

***Exemplo de legibilidade média do GRCP*. Problema GEO0021 contido no repositório TGTP:**

Teorema (Circuncentro de um Triângulo). *O circuncentro de um triângulo pode ser encontrado como a intersecção das três bissectrizes perpendiculares* tem os seguintes valores para os diferentes coeficientes.

$$\mathrm{GEO0021}\begin{cases} \mathrm{CS}_{\mathrm{proof}} = 8554 \\ \mathrm{CS}_{\mathrm{gcl}} = 11 \\ \mathrm{CT}_{\mathrm{proof}} = 591 \\ \mathrm{CS}_{\mathrm{proofmax}} = 2807 \\ \mathrm{CD}_{\mathrm{typeproof}} = 13 \\ \mathrm{CD}_{\mathrm{highproof}} = 3 \end{cases}$$

$$48000 < GRCP = 127408 \leq 135000$$

[18] Os valores obtidos foram arredondados.

Pelo critério *GRCP*, este é um problema de legibilidade média. Pode-se ver que ele possui 13 lemas diferentes, 3 etapas de alta dificuldade, um demonstração longa com uma diferença significativa entre o CS_{proof} e o número de etapas da demonstração (ver Figura 3).

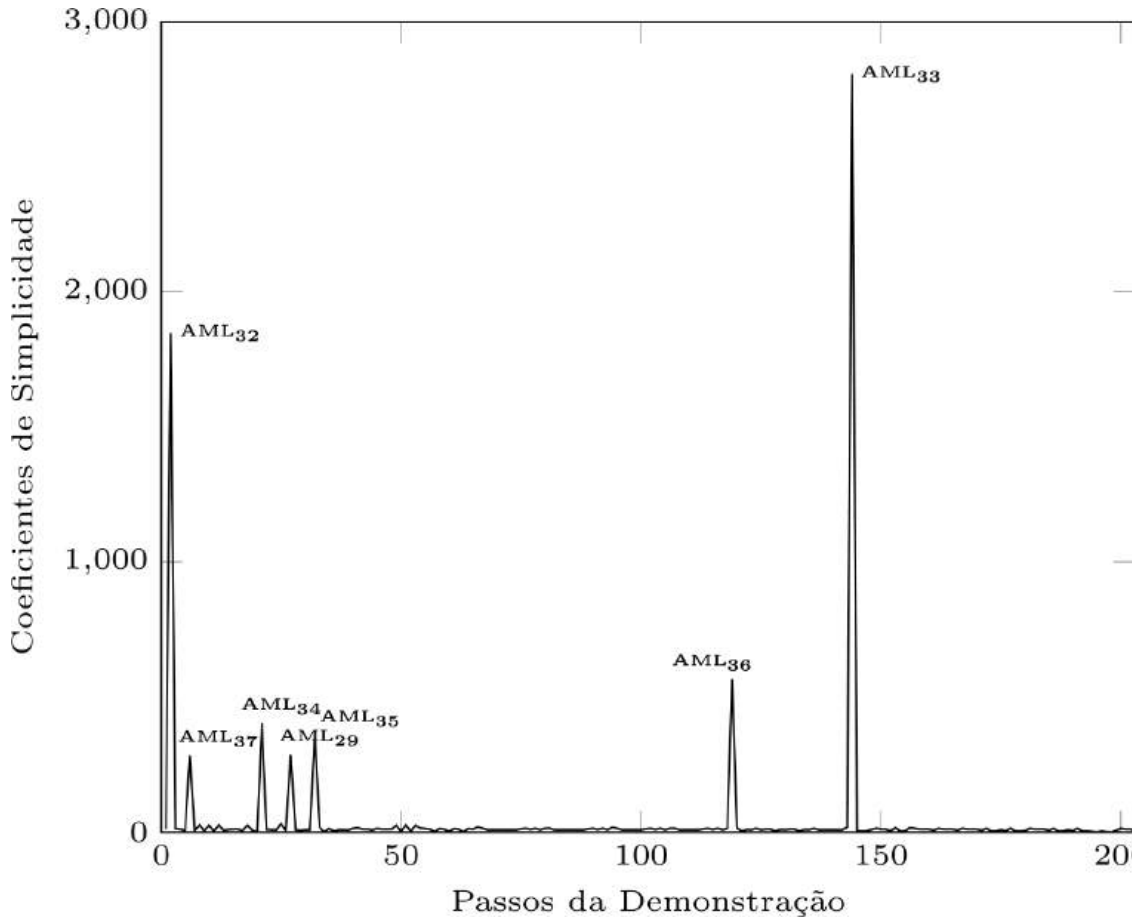


Figura 3. *TGTP, GEO0021, Circuncentro de um Triângulo*

Comparando os Diferentes Critérios

O critério do *GRCP* leva em consideração todos os aspectos significativos de uma demonstração formal, sua dificuldade geral, seu número de passos, o número de passos difíceis e o número de lemas diferentes que devem ser aplicados. Os demais critérios consideram menos aspectos. O *factor de de Bruijn*, dado seu objetivo inicial, leva em consideração apenas o tamanho da demonstração e precisa ter uma demonstração informal para comparar. O *critério ML* considera o número de diferentes lemas aplicados

e usa o número de termos do polinómio máximo como forma de se aproximar da complexidade da demonstração. Tanto no critério ML, como no critério *GRCP*, tem-se em conta o número de lemas numa demonstração: no critério *GRCP* como factor multiplicativo, no critério ML como uma das condições de legibilidade. No critério ML o número de termos no polinómio máximo são considerados, mas, como observaram os seus autores, isso mede o número de cálculos necessários na demonstração, não a sua legibilidade. A legibilidade está fracamente relacionado com o número de etapas da demonstração, o que faz é dar uma aproximação ao número de passos necessários para decompor esses longos polinómios que ocorrem na demonstração em um expressão simples.

Independentemente desta comparação dos diferentes critérios, queremos enfatizar que a visão Geometrográfica tem um alcance mais geral. A qualidade da abordagem Geometrográfica, através da análise de vários coeficientes das demonstrações, aos que se pode adicionar o gráfico da demonstração (ver Figura 3), é dada pela possibilidade de definir uma linguagem que pode ser usado por não especialistas para formular outros critérios mais fracos ou mais fortes do que aquele acima proposto. Temos a possibilidade de ir além de um dado critério geometrográfico, considerando uma abordagem geometrográfica do problema da medição da legibilidade de demonstrações formais produzidas por GATP. Uma abordagem que oferece um ambiente para analisar as demonstrações em detalhes, propor e testar critérios de legibilidade. Até onde sabemos, é o primeiro momento em que a comunidade tem acesso a uma ferramenta tão geral para o formular e o estudar da legibilidade de demonstrações formais em dedução automática em geometria. É também interessante notar que o critério *GRCP* oferece uma classificação de demonstrações que está em linha, quando considerados os pontos fundamentais, com as classificações dadas pelos outros dois critérios referidos. Isto é, as demonstrações que são classificadas como difíceis de ler de acordo ao novo critério também são classificados como de difícil leitura para os demais, e os o mesmo se aplica a demonstrações fáceis de ler (Tabela 2).

Finalmente, temos que dizer que todos os critérios aqui propostos não têm nenhuma validação através da submissão de testes a estudantes, especialistas etc. a grande vantagem que a nossa abordagem oferece é que nos permite formular critérios que podem ser, posteriormente, implementados em repositórios como *TGTP*, podendo ser avaliado experimentalmente de uma forma muito simples.

TGTP	ML	de Bruijn	GRCP
GEO0001	3 < 5, passos deductivos _____ fácil	1.6 < 2 _____ fácil	564 ≤ 48000 _____ fácil (alta)
GEO0021	13 > 5 passos deductivos & número de termos > 5 _____ difícil	37.63 > 2 _____ difícil	127408 ≤ 135000 _____ difícil (média)
GEO0020	13 > 5 passos deductivos & número de termos > 5 _____ difícil	47.31 > 2 _____ difícil	269790 > 135000 _____ difícil (baixa)

Tabela 2. Comparação dos Três Critérios

4. Teoremas Matemáticos Interessantes

Como foi já referido Hilbert destacou a relevância da existência de critérios gerais para a classificação de um problema matemático como um bom problema matemático. Destacou também a importância da clareza e facilidade de compreensão, pode-se dizer, simplicidade, assim como a legibilidade, como características de um bom (interessante) problema. Larry Woss encarou, do ponto de vista computacional, esta procura da existência de critérios gerais para a classificação do quão interessante um problema matemático pode ser considerado.

Larry Woss, listou o problemas da geração automática de problemas interessantes como um dos problemas básicos na área da dedução automática (Wos, 1988).

> [31º problema de Wos] Que propriedades podem ser identificadas para permitir a construção de um programa de dedução automática capaz de

encontrar teoremas novos e interessantes, em vez de simplesmente demonstrar conjecturas?

A busca de quais são as propriedades que podem ser identificadas para permitir um programa de dedução automática encontrar teoremas interessantes é um objetivo de pesquisa interessante (trocadilho intencional) (Colton, Bundy & Walsh, 2000; Gao & Cheng, 2017; Gao, Li & Cheng, 2018, 2019; Puzis, Gao & Sutcliffe, 2006; Quaresma, Graziani & Nicoletti, 2023). O 31º problema de Wos, refere os problema da geração e da classificação. No que se segue aborda-se somente a questão da classificação de um teorema como um teorema interessante, i.e. do encontrar critérios para a classificação de um problema geométrico como interessante.

A questão de gerar teoremas interessantes já foi abordado por vários autores (Colton, Bundy & Walsh, 2000; Gao & Cheng, 2017; Gao, Li & Cheng, 2018, 2019; Puzis, Gao & Sutcliffe, 2006). As diferentes abordagens encontradas na literatura (Colton, 2002; Gao, Goto & Cheng, 2014; Puzis, Gao & Sutcliffe, 2006) compartilham, em suas linhas gerais, o mesmo algoritmo: para um determinado fragmento lógico, seleciona-se um conjunto inicial de factos, referentes às propriedades intrínsecas à construção geométrica, de seguida tem-se um ciclo de geração/filtragem, até que uma da dada condição de paragem seja correspondida.

Filtrando Teoremas Interessantes

Um primeiro nível de filtragem deve descartar as tautologias óbvias e também as conjecturas provadas falsas por evidências empíricas.

A filtragem por teoremas interessantes ou por teoremas desinteressantes, dois lados da mesma moeda, é feito através da aplicação de uma série de filtros. Os filtros até agora encontrados na literaturaque aborda o assunto, são de natureza especulativa, não tendo ainda sido validados de nenhuma forma (Colton, Bundy & Walsh, 2000; Gao, Li & Cheng, 2018, 2019; Puzis, Gao & Sutcliffe, 2006).

Óbvio:
o número de inferências na derivação. Este filtro tenta estimar o quão óbvio é uma dada fórmula, através do número de inferências usadas na derivação da sua demonstração.

Peso:
esforço necessário para ler uma fórmula. O "peso" de uma fórmula pode ser estimado pelo número de símbolos contidos na fórmula.

Complexidade:
o esforço necessário para compreender uma fórmula. O número de símbolos de funções e predicados contidos na fórmula.

Surpresa:
mede novas relações entre conceitos e propriedades.

Intensidade:
mede o quanto uma fórmula resume as informações dos nós ancestrais em sua árvore de derivação.

Adaptabilidade:
mede quão rigorosamente as variáveis universalmente quantificadas estão restritas (para fórmulas na forma clausal).

Foco:
mede até que ponto uma fórmula gera resultados positivos ou negativos no domínio de aplicação.

Utilidade:
mede o quanto um teorema interessante contribuiu para a demonstração de outros teoremas interessantes.

Apesar da relevância destas métricas, seria apropriado construir um inquérito, com um número significativo de especialistas, com vista à sua validação. Este tipo de inquérito seria relevante não apenas para enfrentar o problema de Wos, mas também para melhor entender como construir e avaliar os programas que geram/encontram teoremas interessantes.

Resultado de Indecidibilidade

Na secção anterior foi discutida a aplicação de filtros, esses filtros são baseados em algumas medidas de interesse que ainda precisam ser validadas e que são aplicado de forma heurística. Põe-se a questão: é possível ter uma abordagem determinística, isto é, é possível escrever um programa de computador que, de forma determinística, encontre teoremas interessantes? De seguida mostramos ser não decidível determinar, para uma dada Máquina de Turing, se a linguagem por ela reconhecida tem a propriedade (não-trivial) de encontrar teoremas interessantes (Quaresma, Graziani & Nicoletti, 2023). A demonstração deste resultado usa o *Teorema de Rice* (ver Lema 1) (Rice, 1953; Rogers Jr, 1987; Sipser, 1997).

Definição 1 (Propriedade Não-Trivial). *Uma propriedade p de uma linguagem formal é não-trivial se:*

- *existe uma linguagem recursivamente enumerável com a propriedade p;*
- *existe uma linguagem recursivamente enumerável que não possui a propriedade p.*

Lema 1 (Teorema de Rice). *Seja p qualquer propriedade não-trivial da linguagem de uma máquina de Turing. O problema de determinar se uma determinado linguagem de uma máquina de Turing tem a propriedade p é indecidível.*

Teorema 1 (Resultado de Indecidibilidade). *Para qualquer Máquina de Turing, é indecidível determinar se a linguagem por ela reconhecida tem a propriedade de encontrar teoremas interessantes.*

Demonstração. Todos os programas (máquinas de Turing) capazes de demonstração automática de teoremas em geometria e, por extensão, gerar/encontrar teoremas geométricos, dependem de uma linguagem formal para descrever as construções geométricas, conjecturas e demonstrações.

Por exemplo, podemos considerar a linguagem para a lógica de primeira ordem, *First-Order Form* (FOF)[19] definida no repositório *TPTP* (Sutcliffe, 2017). Considerando de seguida as teorias axiomáticas formais para a geometria escritas nessa linguagem.

[19] http://tptp.cs.miami.edu/TPTP/QuickGuide/

Seja p a propriedade daquela linguagem que diz que o teorema t é interessante, para qualquer definição concebível de interesse, então existe um linguagem enumerável com a propriedade p. Será suficiente restringir o linguagem de tal forma que o teorema t, e somente este, fosse reconhecido. Mas, também existe uma linguagem recursivamente enumerável que não possui a propriedade p. Bastaria restringir essa linguagem de tal forma que apenas as tautologias seriam reconhecidas. Tautologias são, para qualquer definição concebível de interesse, desinteressantes. Provamos que p, a propriedade que pode estabelecer se um determinado teorema é interessante, é uma propriedade não-trivial.

Tendo estabelecido que a propriedade p é não-trivial, então, pela aplicação de Teorema de Rice, é indecidível determinar para qualquer máquina de Turing M, se a linguagem reconhecida por M tem a propriedade p.

Por outras palavras, é indecidível ter um programa determinístico que possa encontrar problemas interessantes. Na melhor das hipóteses, esta é uma tarefa a ser abordada por programas baseado em algoritmos guiados por critérios heurísticos.

À luz do nosso resultado de indecidibilidade, verificar se um teorema é interessante continua a ser uma tarefa inerentemente humana ou, na melhor das hipóteses, para um determinado programa de computador guiado por heurísticas. No entanto, para que este meta-nível (heurístico) possa ser definido, é preciso chegar a um grau mínimo de concordância sobre a definição de teoremas interessantes. Para alcançar algo que possa ser a base da definição de um tal meta-nível, acreditamos que uma exploração empírica da noção de interessante e do que implica concretamente é primordial. Abordaremos esta possibilidade nas conclusões.

5. Conclusões

Como já foi referido acima este é um projecto de longo prazo o qual visa abordar o problema da simplicidade e legibilidade das demonstrações geométricas assim como o problema do analisar de quão interessante um teoremas e/ou demonstração geométrica possa ser, sempre na perspetiva de um sistemas automático de demonstrações geométricas. Por esta razão, essas questões serão analisadas considerando diferentes perspetivas. A primeira etapa adotará uma visão qualitativa: utilizaremos o método da epistemologia histórica para analisar os três problemas. A epistemologia histórica é "a história do categorias que estruturam nosso pensamento, modelam nossos argumentos e demonstrações e certificam nossos padrões para a compreensão do factos. A epistemologia histórica pode (na verdade, deve ser) instanciada pela história das ideias,

mas coloca um tipo diferente de questão: não a história disto ou daquilo uso particular de, digamos, infinitesimais nas demonstrações matemáticas dos séculos XVI e séculos XVII, mas a história das mudanças nas formas e padrões da matemática manifestadas durante este período" (Daston, 1993). Usando esse método analisaremos a história das mudanças nas formas e padrões de demonstrações geométricas através de diferentes culturas matemáticas e épocas, destacando as ideias relacionadas à simplicidade e legibilidade das demonstrações geométricas, e sobre o interesse de um dado problema geométrico. Usando os dados colectados nesta primeira etapa, tentaremos definir alguns desideratos relativos aos problemas em questão. A segunda etapa adotará uma visão quantitativa: conceberemos inquéritos para serem respondidos por especialistas para melhor entender o uso dos adjectivos simples, legível e interessante, em relação às demonstrações e teoremas. Uma etapa final acabará por ser a aplicação dos resultados achados na concepção e implementação de programas que possam ir ao encontro da descoberta de novos problemas interessantes, com demonstrações simples e legíveis.

Referências

Chou, Shang-Ching, Xiao-Shan Gao & Jing-Zhong Zhang. Machine Proofs in Geometry. **World Scientific**, 1994. https://doi.org/10.1142/2196.

Chou, Shang-Ching, Xiao-Shan Gao & Jing-Zhong Zhang. Automated Generation of Readable Proofs with Geometric Invariants, I. Multiple and Shortest Proof Generation. **Journal of Automated Reasoning**, 17 (3): 325–47, 1996a. https://doi.org/10.1007/BF00283133.

Chou, Shang-Ching, Xiao-Shan Gao & Jing-Zhong Zhang. Automated Generation of Readable Proofs with Geometric Invariants, II. Theorem Proving with Full-Angles. **Journal of Automated Reasoning**, 17 (3): 349–70, 1996b. https://doi.org/10.1007/BF00283134.

Colton, Simon. The HR Program for Theorem Generation. In: **Automated Deduction—CADE-18**, Andrei Voronkov (Ed.), 285–89, 2002. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45620-1_24.

Colton, Simon, Alan Bundy & Toby Walsh. On the Notion of Interestingness in Automated Mathematical Discovery. **International Journal of Human-Computer Studies**, 53 (3): 351–75, 2000. https://doi.org/10.1006/ijhc.2000.0394.

Daston, Lorraine. Historical Epistemology. In: **Questions of Evidence:** Proof, Practice, and Persuasion Across the Disciplines. James Chandler, Arnold I. Davidson & Harry D. Harootunian (Eds.), 282–89, 1993. The University of Chicago Press.

de Bruijn, N. G. Selected Papers on Automath. In: R. C. Vrijer, R. P. Nederpelt & J. H. Geuvers (Eds.), 133:41–161, 1994. **Studies in Logic and the Foundations of Mathematics**. Amsterdam: North-Holland.

Gao, Hongbiao & Jingde Cheng. Measuring Interestingness of Theorems in Automated Theorem Finding by Forward Reasoning: A Case Study in Peano's Arithmetic. In: **Intelligent Information and Database Systems**. Ngoc Thanh Nguyen, Satoshi Tojo, Le Minh Nguyen & Bogdan Trawinski (Eds.). 10192:115–24, 2017. Lecture Notes in Computer Science. Springer International Publishing. https://doi.org/10.1007/978-3-319-54430-4_12.

Gao, Hongbiao, Yuichi Goto & Jingde Cheng. A Systematic Methodology for Automated Theorem Finding. **Theoretical Computer Science**, 554 (October): 2–21, 2014. https://doi.org/10.1016/j.tcs.2014.06.028.

Gao, Hongbiao, Jianbin Li & Jingde Cheng. Measuring Interestingness of Theorems in Automated Theorem Finding by Forward Reasoning: A Case Study in Tarskis Geometry. In: **2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation** *(SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE. 2018. https://doi.org/10.1109/SmartWorld.2018.00064.

Gao, Hongbiao, Jianbin Li & Jingde Cheng. Measuring Interestingness of Theorems in Automated Theorem Finding by Forward Reasoning Based on Strong Relevant Logic. In: **2019 IEEE International Conference on Energy Internet** *(ICEI)*, 356–61. IEEE. 2019. https://doi.org/10.1109/ICEI.2019.00069.

Hilbert, David. Mathematische Probleme. **Arch. Für Math. Phys**, 3 (1): 44–63, 213–37, 1901.

Hilbert, Davida. Mathematical Problems. **Bulletin of the American Mathematical Society**, 8: 437–79, 1902.

Hipólito, Inês & Reinhard Kahle (Eds.). **The notion of simple proof – Hilbert's 24th problem**. Vol. 377, 2019. Philosophical Transactions of the Royal Society A.

Hohenwarter, M. GeoGebra - a Software System for Dynamic Geometry and Algebra in the Plane. **Master's thesis, Austria:** University of Salzburg, 2002.

Janičić, Predrag. GCLC — A Tool for Constructive Euclidean Geometry and More Than That. In: **Mathematical Software - ICMS 2006**. Andrés Iglesias & Nobuki Takayama (Eds.), 4151:58–73, 2006. Lecture Notes in Computer Science. Springer. https://doi.org/10.1007/11832225_6.

Janičić, Predrag, Julien Narboux & Pedro Quaresma. The Area Method: A Recapitulation. **Journal of Automated Reasoning**, 48 (4): 489–532, 2012. https://doi.org/10.1007/s10817-010-9209-7.

Janičić, Predrag & Pedro Quaresma. System Description: GCLCprover + GeoThms. In: **Automated Reasoning**. Ulrich Furbach & Natarajan Shankar (Eds.), 4130:145–50, 2006. Lecture Notes in Computer Science. Springer. https://doi.org/10.1007/11814771_13.

Jiang, Jianguo & Jingzhong Zhang. A Review and Prospect of Readable Machine Proofs for Geometry Theorems. **Journal of Systems Science and Complexity**, 25 (4): 802–20, 2012. https://doi.org/10.1007/s11424-012-2048-3.

Johnson, Donovan A.The Readability of Mathematics Books. **The Mathematics Teacher**, 50 (2): 105–10, 1957. https://doi.org/10.5951/MT.50.2.0105.

Kane, Robert B. The Readability of Mathematical English. **Journal of Research in Science Teaching**, 5 (3): 296–98, 1967. https://doi.org/10.1002/tea.3660050319.

Lemoine, Émile. **Géométrographie Ou Art Des Constructions Géométriques**. Vol. 18, 1902. Phys-Mathématique. Scentia. http://catalogue.bnf.fr/ark:/12148/cb36049032t.

Li, Chuan-Zhong & Jhing-Zhong Zhang. Readable Machine Solving in Geometry and ICAI Software MSG. In: **Automated Deduction in Geometry**, 67–85, 1999. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-47997-x_5.

Loria, Gino. La Geometrografia e Le Sue Trasfigurazioni. **Period. Math.**, 3 (6): 114–22, 1908.

Mackay, J. S. The Geometrography of Euclid's Problems. **Proceedings of the Edinburgh Mathematical Society**, 12: 2–16, 1893. https://doi.org/10.1017/S0013091500001565.

Merikoski, K. Jorma & Timo Tossavainen. Two Approaches to Geometrography. **Journal for Geometry and Graphics**, 13 (1): 15–28, 2010.

Noonan, James. Readability Problems Presented by Mathematics Text. **Early Child Development and Care**, 54 (1): 57–81, 1990. https://doi.org/10.1080/0300443900540104.

Pinheiro, Virgílio Athayde. **Geometrografia 1**. Bahiense, 1974.

Puzis, Yury, Yi Gao & G. Sutcliffe. Automated Generation of Interesting Theorems. In: **FLAIRS Conference**, 2006.

Quaresma, Pedro & Pierluigi Graziani. The Geometrography's Simplicity Coeficient for the Axioms and Lemma of the Area Method. **Technical Report TR 2021-001**. Center for Informatics; Systems of the University of Coimbra, 2021.

Quaresma, Pedro & Pierluigi Graziani. Measuring the Readability of Geometric Proofs—the Area Method Case. **Journal of Automated Reasoning**, 67 (1), 2023. https://doi.org/10.1007/s10817-022-09652-0.

Quaresma, Pedro, Pierluigi Graziani & Stefano M. Nicoletti. Considerations on Approaches and Metrics in Automated Theorem Generation/Finding in Geometry. **ADG2023**. 2023.

Quaresma, Pedro, Vanda Santos, Pierluigi Graziani & Nuno Baeta. Taxonomy of Geometric Problems. **Journal of Symbolic Computation**, 97 (March): 31–55, 2020. https://doi.org/10.1016/j.jsc.2018.12.004.

Rice, H. G. Classes of Recursively Enumerable Sets and Their Decision Problems. **Transactions of the American Mathematical Society**, 74 (2): 358–66, 1953. https://doi.org/10.2307/1990888.

Rogers Jr, Hartley. **Theory of Recursive Functions and Effective Computability**. MIT press, 1987.

Santos, Vanda, Nuno Baeta & Pedro Quaresma. Geometrography in Dynamic Geometry. **International Journal for Technology in Mathematics Education**, 26 (2): 89–96, 2019. https://doi.org/10.1564/tme_v26.2.06.

Sipser, Michael. **Introduction to the Theory of Computation**. PWS Publishing Company, 1997.

Smith, Frank. The Readability of Junior High School Mathematics Textbooks. **The Mathematics Teacher**, 62 (4): 289–91, 1969. https://www.jstor.org/stable/27958126.

Stojanović, Sana, Vesna Pavlović & Predrag Janičić. A Coherent Logic Based Geometry Theorem Prover Capable of Producing Formal and Readable Proofs. In: **Automated Deduction in Geometry**. Pascal Schreck, Julien Narboux & Jürgen Richter-Gebert (Eds.), 6877:201–20, 2011. Lecture Notes in Computer Science. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-25070-5_12.

Sutcliffe, Geoff. The TPTP Problem Library and Associated Infrastructure. **Journal of Automated Reasoning**, 59 (4): 483–502, 2017. https://doi.org/10.1007/s10817-017-9407-7.

Thiele, Rüdger. Hilbert's Twenty-Fourth Problem. **The American Mathematical Monthly**, 110 (1): 1, 2003. https://doi.org/10.2307/3072340.

Thiele, Rüdger & Larry Wos. Hilbert's Twenty-Fourth Problem. **Journal of Automated Reasoning**, 29 (1): 67–89, 2002. https://doi.org/10.1023/A:1020537107897.

Wang, Ke & Zhendong Su. Automated Geometry Theorem Proving for Human-Readable Proofs. In: **Proceedings of the 24th International Conference on Artificial Intelligence**, 1193–99, 2015. IJCAI'15. AAAI Press. http://dl.acm.org/citation.cfm?id=2832249.2832414.

Wang, Ke & Zhendong Su. Automated Geometry Theorem Proving for Human-Readable Proofs. **ArXiv**, 2017.

Wang, Ying, Yongsheng Rao, Hao Guan & Yu Zou. NetPad: An Online DGS for Mathematics Education. In: **2017 12th International Conference on Computer Science and Education** (ICCSE). IEEE, 2017. https://doi.org/10.1109/ICCSE.2017.8085507.

Wiedijk, Freek. The de Bruijn Factor. **Poster at International Conference on Theorem Proving in Higher Order Logics** (TPHOL2000), 2000.

Wos, Larry. **Automated Reasoning:** 33 Basic Research Problems. Prentice-Hall, 1988
Ye, Zheng, Shang-Ching Chou & Xiao-Sha Gao. Visually Dynamic Presentation of Proofs in Plane Geometry, Part 2. **Journal of Automated Reasoning**, 45: 243–66, 2010. https://doi.org/10.1007/s10817-009-9163-4.

Zou, Yu & Jingzhong Zhang. Automated Generation of Readable Proofs for Constructive Geometry Statements with the Mass Point Method. In: **Proceedings of the 8th International Conference on Automated Deduction in Geometry**, 221–58, 2011. ADG'10. Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-25070-5_13.

DA "VIRADA NATURALISTA" À "VIRADA INFORMACIONAL" NA FILOSOFIA

João Antonio de Moraes[1]

Rafael Rodrigues Testa[2]

Resumo:

Neste artigo discutimos a passagem da "virada naturalista" à "virada informacional" na Filosofia, ao argumentar que o processo de desconstrução da metafísica da subjetividade ocorrido na primeira virada teria contribuído para a emergência da segunda. Evidenciamos, com isso, como a concepção cartesiana de ser humano como único possuidor de alma e medida de todas as coisas passou para um cenário no qual ele é concebido como apenas mais uma animal dentre outros (pela influência do darwinismo) e, posteriormente, como a exclusividade e superioridade do ser humano foi novamente afetada com o desenvolvimento de "modelos mecânicos da mente". Por fim, justificamos a relevância

Abstract:

In this paper it is discussed the passage from the "naturalist turn" to the "informational turn" in Philosophy. We argue that the process of deconstruction of the metaphysics of subjectivity that occurred in the first turn would have contributed to the emergence of the second. Accordingly, we show how the Cartesian conception of the human being as the sole possessor of soul and measure of all things moved to a scenario in which the human being is conceived as just another animal among others (due to the influence of Darwinism) and, later, how the exclusivity and superiority of the human being was again affected with the development of "mechanical models of the mind". Finally, we

1 Doutor em Filosofia pela Universidade Estadual de Campinas (UNICAMP). Coordenador e Docente do Curso de Filosofia presencial da Faculdade João Paulo II (FAJOPA). Autor dos livros "Implicações éticas da 'virada informacional na Filosofia'" (EDUFU, 2014) e "O Paradigma da Complexidade e a Ética Informacional" (CLE-UNICAMP, 2019). Presidente da Sociedade Brasileira de Ciências Cognitiva (SBCC) nos anos de 2019 a 2021. E-mail: moraesunesp@yahoo.com.br ORCID: https://orcid.org/0000-0001-6057-5138

2 Doutor em Filosofia pela Universidade Estadual de Campinas (UNICAMP). Pesquisador associado do Centro de Lógica, Epistemologia e História da Ciência (CLE-UNICAMP). E-mail: mail@rafaeltesta.com.br ORCID: https://orcid.org/0000-0002-7052-1376

da análise da subjetividade também pela ótica da Ética Informacional, tendo em vista a crescente incorporação das tecnologias no cotidiano dos indivíduos, bem como as características subjacentes a tais tecnologias.

Palavras-chave:

Metafísica da subjetividade. Virada naturalista. Virada informacional. Ciência Cognitiva. Ética Informacional.

justify the relevance of analyzing subjectivity also from the perspective of Information Ethics, in view of the increasing incorporation of technologies in the daily lives of individuals, as well as the underlying characteristics of such technologies.

Keywords:

Metaphysics of subjectivity. Naturalistic turn. Informational turn. Cognitive Science. Information Ethics.

1. A construção da metafísica da subjetividade no pensamento moderno

O pensamento moderno tem sua origem no séc. XVII, resultante de um processo de reavaliação do lugar do homem no mundo. Com o grande número de teorias científicas desenvolvidas neste período (como por exemplo a lei do movimento dos corpos e o método experimental na física, de Galileu Galilei; a lei da gravitação universal, fundamental para a compreensão da física clássica e da mecânica, de Isaac Newton; o método cartesiano e a dualidade mente-corpo, proposto por Descartes, dentre outras), há uma alteração no modo de compreender o próprio mundo: recusa-se a concepção de *cosmos* imutável para adoção de um mundo no qual as leis da física poderiam, de certa forma, explicar seu funcionamento. Adota-se uma concepção mecanicista de universo, a partir da qual este é entendido como constituído por pequenas partículas que mantêm relações externas entre si e se movem em conformidade com determinadas leis universais, enfatizando-se, posteriormente, a capacidade da razão humana de investigar de forma crítica e empírica tais leis que regem a natureza e a sociedade.

Na Modernidade, o *ser humano*, através da razão, possui a expectativa de controle sobre a natureza, sendo considerado o único ser "digno de respeito". Um exemplo dessa concepção, destacado por exemplo por Ferry (2007, p. 126), ressalta a publicação da *Declaração dos Direitos Humanos* em 1789, que "instala o homem no centro do mundo. [...] Ela faz dele não apenas o único ser sobre a Terra, verdadeiramente digno de respeito, mas também propõe a igualdade de todos os seres humanos". Com isso, completa o autor, a filosofia moderna é, antes de tudo um, humanismo. Ademais, ao destacar que o *ser humano* se julga o único ser digno de respeito moral, o pensamento moderno consolida um

antropocentrismo que fará crescer as raízes da subjetividade – com o início de um debate sobre a diferença entre o ser humano e os outros animais.

Em seu estudo sobre o que diferenciaria o ser humano do animal, Descartes (1973a; 1973b) concebe que a natureza humana é constituída por duas substâncias distintas, corpo e alma, aplicando seu "método de análise e síntese", que é composto por quatro preceitos: (I) *duvidar* do objeto de investigação; (II) *dividir*, na análise, cada uma das dificuldades em quantas partes forem necessárias até chegar a algo conhecido, transformando o problema inicial em subproblemas[3] que, eventualmente, possuam uma solução conhecida (o que constitui o ponto de parada da análise); (III) realizar a *síntese*, na qual os problemas devem ser enumerados a partir dos mais fáceis até os mais difíceis, para que seja possível retraçar o caminho de volta até o problema inicial; (IV) *recompor* os subproblemas de maneira ordenada para que nada fique em dúvida e, portanto, o problema maior seja solucionado.

A partir deste mesmo método, logo no início das *Meditações Metafísicas,* Descartes propõe que duvidemos de nossos sentidos, pois "é prudência nunca se fiar inteiramente em quem já nos enganou uma vez", o que coloca em questão até mesmo a existência do corpo, enquanto "objeto" percebido pelos sentidos. A primeira certeza só é conquistada quando o filósofo reconhece que o próprio "duvidar" exige que ele exista enquanto "algo" (uma substância) que seja capaz de duvidar, pois, para duvidar, é preciso existir um ser pensante em algum sentido. Esse ser capaz de duvidar é, para Descartes (1973b, p. 130), uma *substância pensante* "que concebe, que afirma, que nega, que quer, que não quer, que imagina também e que sente", e que possui um estatuto ontológico distinto do corpo. Tal conclusão constitui o *cogito* cartesiano ("*penso, logo existo*"), sendo este o ponto de parada da análise que dará início à síntese no processo de fundamentação do conhecimento. No entendimento de Ferry (2007, p. 159), é o *cogito* que inaugura a época do humanismo moderno, na qual reina o que é denominado por *sujeito*.

O método cartesiano no estudo do *cogito* forneceria um critério de verdade com base na consciência. Segundo Ferry (2007, p. 160), é o estado da nossa consciência "[...] que vai se tornar o novo critério de verdade. Isso já mostra o quanto a *subjetividade* se torna importante para os Modernos". Segundo a concepção cartesiana, a subjetividade pertenceria ao ser humano, dado que só ele possuiria, além do corpo, alma. É justamente

3 A escolha de tais subproblemas seria assegurada, segundo Descartes, pelo bom senso partilhado por todos os homens, visto que são dotados de alma – sendo este, na metafísica cartesiana, o critério de relevância na escolha de suas ações.

a alma que tornaria possível a superioridade do ser humano, pois é ela que possibilita o pensamento. Supondo que a natureza e os outros animais não têm alma, eles não teriam a capacidade de pensar (ou sentir), o que permitiria sua utilização sem consequências morais (tendo em vista sua natureza mecânica, eles poderiam ser facilmente substituídos).

Entendemos que com a "virada naturalista", entretanto, o método cartesiano de análise e síntese começa a ser utilizado, mesmo sem que Descartes o tenha desejado, para um estudo mecânico do pensamento (equivalente, portanto, ao próprio corpo que, para o filósofo, seria mecânico por definição).

2. A "virada naturalista na Filosofia"

O termo naturalismo, segundo Papineau (2020), não possui um significado preciso. Num sentido geral, esse termo é utilizado para fazer referência a uma vertente da Filosofia que se situa mais próxima da Ciência. A abordagem naturalista na Filosofia descarta o sobrenatural na explicação da natureza e da mente, pois, em geral, concebe a realidade constituída apenas por elementos e leis *naturais*[4], as quais são explicadas através de métodos científicos.

No mesmo viés de Papineau (2020), Abrantes (2004) afirma que não há somente um tipo de naturalismo, mas que há diversos tipos, que se configuram de acordo com o conjunto de teses adotadas. Abrantes (2004, p. 5) ressalta que dentre tais teses destacam-se: (i) a defesa do Fisicalismo no estudo da mente; (ii) a rejeição do fundacionalismo; (iii) a recusa de justificação *a priori*; e (iv) o monismo metodológico.

A tese (i) consiste na concepção de que todas as coisas existentes são físicas: tais coisas expressam propriedades físicas ou estão relacionadas com sua natureza física. A tese (ii) é uma refutação à grande parte das teorias modernas, que são fundamentadas em bases transcendentais. Um exemplo de teoria que compõe o fundacionalismo é a proposta por Descartes, que está fundamentada numa metafísica do *cogito* pressupondo a existência de um Deus. A tese (iii), por sua vez, diz respeito à rejeição de justificação *a priori* para

4 Conforme Moraes (2014), utiliza-se o termo "natural" (em vez de físico) para não reduzir o naturalismo ao Fisicalismo (que seria apenas umas das vertentes do naturalismo ontológico). Além do físico, algumas vertentes do naturalismo também adotam uma perspectiva informacional, na qual a informação é o elemento fundamental para a explicação da mente. Neste sentido, o termo natural englobaria outros termos como "físico", "biológico" ou "informacional" – que expressariam uma rejeição a pressupostos transcendentais na fundamentação do conhecimento *a priori*.

crenças e contestação do pretenso status *a priori* da epistemologia. Segundo a vertente naturalista, quando lidamos com a natureza do conhecimento e da crença é necessária uma justificação *a posteriori*, para que possamos distinguir o conhecimento da mera opinião verdadeira. Por fim, a tese (iv) decorre do Fisicalismo: ela sustenta que, uma vez que os elementos existentes no mundo são constituídos por elementos físicos, não é preciso utilizar diferentes métodos para explicar os processos e eventos (presentes no mundo físico), mas apenas o método científico empregado na Física. Grosso modo, as quatro teses apresentadas têm em comum a rejeição ao transcendente como elemento explicativo, metodológico, ou como pressuposto no qual uma teoria da mente poderia se fundamentar.

Ainda de acordo com Abrantes (2004), podemos destacar três posturas naturalistas: naturalismo ontológico, metodológico e epistemológico. O *naturalismo ontológico* pressupõe uma concepção realista de mundo: o que é real e inteiramente existente é aquilo que é natural, sem recurso ao transcendente. O *naturalismo metodológico* busca unificar métodos de explicação da natureza da mente e do comportamento em sua análise filosófica; considera que se a Filosofia tem a pretensão de dizer algo relevante sobre o mundo terá de fazê-lo a partir de métodos e dados das ciências naturais (por exemplo, oriundas da Física ou Biologia). O objetivo direcionador das pesquisas filosóficas seria uma aproximação da ciência em sua prática, com seus métodos e resultados. Já o *naturalismo epistemológico* assume a tese epistêmica de uma epistemologia evolucionária, herdeira da tradição darwinista[5].

Entendemos que as origens da alteração do paradigma moderno para o naturalista, responsáveis, por exemplo, pela formulação das quatro teses indicadas, podem ser ilustradas principalmente nas propostas de La Mettrie (1748) e de Dewey (1909). La Mettrie, filósofo e médico, desenvolve sua concepção sobre o funcionamento da mente a partir da rejeição ao dualismo cartesiano e da concepção de que a alma é apenas mais uma parte material do corpo. Dewey, por sua vez, desenvolve uma análise da influência do darwinismo na Filosofia, destacando o nascimento de uma nova lógica de investigação da vida e do conhecimento, a qual possibilita a presença do acaso enquanto recurso explicativo para a evolução dos organismos. A concepção materialista de ambos inicia a desconstrução da metafísica da subjetividade, uma vez que a concepção de ser humano é reconstruída e

5 Uma tese de Darwin (1859) relevante para nossa discussão consiste na concepção de que a evolução dos organismos ocorreria por sua relação com o meio, de modo que o meio e suas variações atuariam como seletor natural dos organismos que sobreviveriam. Assumir tal tese poderia acarretar que, para compreendermos os estados mentais, também seja preciso conceber sua ordem evolutiva.

este passa a ser entendido apenas como mais um animal dentre outros, que se diferencia apenas por seu grau evolutivo, sem qualquer apelo a uma entidade transcendental que justifique sua existência.

La Mettrie, em 1748, publica seu livro *L'Homme Machine* ("*O homem m*áquina"), no qual podemos identificar características dos pressupostos que subjazem o naturalismo na Filosofia da Mente contemporânea. Entendemos que, com essa obra, este filósofo inicia uma "virada naturalista na Filosofia", pois propõe que analisemos a natureza do homem a partir de uma perspectiva estritamente materialista, empírica e não-transcendental. Para La Mettrie, esse estudo deve se pautar na experiência e observação, cabendo tal trabalho aos filósofos e cientistas. É preciso abandonar as teorias que se pautam em causas secundárias (que constituiria o que Abrantes (2004) denomina por "fundacionalismo") e assumir uma metodologia materialista, empregando apenas leis e pressupostos físicos para explicar a natureza do ser humano e dos eventos presentes no mundo.

O ser humano, no entendimento de La Mettrie (1748), é uma máquina complexa de difícil compreensão, e seu corpo "é uma máquina que se encerra em si mesma", ou seja, não há qualquer elemento constituinte do corpo-máquina que extrapola o âmbito material. Neste contexto, a alma seria mais uma parte material do corpo, cuja explicação não seria um problema, pois, no entendimento do filósofo, esta é concebida para fazer referência à parte material do ser humano que pensa e é seu princípio de movimento. O corpo, por sua vez, seria como um relógio, que requer um bom estado de todas as peças para seu funcionamento correto.

Outro aspecto que podemos identificar na análise de La Mettrie acerca da natureza humana é uma postura que apresenta elementos da teoria evolucionária, que seria desenvolvida por Darwin um século depois. Para La Mettrie, a alma humana possui suas faculdades dada a complexificação resultante de um processo evolutivo. Nesse sentido, a relação da alma com o corpo não pressupõe um dualismo substancial, mas poderia ser reduzida a elementos físicos e biológicos, sem a necessidade de recorrer a entidades transcendentais.

Para explicar a natureza da fusão entre os estados da alma e do corpo, La Mettrie se apoia em dados fisiológicos. Tal relação, segundo o filósofo, adquire consistência a partir do grau de evolução do organismo: quanto mais complexo for o cérebro, mais estreita e firme será a relação entre alma e corpo e mais "racional" será o organismo. Tal complexidade tem como elemento central a *organização* do cérebro, sendo ela a primeira qualidade humana. Nessa perspectiva, La Mettrie considera que a "máquina ser humano" poderia

ser explicada em sua totalidade, uma vez que as habilidades da alma são referentes ao grau de organização específico do cérebro e que a totalidade do corpo não é mais do que o produto emergente da organização dos seus órgãos.

Aspectos naturalistas da teoria de La Mettrie podem ser destacados também em sua concepção acerca da linguagem. Em relação à linguagem, o filósofo ([1748] 2010, p. 9) levanta a seguinte questão: "o que *era* o homem antes de ter inventado palavras e aprendido linguagens?". La Mettrie considera que, antes da invenção da linguagem, o ser humano era apenas um animal complexo dentre outros. Foi através da troca de signos e graças à organização física de seu cérebro que ele teria inventado as palavras e aprendido uma linguagem. Nesse sentido, a linguagem pode ser entendida como um dos resultados do processo evolutivo do cérebro humano.

Em síntese, La Mettrie assume uma postura evolucionária para desenvolver sua investigação acerca da natureza do homem. Tal postura pode ser identificada por sua concepção de que seres existentes dependem de seus graus de organização e complexificação. Dessa maneira, o ser humano é entendido como um grande e complexo relógio, sendo que seus movimentos e parte racional são atribuídos à alma, que é apenas mais uma dentre as partes materiais do corpo, localizada no cérebro. Esse filósofo considera que não é preciso distinguir entre alma e corpo como constituindo duas substâncias distintas – a alma reflete aquilo que é possível segundo a organização do corpo. Ademais, não haveria nada contraditório sobre: "(1) ser uma máquina e (2) ser capaz de sentir, pensar e dizer o que é correto a partir do errado" (LA METTRIE, 1748). Esta passagem destaca a concepção de homem-máquina de La Mettrie enquanto uma organização mais complexa em relação aos outros animais.

Contudo, qual a relação presente entre uma postura evolucionária, que já podemos identificar em algumas partes da proposta de La Mettrie, e o naturalismo presente na Filosofia da Mente contemporânea? Entendemos que, em 1859, com a publicação de *A Origem das Espécies*, Darwin teria fortalecido (indiretamente), na Filosofia, a herança deixada por La Mettrie.

A *Origem das Espécies* propiciou uma alteração do paradigma "do fixo e do transcendente" na Filosofia dando lugar a uma concepção que aceita a atuação do acaso nas mudanças apresentadas pelos organismos. Conforme indicamos, o naturalismo contemporâneo envolve a teoria evolucionária, denominado, por Abrantes (2004), de *naturalismo epistemológico*. Comentaremos as ideias darwinistas a partir do entendimento de Dewey (1910) em seu artigo *The influence of Darwinism in Philosophy*.

Segundo Dewey (1910), a publicação da obra de Darwin teve grande influência no pensamento filosófico e científico vigente, em especial, no que diz respeito à rejeição do pressuposto do recurso à transcendência na compreensão da vida. O filósofo destaca o surgimento de uma *nova lógica* de investigação na Filosofia e na Ciência, a qual rejeitaria uma visão imutável e fixa dos elementos – visão esta que implica a existência de instâncias metafísicas e transcendentais em suas explicações. Esta nova lógica daria lugar a uma concepção dinâmica de aquisição do conhecimento que admite a ocorrência da mudança em decorrência do acaso e das condições locais, contextuais.

Dewey (1910) destaca que em toda a história os seres humanos se impressionaram com os fatos da vida. No entendimento da tradição filosófica (La Mettrie é uma das poucas exceções a esta tradição), a alteração das coisas vivas (em forma, tamanho ou qualidade) eram atribuídas a um *telos*, um fim perfeito, cuja função é manter a ordem em tais mudanças. Quando um grupo de coisas vivas visavam o mesmo *telos*, ele recebia o nome de *espécie*. De acordo com Dewey (1910), esses termos foram aprofundados e ganharam força pela capacidade de serem aplicados a tudo que apresenta uma ordem em seu fluxo, mantendo um padrão durante a mudança. A natureza passa a ser entendida como uma realização progressiva de propósitos e metas visando um fim perfeito, ideal.

No contexto da tradição filosófica, o conceito de *espécie* envolve os conceitos de causa final e de forma fixa; o que favorece a suposição de uma entidade transcendental para a qual as coisas se direcionavam. A mudança era entendida como um lapso nesse percurso. O conhecimento da natureza e de seus eventos e processos eram obtidos a partir de uma inteligência puramente contemplativa. Para Dewey (1910), contudo, a experiência do mundo extrapola a simples contemplação, uma vez que a experiência humana está em fluxo. Entretanto, a investigação vigente na época (metade do século XIX) era impelida a buscar uma realidade em formas e entidades que, supostamente, transcendem os modos naturais de percepção. Tal necessidade constitui o que é conhecido por *problema da transcendência no conhecimento*, que é objeto de crítica de Dewey.

Como superar o problema da transcendência no conhecimento que, segundo Dewey (1910), não fornece explicações que satisfaçam nossa percepção? O filósofo destaca que há dois caminhos a seguir: "[1] Temos que procurar os objetos e os órgãos apropriados do conhecimento nas interações mútuas de alteração das coisas, ou então, [2] para escapar à infecção da mudança, devemos procurá-los em alguma região transcendente e sobrenatural".

Dewey considera que, se optarmos pelo caminho (2), continuaremos com dificuldade para compreender a natureza e seus eventos, pois, ainda, haverá a necessidade de buscar instâncias e entidades transcendentais para atuar como causa das alterações e mudanças. O problema de seguir tal caminho é: como explicar, filosófica ou cientificamente, a atuação de tais entidades se elas extrapolam os limites de nossa percepção? Ele ressalta que é possível identificar tal problema nos esforços escolásticos de interpretar a natureza e a mente em termos de formas e faculdades ocultas.

A escolha da opção (1), segundo Dewey, abre um caminho para uma nova lógica de investigação, uma *lógica da mudança*, que retira a necessidade de referência ao transcendental e torna possível a investigação acerca da natureza e da vida no âmbito filosófico e científico. Nas palavras de Dewey (1910): "a influência de Darwin sobre a filosofia reside em sua conquista do fenômeno da vida segundo seu princípio de transição e que, consequentemente, liberou a nova lógica para ser aplicada à mente, à moral e à vida."

Em sua proposta de uma nova lógica de investigação, Dewey (1990) ressalta que ainda reside um problema de longa ocorrência histórica: o debate entre "*design*" *versus* "acaso". Tal debate possui a seguinte questão subjacente: a origem das coisas já estaria pré-determinada visando um fim ideal ou o que ocorre seria a atuação do acaso em seu desenvolvimento? Em outras palavras, há um criador que é ideal, perfeito, que transcende a natureza ou o que existe são processos guiados pelo acaso? De acordo com o autor, a proposta darwinista não poderia estar associada ao *design,* pois Darwin não aceitaria a concepção de uma natureza que resulta da busca por uma necessidade ideal, mas também não concordaria que o puro acaso pudesse ser sua causa.

No entendimento de Gonzalez e Broens (2011), a proposta de Darwin fortalece a concepção de que há processos guiados pelo acaso atuando na evolução dos organismos. A partir de tal entendimento, o objeto do conhecimento passa a ser investigado numa perspectiva relacional. As filósofas (GONZALEZ & BROENS, 2011, p. 182) enfatizam que "a nova lógica focaliza a interação entre os seres de uma mesma espécie e as variáveis externas de diferentes ecossistemas; interação essa que envolve o acaso e se desenvolve em uma rede que molda e é simultaneamente moldada por variações algumas vezes imprevisíveis". Desta forma, temos que diferentes redes são geradas pelos distintos polos relacionais que se estabelecem (ou não) no complexo sistema da vida. Portanto, o organismo e a natureza passam a ser concebidos num âmbito relacional, o que possibilita a influência do meio no organismo dando origem a processos evolutivos.

A nova lógica de investigação tem seu impacto na Filosofia, conforme ressalta Dewey (1990), ao alterar a relevância de certas questões. Uma vez entendido que a natureza não busca um fim ideal (transcendente) e que a concepção do *design* é abandonada, questões como: "quem criou o mundo?" e "para que o mundo foi feito?" perdem sua relevância. Ao admitir a lógica da mudança, outras questões, que anteriormente eram ignoradas, ganham destaque, como, por exemplo: "que tipo de mundo é esse?". Esta última questão ilustra a alteração de paradigma ocasionada pela nova lógica: em vez de procurar o que está por trás da geração do mundo, procura-se, agora, entender *quais* e *o qu*ê são os processos responsáveis por suas mudanças.

A admissão de uma nova lógica no processo de investigação filosófica não é simples. De acordo com Dewey (1910), essa nova lógica introduz a responsabilidade na vida intelectual. Nesse sentido, a responsabilidade da Filosofia cresce à medida que ela se encarrega de "tornar-se um método de localizar e interpretar os mais sérios dos conflitos que ocorrem na vida" (DEWEY, 1910).

Enfim, Dewey (1990), na condição de um dos arautos da "virada naturalista na Filosofia", procura destacar o nascimento de uma nova lógica de investigação, na Filosofia e na Ciência, gerada a partir da publicação de *A Origem das Espécies*. Tal obra possibilitou o rompimento com a concepção tradicional de que as espécies naturais são *fixas* e *imutáveis* e com a necessidade de apelo ao transcendente para explicar qualquer tipo de alteração em tais espécies. Com a nova lógica, a mudança e a alteração na natureza e nos organismos são explicadas considerando a atuação do acaso em seus processos evolutivos. Segundo o princípio regulador da nova lógica, o progresso intelectual ocorre através da substituição de questões: não as resolvemos, nós as superamos (DEWEY, 1910).

Em resumo, destacamos, até o momento, pressupostos que caracterizam o que entendemos constituir as bases da "virada naturalista na Filosofia", nas quais podemos identificar elementos que compõem o naturalismo na Filosofia da Mente contemporânea. Compartilhamos com outros estudiosos do naturalismo que a refutação de hipóteses de um dualismo de substância e a busca por soluções ao problema da relação mente/corpo são as principais responsáveis pela ocorrência e desenvolvimento dessa virada. Apoiada no "método de análise e síntese" cartesiano, mas destituído de sua metafísica transcendente, inicia-se uma tentativa de compreensão naturalista do pensamento.

Neste sentido, o estudo da natureza do pensamento parte da questão "o que é pensar" e a subdivide em subproblemas "menores" do tipo: "quais as funções do pensamento?", "onde ele está localizado?", "que neurônios são responsáveis pela função 'x'?", entre outros.

Entendemos que tal empreitada metodológica foi uma das responsáveis pelo desenrolar da desconstrução da metafísica da subjetividade implantada na Ciência Cognitiva.

Julgamos que a aplicação do "método de análise e síntese" no estudo do pensamento tem seu ápice na "virada informacional na Filosofia", principalmente com o movimento da Cibernética e dos estudos de Inteligência Artificial (IA), que gerou a concepção de que sistemas artificiais também poderiam possuir estados mentais[6]. Esta concepção situa a desconstrução do subjetivismo não apenas em relação ao corpo, mas também em relação ao pensamento, colocando em xeque a tese antropocêntrica de que só os seres humanos possuiriam mente.

3. Da "virada naturalista" à "virada informacional" na Filosofia

Como vimos, as principais características que marcaram a "virada naturalista na Filosofia" são: a recusa do apelo a entidades transcendentais; a rejeição da fundamentação do conhecimento em bases *a priori*; e a aproximação entre Filosofia e Ciência (em especial, a Biologia e sua perspectiva evolucionista) – fazendo com que a Filosofia comece a se apoiar em dados científicos para o desenvolvimento de suas teorias. Tais características estão também presentes no cenário que possibilitou a ocorrência da "virada informacional na Filosofia". Além desses elementos, convém ressaltar a abordagem mecanicista da natureza – inspirada no "método de análise e síntese" –, que, no pensamento moderno, era aplicada na Ciência e na Filosofia e que, com o início da "virada naturalista", também é aplicada no estudo do pensamento.

A "virada informacional na Filosofia" deu início a uma corrente de investigação sobre a natureza ontológica e epistemológica da informação na Filosofia e na Ciência Cognitiva, fortalecendo o projeto naturalista da mente (MORAES, 2014). Na perspectiva naturalista, a informação, entendida como um elemento objetivo existente na natureza, é

6 Tal tese refere-se àquilo que, historicamente, foi chamado de "IA forte", qual seja, a ideia de criar sistemas de inteligência artificial capazes de realizar toda tarefa cognitiva que um ser humano pode fazer, incluindo tarefas que envolvem compreensão, criatividade e autoconsciência. Atualmente, costuma-se distinguir a IA entre completa (geral) e estreita (específica): a IA completa seria capaz de executar qualquer tarefa intelectual que um ser humano pode realizar, abrangendo diversas áreas do conhecimento; já a IA estreita refere-se a sistemas projetados para tarefas específicas e limitadas, não possuindo a capacidade de executar tarefas fora desse escopo restrito. A maioria dos sistemas de IA atualmente são exemplos de IA estreita, enquanto a busca por uma IA completa continua sendo um desafio em aberto na área da inteligência artificial (Russell; Norvig, 2021).

admitida como ingrediente fundamental para a análise de problemas filosóficos (como, por exemplo, a natureza da mente, a natureza do comportamento intencional, a natureza do significado, entre outros).

Segundo Adams (2003), as propostas de Turing (1950) – e sua influência na origem da IA – e de Wiener (1948; [1954] 1965) – com a criação da Cibernética – fornecem subsídios que permitem considerar a "virada informacional na Filosofia" como a consolidação do processo de desconstrução da metafísica da subjetividade, a qual entendemos já estava patente na "virada naturalista". A aplicação do "método de análise e síntese" no estudo do pensamento vem inaugurar, com o auxílio da tecnologia, um novo entendimento sobre o que significa explicar a natureza dos estados e processos mentais através da produção de modelos mecânicos. Com a "virada informacional", o entendimento de que a mente é um processador de informações corrobora a aplicação do "método de análise e síntese". Este entendimento fortalece o processo de desconstrução da metafísica da subjetividade, uma vez que abre a possibilidade teórica de sistemas artificiais poderem apresentar estados mentais.

A aplicação do "método de análise e síntese", entretanto, uma vez desvinculada de sua metafísica, retira também o elemento que exercia a função de "bom senso" na escolha da subdivisão do problema a ser analisado; ou seja, o elemento responsável pelo critério de relevância na escolha das etapas é perdido. Isso porque, no entendimento de Descartes, se retiramos a alma da constituição do ser humano, o que temos é apenas o corpo mecânico que não possui a capacidade de tomar decisões, que não possui critérios apropriados para decidir os caminhos a tomar. Surge, então, um problema para a Ciência Cognitiva que adota o "método de análise e síntese" na construção de modelos, pois não haveria um critério de relevância que atribuísse certa plausibilidade às possibilidades de escolha. É neste sentido que se destacam as críticas aos modelos mecânicos que resolveriam problemas pautados no critério de relevância do programador.

Dentre as críticas aos modelos mecânicos destituídos de um critério de relevância, destaca-se a seguinte: dado que a Ciência Cognitiva possui como ferramenta metodológica e explicativa a construção de modelos mecânicos, e tendo em vista que um dos principais aspectos do pensamento é a capacidade de escolhas, a explicação deste aspecto, neste âmbito de investigação, se daria via construção de modelos que possuíssem a capacidade de escolha. Entretanto, críticos como Dascal (1990) argumentam que o desempenho dos modelos na execução de escolhas pressupõe o que já está programado pelo engenheiro. Dessa maneira, uma vez que os critérios de relevância de escolha remetem ao critério

do programador, tal explicação seria deficiente. Discussões acerca desta deficiência são desenvolvidas no âmbito da Teoria da Auto-Organização (GONZALEZ, 2004, HASELAGER; GONZALEZ, 2009) e do desenvolvimento de algoritmos genéticos (FALKENAUER, 1997; FOGEL, 2000).

Não é nosso interesse aprofundar a discussão acerca da problemática da aplicação do método cartesiano sem um critério de relevância, mas apenas destacar uma das dificuldades que podem surgir na sua aplicação. Seguimos, portanto, com a apresentação da "virada informacional na Filosofia", entendida como decorrente da "virada naturalista", a qual constitui, para nós, a consolidação do processo de desconstrução da metafísica da subjetividade.

A partir da publicação dos artigos seminais de Turing (1937; 1950), os quais estabeleceram as bases da teoria da computabilidade e do próprio conceito de IA, a mente passa a ser entendida, na Ciência Cognitiva, como um *sistema mecânico de processamento de informação,* cuja estrutura seria dada por um conjunto de algoritmos que, mecanicamente, operam sobre símbolos. Além disso, a tese de Turing (que afirma que qualquer problema matemático que pode ser resolvido por um processo algorítmico pode ser resolvido por uma máquina de Turing[7]) se destacou nos estudos cognitivos por ter propiciado a construção de modelos da atividade mental subjacentes ao pensamento inteligente, o que, supostamente, possibilitaria explicar (e conhecer) a natureza deste tipo de pensamento. Destaca-se, aqui, o pressuposto norteador da Ciência Cognitiva, segundo o qual *"conhecer é fazer"*, através da construção de modelos. Um modelo, lembra Dupuy (1996, p. 23), "se trata de uma idealidade, no mais das vezes formalizada e matematizada, que sintetiza um sistema de relações", sendo, portanto, "como uma forma abstrata que vem encarnar-se ou realizar-se nos fenômenos".

Na "virada informacional na Filosofia", o modelo adquire o status de ferramenta explicativa da natureza e da dinâmica organizadora do pensamento. O modelo mecânico do pensamento explicaria, quando bem-sucedido, os processos que caracterizam o pensar. Este pressuposto embasa o método conhecido como "método sintético de análise", o qual apresenta as seguintes etapas (GONZALEZ, 2005): 1) Enuncie, com clareza, o problema a ser analisado; 2) Divida-o em subproblemas, se necessário for; 3) Identifique as funções, bem como as regras de operação, que possibilitam a solução desses subproble-

7 Uma máquina de Turing é um modelo teórico de computação capaz de simular qualquer algoritmo computacional através de uma fita de leitura e escrita e uma cabeça de leitura que pode se mover e alterar símbolos na fita.

mas; 4) Integre as funções das partes menores, identificando uma função mais abrangente que as reúna.

Seguindo tais passos, busca-se a resolução de um problema complexo através de algoritmos que possibilitam a construção de modelos que realizem tarefas, tais como: problemas matemáticos, problemas de jogos, estruturar um diagnóstico médico, entre outros. Conforme ressalta Gonzalez (2005, p. 567), os adeptos do "método sintético de análise" entendem que a explicação de um evento é fornecida a partir da construção de modelos que simulam ou reproduzem, por meio de leis mecânicas, as funções desempenhadas pelo evento original. Neste método, o computador é empregado como uma ferramenta fundamental.

O "método sintético de análise", fundamentado no pressuposto de que "*conhecer é modelar*", tornou possível o desenvolvimento de modelos mecânicos da mente que geraram, inicialmente, dois desdobramentos na Ciência Cognitiva. Na IA forte, os modelos, quando bem-sucedidos, além de simular, possuem estados mentais. Na IA fraca, por sua vez, os modelos mecânicos apenas simulariam, enquanto recursos explicativos, os estados mentais, mas não possuiriam tais estados. Em ambos os casos, a construção de modelos substituiria o papel das teorias explicativas[8].

Uma das implicações filosóficas ocasionada pela concepção mecanicista da mente seria que se, por hipótese, for possível explicar a atividade mental por meio de modelos que atuam seguindo operadores lógicos, também seria possível solucionar o problema da relação mente/corpo através desta abordagem. Isto porque a mente, compreendida como a faculdade de processar mecanicamente informações, adquiriria, para alguns, um lugar no mundo físico. Como ressalta Dupuy (1996, p. 27), tal suposição considera que "a mente, entendida como o modelo da faculdade de modelizar reencontrou seu lugar no universo material". Em outras palavras, "há informação (e até sentido) no mundo físico", sendo as faculdades da mente apenas as propriedades de sistemas de processamento de informação. Dessa forma, temos que, de acordo com Dupuy (1996), a Ciência Cognitiva surge para desconstruir a metafísica da subjetividade, de modo a explicar os fenômenos mentais em termos de informação.

8 Atualmente, as diferentes técnicas de IA, como Aprendizado de Máquina, Redes Neurais, Processamento de Linguagem Natural e outros, são fascinantes exemplos de como a compreensão do funcionamento da mente humana pode inspirar a criação de sistemas inteligentes que simulam processos cognitivos como aprendizado, raciocínio e tomada de decisões. Por outro lado, a utilização desses métodos de IA nos ajuda a explorar como a mente processa informações, aprende com experiências e se adapta a novos cenários.

Dando sequência ao processo de desconstrução da metafísica da subjetividade, surge outra vertente da Ciência Cognitiva, a Cibernética, que possui como seu criador Wiener (1948; [1954] 1965). Wiener cria a Cibernética com o intuito de desenvolver uma teoria do controle e da comunicação tanto em animais quanto nas máquinas. Este interesse conduziu Wiener ([1954] 1965, p. 18) ao entendimento de que: "os comandos através dos quais exercemos nosso controle sobre o nosso meio são um tipo de informação que impomos a ele". Para o autor, podemos conceber informação como o conteúdo daquilo que pode ser trocado com o mundo externo para nos ajustarmos a ele. Neste sentido, seria por meio da troca de informação com o meio que ocorreria o processo de controle da ação: informações diferentes geram ações diferentes, sendo que é em função da informação disponível no meio que a máquina ou animal desempenha uma ação.

Na explicação sobre a troca de informação do organismo com o meio, destaca-se o conceito de *feedback*, enquanto base da Cibernética. O *feedback* pode ser entendido como uma propriedade dos sistemas de ajustar os seus futuros comportamentos em função das performances passadas. Os processos de *feedback*, segundo Wiener ([1954] 1965), estão presentes nos sistemas neurais, artificiais e orgânicos, na habilidade de preservar, na memória, os resultados das operações realizadas no passado para uso no futuro. Wiener (1965, p. 121) destaca dois empregos fundamentais da memória: i) ela é necessária para manter os processos sinápticos correntes; e ii) é admitida como parte de arquivos da máquina ou do cérebro que contribuem para as bases de comportamentos futuros. No caso dos organismos, o estudo da memória será feito a partir do estudo do armazenamento de informação no cérebro através da atividade sináptica.

Além dos estudos sobre o papel dos processos de *feedback* no comportamento, entendemos que a principal contribuição de Wiener ([1954] 1965, p. 132) para a "virada informacional" surge a partir de sua controversa definição, segundo a qual: "Informação é informação, não é matéria ou energia. Nenhum materialismo que não admitir isso poderá sobreviver nos dias de hoje" . Conforme Moraes (2014), tal afirmação fortalece o projeto naturalista, uma vez que a informação é aí entendida como uma propriedade constituinte do mundo, ao lado da matéria e da energia. Nesse viés, explicações dos fenômenos mentais via informação ganham força.

É importante ressaltar que a citação acima, aparentemente tautológica, na verdade é uma estratégia para indicar a dificuldade de se explicar o plano ontológico da informação, que não se reduz à matéria nem à energia. Tal concepção indica um pressuposto metafísico de Wiener ([1954] 1965), segundo o qual todo o universo, incluindo os seres

humanos, é composto pela relação entre informação, matéria e energia. Neste contexto, os organismos podem ser compreendidos como padrões atuais de informação, os quais mantêm uma estabilidade na troca matéria-informação-energia.

A proposta de Wiener ([1954] 1965), por estar pautada numa abordagem biológico-informacional, auxilia o projeto de modelagem do pensamento. Entretanto, ela não foi prontamente adotada (na década de 1950), pois o interesse da Ciência Cognitiva da época era o de desenvolver a hipótese segundo a qual *"conhecer é modelar"* através de processos estritamente simbólicos, a partir do "método sintético de análise". Uma vez que a proposta de Wiener ([1954] 1965) possui um viés biológico, de atuação da informação nos sistemas complexos, ela é mais difícil de modelagem no domínio simbólico. Fenômenos biológicos aparentemente não apresentam uma relação estritamente determinista, mas podem envolver variações e novidades que escapam do universo causal determinista, dificultando sua reprodução através de algoritmos. Essa complicação já estava presente nas técnicas disponíveis para construção de modelos da época.

Em síntese, procuramos até aqui trazer subsídios para nossa hipótese segundo a qual a "virada informacional na Filosofia" consolida o processo de desconstrução da metafísica da subjetividade. A aproximação da Filosofia com a Ciência possibilitou também a aproximação de problemas filosóficos tratados a partir de metodologias científicas e computacionais. Desse modo, a tese de Turing (interpretada como a tese de que pensar é processar informação por meio de algoritmos) impulsionou o desenvolvimento de computadores que fossem capazes de reproduzir estes mesmos algoritmos. Diante disso, o pressuposto norteador da Ciência Cognitiva de que "*conhecer é fazer*", através da construção de modelos, ganha corpo e auxilia o aprimoramento de sistemas artificiais que simulam (reproduzem?) aspectos da mente humana, dando origem e fortalecendo os estudos da IA.

Destarte, consolida-se a desconstrução da metafísica da subjetividade, pois o homem perde seu lugar de "único ser com mente", como era concebido pelo pensamento moderno. Vimos que Wiener (1948, [1954] 1965) também contribui para tal desconstrução, uma vez que promove uma análise informacional da mente inspirada na Biologia, visando o desenvolvimento de máquinas que pudessem apresentar características semelhantes àquelas dos organismos (e, consequentemente, funções semelhantes).

4. Considerações finais e repercussões atuais

Procuramos argumentar que a ocorrência da "virada naturalista na Filosofia" tem como eixo norteador a desconstrução da metafísica da subjetividade. Na busca por tal desconstrução, foram geradas diversas teses que atualmente fundamentam o naturalismo na Filosofia da Mente. Entendemos que este cenário propiciou a emergência da "virada informacional na Filosofia". Neste sentido, defendemos a hipótese de que a "virada informacional" pode ser concebida como a consolidação do processo de desconstrução iniciado na "virada naturalista"[9].

Como ressaltamos, a concepção cartesiana referente à natureza do ser humano atribui a ele um status de superioridade em relação à natureza e aos outros animais, pois, supostamente, seria o único ser dotado de alma e, consequentemente, o único ser capaz de pensar e sentir. Entretanto, com ambas as viradas, tal concepção é refutada e substituída pelo entendimento de que o ser humano é um organismo complexo dentre outros, e que, retirado do centro do universo, passa a ser estudado como um sistema processador de informações, de modo a resolver problemas mecanicamente. Uma vez que perde seu status de superioridade *epistêmica*, o estudo acerca da natureza do pensamento humano adquire um viés biológico e/ou informacional, pautado em bases naturalistas.

A aproximação da Filosofia com a Ciência visa, agora, fornecer explicações objetivas dos eventos mentais. a Filosofia proporciona aos teóricos um conhecimento, em princípio, passível de verificação. A perspectiva evolucionária, inerente à "virada naturalista", também auxilia no desenvolvimento de explicações "objetivas", sem o apelo a entidades transcendentais. Assim, o organismo é entendido como um produto da relação que estabelece com o seu meio, que, por sua vez, atua sobre ele. A atuação do acaso na relação meio/organismo é utilizada como recurso explicativo para alterações que o organismo apresente em relação à sua espécie. Esta perspectiva naturalista pode ser, sob certa perspectiva, entendida como um aspecto positivo das viradas "naturalista" e "informacional", uma vez que, em princípio, ela evitaria o recurso a entidades cuja verificação extrapola o alcance de nossa percepção.

9 É importante destacar que no estágio inicial da "virada informacional na Filosofia" os pesquisadores adeptos desta virada focalizavam na desconstrução da metafísica da subjetividade. Contudo, no desenrolar de tal virada, é possível encontrar estudos acerca da subjetividade via informação (a abordagem dos *qualia* proposta por Dretske [1995], por exemplo).

A ocorrência das viradas "naturalista" e "informacional" também propiciaram o surgimento de novas áreas de investigação sobre a natureza do pensamento: a Filosofia da Mente e a Ciência Cognitiva. Surgem, então, diferentes perspectivas para se abordar problemas filosóficos clássicos, sendo que novos problemas estão sendo formulados (e.g.: o que é informação? Sistemas artificiais manipulam informação significativa? A epistemologia pode ser fundamentada em informação?, entre outros).

Dessa forma, por exemplo, o problema da relação mente/corpo, gerado pela concepção cartesiana de um dualismo de substância, motiva o desenvolvimento de inúmeras teorias na Filosofia da Mente, produzindo impactos também na Ciência Cognitiva. Nesta ciência, o problema da relação mente/corpo é, por exemplo, considerado em analogia à relação hardware/software: o corpo corresponderia ao hardware enquanto a mente corresponderia ao software. Destaca-se, ainda, uma grande transformação metodológica: o "método de análise e síntese" cartesiano é reformulado dando lugar ao "método sintético de análise", que assegura etapas similares do método cartesiano, em particular, no processo de análise, mas enfatiza os elementos sintéticos na construção de modelos artificiais (MACKAY, 1963; RUMELHART; MACLELLAND, 1989; GONZALEZ, 1996; SIMON, 1998; BROOKS, 2002; GALLAGHER, 2007).

O "fascínio pela modelagem" é inspirado no célebre e conhecido princípio do *verum-factum* de Vico (1984, p. 31), o qual é constituído pela ideia de que o que é verdadeiro e o que se faz é aquilo de que somos causa, do que fabricamos. Neste sentido, obtém-se o entendimento de que "o fazer" fornece ao ser humano o conhecimento racional (por meio de regras) daquilo que foi feito; o que o conduz a pensar que possui um controle explicativo e preditivo dos objetos de conhecimento.

As explicações pautadas em modelos parecem ressuscitar aquele espírito de controle e dominação da natureza presente, com maior impacto, no pensamento moderno. Isso porque, como ressalta Dupuy (1996, p. 36), os impactos da modelagem na Filosofia fazem surgir questões como: "Trata-se de desvalorizar o homem? De elevar a máquina? Ou, pelo contrário, de fazer do homem um demiurgo capaz de criar um cérebro ou um espírito artificial?". Segundo o próprio autor, "cada uma destas interpretações tem, certamente, sua parte de verdade, maior ou menor conforme os indivíduos e as épocas" (DUPUY, 1996, p. 36).

Com isso, concebe-se a possibilidade de implicações éticas decorrentes do uso da modelagem no estudo da mente; da aplicação do "método sintético de análise" para a compreensão do humano inserido no meio ambiente. O final da citação nos conduz a

refletir: em que época estamos? Segundo Ferry (2007, p. 250), a resposta seria a seguinte: "pela primeira vez na história da vida, uma espécie viva detém os meios de destruir todo o planeta; e essa espécie não sabe para onde vai". Inspirado em Heidegger, Ferry (2007, p. 239) considera que a humanidade "não sabe para onde vai", pois a concepção mecânica de mundo, quando levada ao extremo, extrapola aquela concepção oriunda do pensamento moderno, constituindo uma "sociedade da técnica".

Nesse contexto, o desenvolvimento científico, que até o Modernismo tinha como pressuposto a segurança e a liberdade do ser humano diante das contingências da natureza, adquire outra conotação: o do "mero usar". Os ideais da modernidade são unidos a uma noção de competição (de concorrência), centrando um maior interesse nos meios (no aprimoramento da técnica), os quais gerariam um esvaziamento dos fins: o progresso não tem outro fim senão a si mesmo. No entendimento de Ferry (2007, p. 247), o poder que o ser humano tem sobre o mundo aumentou, mas "[...] é simplesmente o resultado inevitável da competição. Nesse ponto [...] a técnica é realmente um processo sem propósito, desprovido de qualquer espécie de objetivo definido". Em outras palavras, ocorre um esvaziamento do propósito no sentido da perda de foco, que não está mais na dominação da natureza visando o "bem" do ser humano, mas no desenvolvimento da técnica em busca de uma suposta superioridade em um meio competitivo.

Uma vez que a ação humana estaria pautada na dominação "sem propósito", o homem, segundo Ferry (2007, p. 249), supostamente, perderia por completo seu lugar no mundo e, também, perderia o controle dos objetivos a serem buscados, propiciando, segundo nosso entendimento, uma *diluição da metafísica da subjetividade*. A subjetividade se dilui, pois, o sujeito epistemológico não está no centro, nem é mais um dentre os outros existentes, ele se dilui na conquista cega de especializações. Diferentemente do pensamento moderno – segundo o qual o ser humano era entendido como o único ser digno de valor, sendo que a suposta dominação da natureza propiciaria uma emancipação do obscurantismo medieval e das servidões naturais – , no "mundo da técnica", aquele "homem moderno" perde seu papel de delimitador dos fins, pois o que é gerado é uma "dominação sem propósito", a busca por um progresso que ultrapassa as vontades individuais visando à competição e o mero aprimoramento da técnica.

As repercussões da "virada informacional na Filosofia" envolvem um grande grau de complexidade. A problematização, vale citar, não se trata de alimentar um sentimento de nostalgia dos tempos em que o ser humano possuía um papel de delimitador dos fins da dominação, nem expressar uma tecnofobia, mas apenas indicar a necessidade, apa-

rentemente urgente, do desenvolvimento de uma Ética Informacional. Estudiosos como Rafael Capurro (2010), Luciano Floridi (2011; 2013), Quilici-Gonzalez, Kobayashi, Gonzalez e Broens (2010; 2014), Moraes (2019), entre outros, propõem um maior enfoque nas implicações que o desenvolvimento de tecnologias computacionais têm tido sobre a sociedade. Mais especificamente, sobre a geração de novos padrões de conduta, de cunho moral, decorrentes da inserção de tais tecnologias na vida cotidiana, conforme abordado por Moraes e Testa (2022).

Outra questão de cunho ético que se coloca como repercussão do processo de desconstrução da metafísica da subjetividade propiciado por ambas as viradas, em especial pela "virada informacional", é relativa aos processos técnicos do funcionamento dos sistemas de IA atuais – cujas decisões e ações não podem ser facilmente compreendidas ou justificadas de forma clara e transparente, tornando-as "não-explicáveis" (RUSSELL; NORVIG, 2021). As razões para tanto incluem, por exemplo: (i) *complexidade do modelo*: alguns modelos de IA, como redes neurais profundas, podem ter uma arquitetura complexa com milhões de parâmetros, tornando difícil acompanhar o raciocínio interno da máquina; (ii) *caixa-preta*: em muitos casos, os algoritmos de IA são como "caixas pretas", onde as entradas e saídas são conhecidas, mas o processo interno não é transparente; (iii) *aprendizado de máquina não supervisionado*: em alguns métodos, os padrões aprendidos podem ser difíceis de serem interpretados pelos seres humanos; (iv) *dados não explicáveis*: às vezes, os algoritmos de IA podem ser treinados com grandes conjuntos de dados, tornando difícil identificar como certas conclusões são tiradas com base em padrões; e (v) *limitações de linguagem*: os modelos de IA podem processar informações de maneira muito diferente dos humanos, o que dificulta a explicação usando linguagem humana comum.

A falta de explicabilidade é uma questão cara à Ética Informacional, especialmente em contextos críticos, como cuidados médicos, nos quais a capacidade de entender o raciocínio subjacente de uma IA é crucial para garantir sua confiabilidade e segurança. Além disso, a falta de explicabilidade de tais modelos trazem novos elementos à hipótese segundo a qual "conhecer é modelar" e, principalmente, ao "método sintético de análise", porquanto tais modelos computacionais não mais podem ser entendidos como guiados por processos lógico-dedutivos.

Pensemos, ainda, sobre se o status explicativo de tais modelos é relativa à lógica subjacente dos mesmos – a despeito de representados por uma linguagem lógico-dedutiva, uma questão que raramente é discutida é sobre o status da lógica subjacente. Seria possível que distintas lógicas aduzam a diferentes definições de racionalidade? Em caso

afirmativo, qual o impacto em relação ao próprio conceito de mente?[10] O fato de que tais modelos fazem cada vez mais parte da vida cotidiana dos indivíduos corrobora a urgência de uma Ética Informacional.

Referências

ABRANTES, P. Naturalismo em filosofia da mente. In: FERREIRA, A.; GONZALEZ, M.E.Q; COELHO, J. G. (Orgs.). **Encontro com as ciências cognitivas.** Marília: UNESP, v. 4, p. 5-37, 2004.

ADAMS, F. The informational turn in philosophy. **Minds and Macnhines**. Netherlands: Kluer Academic Publishers, v. 13, p. 471-501, 2003.

CAPURRO, R. Desafíos téoricos y practicos de la ética intercultural de la información. In: **E-Book do I Simpósio Brasileiro de Ética da Informação**. João Pessoa: Idea, p. 11-51, 2010.

CARNIELLI, W.; TESTA, R. Paraconsistent Logics for Knowledge Representation and Reasoning: advances and perspectives. **18th International Workshop on Nonmonotonic Reasoning**. 2020.

DASCAL, M. Artificial intelligence as epistemology? In: VILLANUEVA (Ed.). **Information**, semantics and epistemology. Oxford: Blackwell, 1990, p. 224-41.

DARWIN, C. **On the origin of species** – Or the preservation of favoured races in the struggle for life. The Project Gutemberg. Disponível em: http://www.gutenberg.org/files/1228/1228-h/1228-h.htm . Acesso em 3 ago. 2023 [1859].

DESCARTES, R. Discurso do método. In: DESCARTES, R. **Obras escolhidas**. Rio de Janeiro: Bertrand Brasil, 1973a.

DESCARTES, R. Meditações. In: DESCARTES, R. **Obras escolhidas**. Rio de Janeiro: Bertrand Brasil, 1973b.

10 A sugestão de uma resposta afirmativa à primeira questão é levantada por Testa (2014), que avança em um modelo paraconsistente de agente epistêmico (TESTA; CONIGLIO E RIBEIRO, 2017). Porém, seu desdobramento sob a ótica da Filosofia da Informação ainda não foi explorado, apesar de seu potencial poder informacional ter sido brevemente colocado por Carnielli e Testa (2020).

DEWEY, J. **The influence of darwinism in philosophy.** Disponível em: https://www.gutenberg.org/cache/epub/51525/pg51525-images.html. Acesso em: 03 ago. 2023 [1910].

DEWEY, J. **Reconstrução em filosofia**. Tradução: Antônio Pinto de Carvalho. São Paulo: Cia. Ed. Nacional, 1959.

DRETSKE, F. I. **Naturalizing the mind**. Cambridge: MIT Press, 1995.

DUPUY, J. P. **Nas origens das ciências cognitivas**. Unesp: São Paulo, 1996.

FALKENAUER, E. **Genetic algorithims and grouping problems**. Chichester: John Wiley & Sons Ltd., 1997.

FERRY, L. **Aprender a viver**: filosofia para os novos tempos. Objetiva: Rio de Janeiro, 2007.

FOGEL, D. B. **Evolutionary computation***:* towards a new philosophy of machine intelligence. New York: IEEE Press, 2000.

FLORIDI, L. **The Philosophy of Information**. Oxford: Oxford University Press, 2011.

FLORIDI, L. **The Ethics of Information**. Oxford: Oxford University Press, 2013.

GONZALEZ, M. E. Q. **Informação e conhecimento comum**: uma análise sistêmica dos processos criativos auto-organizados. 2004 (Tese de Livre-Docência).

GONZALEZ, M. E. Q. Information and mechanical models of intelligence: What can we learn from Cognitive Science? **Pragmatics & Cognition**, Amsterdam: Ed. John Benjamin Publishing Company, v. 13, n. 3, p. 565-82, 2005.

GONZALEZ, M. E. Q.; BROENS, M. C. Darwin e a virada naturalista na Filosofia. In: João Quartim de Moraes (Org.). **Materialismo e evolucionismo II**: a origem do homem. Campinas: UNICAMP, 2011, v. 59, p. 175-91 (Coleção CLE).

HASELAGER, W. F. G.; GONZALEZ, M. E. Q. Auto-organização e autonomia. In: D'OTTAVIANO, I. M. L.; BRESCIANI FILHO, E.; GONZALEZ, M. E. Q. (Orgs.). **Auto--Organização:** estudos interdisciplinares. Campinas: UNICAMP, v. 52, p. 223-36, 2008.

LA METTRIE, J. O. **Man a machine**. Disponível em: < https://www.marxists.org/reference/archive/la-mettrie/1748/man-machine.htm>. Acesso em 3 ago. 2023 [1748].

MORAES, J. A. **Implicações éticas da "virada informacional na filosofia"**. Uberlândia: EDUFU, 2014.

MORAES, J. A. **O paradigma da complexidade e a ética informacional**. Campinas: CLE-UNICAMP, 2019.

MORAES, J. A.; TESTA, R. R. A sociedade contemporânea à luz da ética informacional. **Acta Scientiarum. Human and Social Sciences**, v. 42, n. 3, e56496, 2020. DOI: http://dx.doi.org/10.4025/actascihumansoc.v42i3.56496.

PAPINEAU, D. Naturalism. In: ZALTA, E. N. (Ed.). **The Stanford Encyclopedia of Philosophy** (Summer 2021 Edition), 2020. Disponível em: https://plato.stanford.edu/archives/sum2021/entries/naturalism/. Acesso em: 03 ago. 2023.

QUILICI-GONZALEZ, J. A.; KOBAYASHI, G.; BROENS, M. C.; GONZALEZ, M.E.Q. Ubiquitous computing: any ethical implications? **International Journal of Technoethics**, v. 1, p. 11-23, 2010.

QUILICI-GONZALEZ, J. A.; BROENS, M. C.; GONZALEZ, M.E.Q.; KOBAYASHI, G. Complexity and information technologies: an ethical inquiry into human autonomous action. **Scientiae Studia**, v. 12, p. 171-179, 2014.

RUSSELL, S. J.; NORVIG, P. **Artificial intelligence:** A modern approach. 4th ed. Hoboken: Pearson, 2021.

TESTA, R. R. **Revisão de Crenças Paraconsistente baseada em um operador formal de consistência**. Tese (Doutorado em Filosofia), Universidade Estadual de Campinas, 2014. DOI: https://doi.org/10.47749/T/UNICAMP.2014.935185.

TESTA, R. R. Paraconsistency. In: MATTINGLY, J. (Ed.), **The SAGE encyclopedia of theory in science, technology, engineering, and mathematics**. SAGE Publications, Inc., v. 1, p. 629-32, 2023. DOI: https://dx.doi.org/10.4135/9781071872383.n144.

TESTA, R.R.; CONIGLIO, M.E.; RIBEIRO, M.M. AGM-like paraconsistent belief change. **Logic Journal of the IGPL**, v. 25, n. 4, p. 632-72, 2017. DOI: https://doi.org/10.1093/jigpal/jzx010.

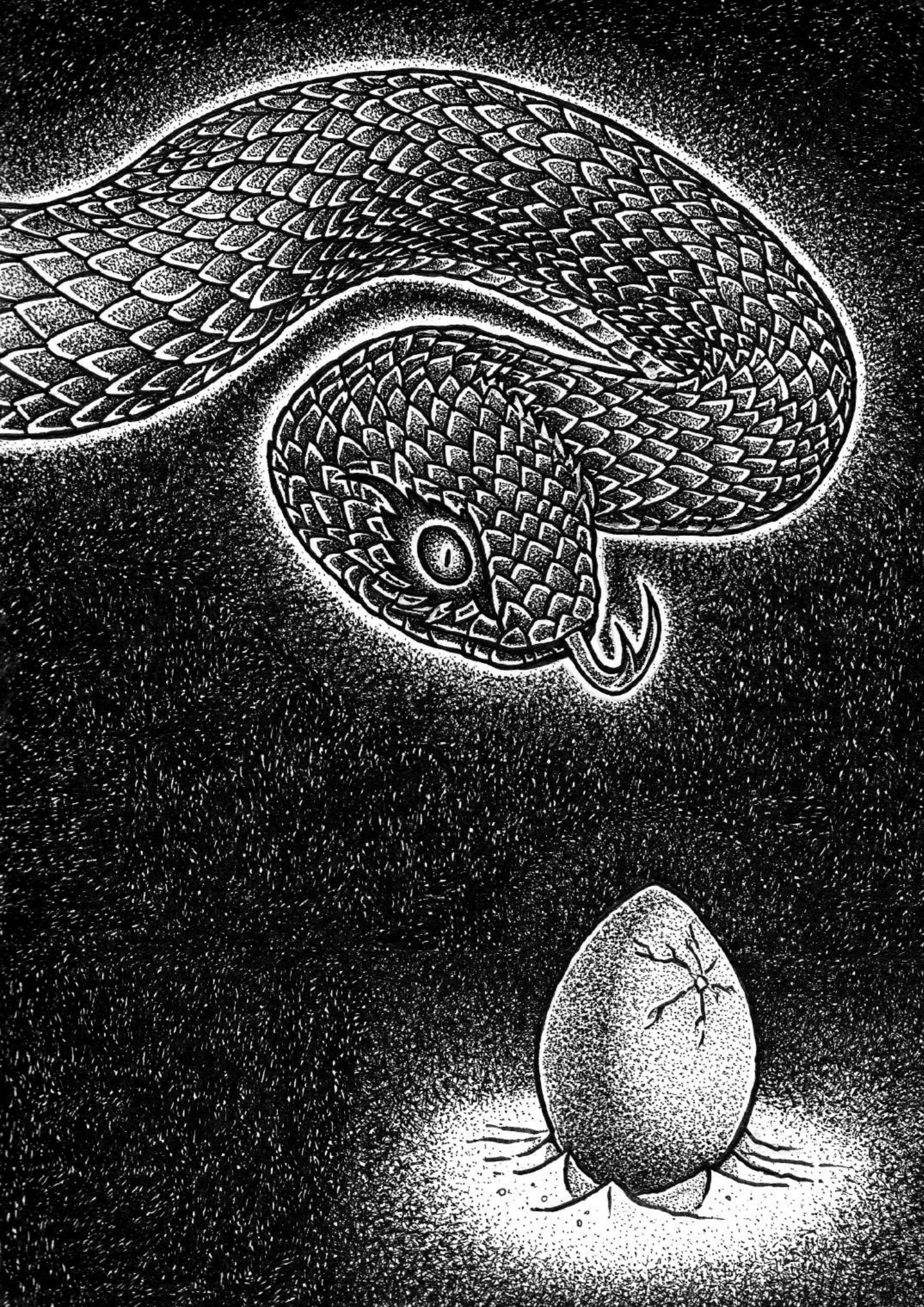
TURING, A. M. On Computable Numbers, with an Application to the Entscheidungsproblem. **Proceedings of the London Mathematical Society**. v. 42, n. 1, p. 230-65, 1937.

TURING, A. M. Computing machinery and intelligence. In: **Mind**, 59, 433-60, 1950.

VICO, G. **Princípio de uma ciência nova:** acerca da natureza comum das nações. São Paulo: Editora Abril, 1984 (Coleção *Os Pensadores*).

WIENER, N. **Cybernetics**. 2ª Ed. Cambridge, MA: MIT Press, 1965 [1948].

WIENER, N. **The human use of human beings:** cybernetics and society. London: Sphere Books LTD, 1968 [1954].



ACERCA DE MENTES NATURAIS E DIGITAIS, OU DE PROMESSAS ARRISCADAS

Luís Moniz Pereira

Departamento de Informática, Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa
2829-516 Monte da Caparica, Portugal
lmp@fct.unl.pt
ORCID 0000-0001-7880-4322

António Barata Lopes

ANQEP – "Agência Nacional para a Qualificação e Ensino Profissional" e
Agrupamento de Escolas de Alvalade
Av. 24 de Julho 138, 1200-771 Lisboa, Portugal
lopesab@msn.com

Resumo:

A Ciência da Computação diz-nos como criar um meta-interpretador para uma linguagem, escrito na própria linguagem, caso do Lisp e Prolog. A metalinguagem sendo igual à linguagem, faz com que ela goze da capacidade reflexiva de falar sobre seus enunciados e procedimentos. Da mesma forma, a Máquina de Turing básica permite fazer emergir, por *bootstrap*, a Máquina de Turing Universal. E portanto, permitir modelar qualquer computação máquinal. Como o Uroboros, a cobra que come a sua própria cauda, uma Máquina de Turing pode emergir para outros níveis de existência. Tais capacidades computacionais são boas demais para serem ignoradas pela

Abstract:

Computer Science tells us how to devise a meta-interpreter for a language, written in the language itself, viz. Lisp, Prolog. The metalanguage being the same as the language, makes it enjoy the reflexive ability to talk about its statements and procedures. Likewise, the basic Turing Machine lets emerge, by bootstrapping, the Universal Turing Machine. And thence permits to model any machine computation. Like the Uroboros, the snake that eats its tail, a Turing Machine can emerge onto other levels of existence. Such computational abilities were too good to be ignored by evolution, therewith fostering the emergence of cognition. Our brains, consciously and adeptly use and can teach

evolução, tendo promovido pois o surgimento da cognição. Os nossos cérebros, consciente e habilmente, usam e podem ensinar tais capacidades: ou seja, autodepuração, explicação e justificação, antecipação e preferência por futuros, visão contra-factual do passado com conhecimento do presente, actualização moral, detecção e remoção de contradições, argumentação, etc. A nossa tese: a computação, complementada por dados da Psicologia Evolucionária e por estudos sobre o comportamento de animais sencientes, é a ferramenta epistémica objectiva, evolutiva e abstracta, por excelência, para modelar a consciência, incluindo a evolução da moralidade nas populações em geral, empregando a Teoria dos Jogos Evolutivos. Esta abordagem epistémica computacional em direcção a uma teoria especulativa objectiva da consciência parece ser o actual paradigma *par excellence*.

Palavras-chave:

Inteligência artificial (IA); Alan Turing; máquina de Turing; Psicologia Evolucionária.

such abilities: namely self-debugging, explanation and justification, anticipating and preferring futures, counterfactually seeing the past with knowledge of the present, moral updating, contradiction detection and removal, argumentation, etc. Our contention: computation, complemented with data from Evolutionary Psychology and studies of sentient animals, is the objective, evolutionary, and abstract epistemic tool par excellence to model consciousness, including the evolution of morality in populations in general, employing Evolutionary Game Theory. The computational epistemic approach towards an objective speculative theory of consciousness appears to be the present paradigm *par excellence*.

Keywords:

Artificial intelligence (AI); Alan Turing; Turing machine; Evolutionary Psychology.

What's past is [*epistemic*] prologue, what to come in yours and my discharge.
William Shakespeare, *The Tempest* (1610-1611)
(Where Prospero bids farewell to black magic and enters a "brave new world")

Introdução

Quando pretendemos abordar conteúdos associados ao termo *mente*, dado que estamos perante um dos conceitos, ou ordem de conceitos, mais difícil de definir, temos sempre de considerar que enfrentamos um campo de investigação recente, e com muita margem para aprofundamento. Podemos, ainda assim, delimitar o seu âmbito a um conjunto agregado de processos de matriz cognitiva, emocional, e conativa, de natureza consciente e inconsciente, orientados para a aquisição, conservação, exploração, e avaliação de informações externas e internas.

Nestes termos, dispor ou não de uma mente não é, de todo, irrelevante. Embora os agentes destituídos dela possam estabelecer relações bem-sucedidas com o meio, o facto é que a capacidade de criar representações, internas e externas, é a condição de possibilidade para interacções mais diversificadas e complexas com o meio externo, e connosco próprios.

A ideia de estarmos perante um conjunto agregado de funções múltiplas não é nova no pensamento ocidental, podendo já ser encontrada no tratado de Aristóteles sobre a Alma[1]. Note-se que a ordem de conceitos com os quais caracterizamos a mente ancora no que tradicionalmente se designou por alma, espírito, ou outros termos próximos de uma concepção antropológica dual, sintetizando imanência e transcendência. Aquele filósofo grego tinha a noção clara de que tal capacidade não era exclusiva dos seres humanos, conjecturando, por isso, a existência de almas qualitativamente diferenciadas, e referindo-se também a uma gradação que, de alguma forma, quando actualizada para o presente, tem condições para ser lida como evolutiva; senão no tempo, pelo menos nas qualidades. É assim que Aristóteles nos fala de almas com faculdade sensitiva; sensitiva e locomotiva; e finalmente, sensitiva, locomotiva e intelectiva.

A temática das faculdades superiores da cognição foi sempre inerente ao discurso filosófico; mas, para construirmos uma noção mais dinâmica dela, foi preciso chegarmos

1 ARISTÓTELES. *Acerca del alma*. Introdução, tradução do grego e notas de Tomás Calvo Martinez. Madrid: Gredos, 1994.

ao século XIX, e a cientistas como Paul Broca, com o qual as faculdades mentais superiores – como o pensamento e a linguagem – começaram a ser vistas, não como aptidões decorrentes de uma misteriosa centelha divina, mas sim como produção de um órgão director que é o cérebro. Um cérebro resultante de pressões evolucionárias desafiantes, e de interacções com meios diversos, a exigirem respostas diferenciadas e complexas.

Ora, foi essa necessidade de diferenciação e sofisticação que veio a resultar na emergência da estrutura orgânica mais complexa, flexível e poderosa que a natureza engendrou. O cérebro encontra-se no topo de um corpo erecto, protegido por uma camada óssea extremamente robusta e rodeada de sensores de vária ordem, como se guardas de uma caixa-forte se tratassem. Por via dos membros superiores e inferiores do corpo onde está alojado, dispõe também de actuadores que, não só permitem interagir com o mundo, como "afinar-se" a si próprio. O seu comparativamente elevado custo energético assinala que sua evolução mereceu a pena.

1. Novas emergências paradigmáticas

Antes dos primeiros estudos científicos sobre o suporte biológico da cognição, tomando como base de investigação lesões ocorridas nos indivíduos e respectivas consequências, como os trabalhos conduzidos por Broca; ou partindo do estudo do comportamento animal, como os estudos conduzidos por Pavlov acerca da possibilidade de se condicionarem respostas comportamentais; ou ainda os primeiros passos da Psicologia Experimental conduzidos por Wundt, a anterior abordagem à mente humana tinha como pressuposto a existência em todos nós de uma dimensão espiritual, a qual era tida como incompatível com a sua abordagem por via das ciências experimentais, ou a construção engenheirística de artefactos sucedâneos. Postulava-se também uma ruptura entre humanos e natureza, segundo a qual os primeiros eram vistos como beneficiários da restante criação divina. A revisão dessa crença implicou que deixámos de nos percepcionar como a pedra angular da criação, para concebermos a vida e cognição de um modo integrado. Nessa integração, a mente humana ocupa, ainda e por enquanto, o topo da hierarquia mental; mas agora, enquanto resultado de um processo em evolução, o qual, atingido este patamar, permitiu que vida e a própria cognição tomassem consciência de si, interpelando não apenas a sua origem e natureza, mas também as suas potencialidades e limites.

Mudado o paradigma, presentemente, os avanços no domínio da Psicologia Evolucionária, e o conhecimento que vamos adquirindo sobre a cognição e comportamento

animal, permitem-nos construir uma visão muito mais coesa e integrada da evolução das espécies e da inteligência/cognição, enquanto processo igualmente evolucionário e distribuído. Além disso, toda a tecnologia aplicada ao universo das neurociências e das ciências cognitivas, como a imagiologia por ressonância magnética, ou engenharia de interfaces cérebros/computadores, entre muitos outros recursos, servem de base a conjecturas mais rigorosas acerca dos suportes biológicos para a inteligência e cognição.

Numa outra vertente, os desenvolvimentos da Inteligência Artificial (IA), associados à capacidade do que poderíamos designar por engenharia filosófica, permitem-nos simular e experimentar no computador processos que anteriormente apenas podiam ser objecto de especulação mediante reflexão e argumentação; embora, por vezes, com base em observações de campo, ou experiências intencionalmente conduzidas. Tais caminhos, devido ao seu potencial de explicitação e análise dos processos mentais humanos e não-humanos, permitem um olhar simultaneamente mais global e detalhado de todos os processos cognitivos. Mais global, porque nos remetem para a mente enquanto faculdade detida por várias espécies, quiçá extraterrestres; e também mais detalhado, pois dispomos de conceitos e instrumentos de investigação que nos permitem "entrar" no domínio da mente em acção, por assim dizer. Mesmo a moral, vista como caso de investigação com recurso à teoria dos jogos evolucionários, é actualmente, objecto de abordagem por via da Psicologia Evolucionária e das Ciências da Computação.

Um dos aspectos que primeiramente devemos abordar diz justamente respeito ao carácter evolucionário dos processos mentais, o qual deve ser considerado em várias vertentes. Desde logo, explorando o facto de partilharmos tais faculdades não apenas com os outros símios, mas também com golfinhos, elefantes e todos aqueles que detêm um sistema nervoso suficientemente complexo para construir representações do meio e linguagens que as expressem (incluso os eventuais extraterrestres); ressalvando, evidentemente, que essas linguagens de outras espécies (pelo menos as terrestres) não têm o potencial das humanas, as quais representam objectos na sua ausência, instauram mundos possíveis, imaginam mundos contra-factuais, diversificam a simbologia, e desdobram-se em meta-linguagens.

É claro que, presentemente, estamos muito longe de Aristóteles e, por exemplo, da consideração segundo a qual apenas os humanos, meta-linguisticamente considerados, riem. Todavia, o riso, também reconhecido noutros símios, é tão só um mecanismo de ligação entre seres com a faculdade de sentir e transmitir afectos. Assim, uma mãe chimpanzé, ou uma mãe humana, ao fazerem cócegas aos seus bebés, obtê-lo-ão como

resposta. Podemos, pois, estabelecer uma conjectura razoável segundo a qual o riso humorístico humano tem uma base ancestral e distribuída por mais outras espécies, para além dos próprios. De igual forma, outras faculdades humanas mais complexas e que exigem uma mente com a capacidade de construir representações abstractas do mundo são também o resultado de competências emergentes, mesmo que tivessem o seu início de modo muito mais rudimentar.

Dois casos estudados por Franz de Waal[2] permitem-nos ilustrar com clareza o que acabámos de afirmar. O primeiro diz respeito às investigações conduzidas com grupos de elefantes asiáticos, na Tailândia, as quais permitem não apenas evidenciar o forte sentido comunitário desta espécie, bem como comportamentos associados a um rudimento de moral, quer em situações de entreajuda entre pares, quer também no modo como os adultos engendram estratégias de protecção das crias face a ameaças.

Foram também investigadas todas as estratégias de comunicação vocal e não vocal (linguagem corporal) que permitem consolidar o sentido de comunidade, a linguagem complexa, as capacidades de avaliar situações de risco individual e colectivo, a construção de estratégias concertadas e as trocas de expressões emocionais após a vivência e superação de situações de risco. Os elefantes em cativeiro, usados como força de trabalho, manifestam inequivocamente comportamentos projectivos, como introduzir forragem num chocalho, a fim de se poderem deslocar silenciosamente, em passeios nocturnos, à revelia dos seus guardadores humanos; e têm os antepassados em consideração por via de um "culto" da morte, que os leva a visitar o local onde pereceram os membros da sua família, como forma de solidificar o intra- e o inter- geracional.

O outro caso diz respeito aos estudos conduzidos por Sarah Brosnan e Frans de Wall, no Brasil, onde os indivíduos testados foram macacos-prego[3]. Essas experiências permitiram validar a capacidade destes símios para avaliação de certas tarefas, e comparação das respectivas recompensas. Foi também possível indagar se o seu comportamento seguia critérios que, minimamente, se pudessem equiparar a um rudimento de moral.

A situação experimental, protagonizada por fêmeas da espécie, consistiu em colocá-las perante a troca de uma simples pedra por uma rodela de pepino. A partir do momento em que um dos sujeitos experimentais começa a ser recompensado com bagos

2 Plotnik JM, de Waal FBM, Reiss D. 2006. Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences of the United States of America* 103:17053-17057

3 https://www.youtube.com/watch?v=JJsKS4DQ95E

de uva, ao passo que o outro continua a receber rodelas de pepino, emerge a percepção da desigualdade. Então, após algumas repetições do mesmo procedimento, o sujeito experimental recompensado com rodelas de pepino opta por recusar o alimento, chegando mesmo ao ponto de o arremessar aos cientistas.

Tal comportamento prova que estes símios são capazes de monitorizar a sua acção e a dos parceiros, de analisar situações relativas, de as percepcionar como injustas porque desiguais, e de construir uma resposta coerente com a sinalização negativa em questão. Não nos devemos esquecer que valorizar é sempre estabelecer uma relação entre itens, a qual é independente dos conteúdos específicos de cada um deles; ou seja, implica sempre um constructo mental, qualquer que ele seja, elaborado pelo sujeito avaliador, no âmbito do qual valoriza uns aspectos em detrimento de outros.

2. A evolução distribuída emergente

Os riscos associados à projecção de características humanas em animais estão suficientemente documentados na literatura científica; no entanto, o pretexto de evitarmos tais riscos não deve, por outro lado, coibir a análise. Ora, as situações experimentais atrás mencionadas evidenciam que a mente é um fenómeno evolucionário, inter-espécies, regida por uma dinâmica de complexificação progressiva. Nos *Homo sapiens*, essa mesma dinâmica permitiu criar representações dos sujeitos em acção, de matriz colectiva e individual, com respectiva avaliação e revisão de estratégias. Essa complexidade está apoiada em linguagens com níveis de abstracção – e inerente artificialidade – cada vez mais diversa.

Desta forma, o jogo evolucionário ocorre quer ao nível dos organismos vivos que vão emergindo, quer ao nível das faculdades que lhes são próprias, incluindo a capacidade de modificar e adaptar-se ao seu eco-nicho. Cabe aqui a introdução do termo "artificialidade", pois, num certo sentido, os suportes dos vários tipos e modos de inteligência podem ser vistos como o hardware, que serve o exercício das funções já referidas; por outro lado, os conteúdos mentais e respectiva organização, que são constituintes de cada mente, podem ser vistos como software.

Nesta linha de pensamento, até dada a reconhecida fragilidade dos nossos suportes biológicos, podemos muito bem conjecturar que um dos maiores desafios enfrentados pelos humanos será proceder à migração de funções cognitivas e, por inerência, mentais, para suportes mais robustos e duráveis relativamente a um corpo biológico sujeito às

agressões de vírus e bactérias, e – por enquanto – com poucos recursos para enfrentar a inevitabilidade da degradação e da morte.

É nesta base que podemos conjecturar a existência de futuras mentes digitais, ou bio-digitais, mercê da hibridização entre organismos biológicos e artefactos de IA. As próteses para sensores, actuadores, e outros órgãos do nosso corpo serão, certamente, complementadas com próteses cognitivas que originarão humanos melhorados, e mesmo simbiontes distribuídos. Não sabemos quais as consequências psicológicas, sociais e económicas de tal caminho. Mas sabemos muito bem que os humanos nunca deixaram de experimentar tudo o que é tecnicamente possível. Acresce que os desafios que colocamos a nós próprios, como sejam a exploração espacial, a manipulação genética, ou a gestão de sistemas e plataformas globais progressivamente mais sofisticadas, multifuncionais e integradas, têm o seu actual desempenho e progresso imputado ao desenvolvimento da IA.

As questões associadas à gestão e integração de dados, com a migração do conceito de aprendizagem para o domínio das Ciências da Computação, autorizam que, por inerência, se tenha iniciado e se desenvolva uma nova literatura em torno da ideia de mente digital. Não que estejamos próximos do conhecimento artificial auto-consciente, mas apenas porque já o podemos conjecturar, e não o eliminar *a prori*. Os humanos sempre especularam sobre a sua relação com inteligências outras; referimo-nos a deuses e a anjos, mas também podemos invocar extraterrestres ou monstros que, sorrateiramente, pudessem coabitar connosco, no mesmo planeta. Mesmo a possibilidade de criarmos autómatos nos quais fosse possível injectar vida esteve sempre presente na especulação humana[4]. Assim sendo, é, e não é, surpreendente que – com quase toda a certeza – a primeira inteligência-outra com a qual interagiremos será uma criação nossa, mas suficientemente empoderada para ser autónoma.

Nestes termos, antecipamos a emergência de um ecossistema cognitivo onde seres biológicos não humanos, humanos melhorados, eventualmente extraterrestres, e máquinas cognitivas formarão um *cluster* destinado à produção e partilha de conhecimento altamente diversificado e distribuído. Sendo que essa diversidade propiciará, certamente, melhores porque mais diversificadas representações tanto do mundo como das próprias modalidades de mente. Sendo também possível imaginarmos a construção de uma mente que aglutine funcionalidades oriundas de domínios diversos.

4 "*AI Narratives – A History of Imaginative Thinking about Intelligent Machines*". Edited by Stephen Cave, Kanta Dihal, Sarah Dillon, Oxford University Press, Oxford, UK 2020.

Presentemente, damos pequenos passos que permitem grandes avanços em várias direcções, nomeadamente no domínio da modelização matemática vectorial que suporta a Linguística Computacional e que, a breve trecho, interferirá drasticamente em todos os domínios das relações humanas. Por outro lado, a alvorada da computação quântica promete-nos saltos qualitativos e quantitativos no tratamento de dados que, por enquanto, apenas conseguimos vislumbrar.

3. Construir para compreender

Richard Feynman tem uma frase famosa: "A prova da compreensão está na construção". Supomos que uma abordagem complementar em direcção a um modelo abstracto da consciência reside na resolução computacional de problemas comuns reconhecidos por exigirem consciência, incluindo processamento reactivo e deliberativo. Partindo então das inovações computacionais que tais problemas e suas combinações nos levem a alcançar, um de nós empregou essa abordagem para modelar a moralidade, tanto para o raciocínio moral individual (começando com os problemas do bonde) quanto para modelar o surgimento e a evolução da moralidade em populações de tais indivíduos, incluindo o sentimento de culpa (Pereira *et al*. 2017; 2023)[5].

Somos a favor de uma postura funcional de Turing em relação à consciência, ou seja, prontamente vista como implementável de acordo com Feynman. No entanto, absorvendo inspiração experimental humana/animal/insecto, consciência distribuída incluída. Com respeito à cognição, enfatizamos todos os 3 níveis indispensáveis detalhados em Pearl (2018)[6], em consonância com o Tri-Processo (Stanovich, 2010)[7]: previsão e reacção por correspondência e aprendizagem; hipótese de cenários com abdução e planeamento; contra-factualizando o passado com o conhecimento presente.

5 T. Cimpeanu, L. M. Pereira, T. A. Han. Co-evolution of social and non-social guilt in structured populations, extended abstract, in 22nd Intl. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (Eds.), *Proceedings ACM Digital Library*, London, UK. May 29 - June 2, 2023.
L. M. Pereira, T. Lenaerts, L. A. Martinez-Vaquero, T. A. Han. Social Manifestation of Guilt Leads to Stable Cooperation in Multi-Agent Systems, in: *Procs. 16th Intl. Conf. on Autonomous Agents and Multiagent Systems* (AAMAS 2017), Das, S. et al. (Eds.), pp. 1422-1430, 8–12 May 2017, São Paulo, Brazil.

6 Judea Pearl, Dana Mackenzie. "*The Book of Why: The New Science of Cause and Effect*". Basic Books, New York, NY, USA 2018.

7 Keith Stanovitch. "*Rationality and the Reflective Mind*". Oxford University Press, Oxford, UK 2010.

No entanto, mesmo considerando todos os actuais avanços, estamos ainda muito longe de criar uma mente digital; embora essa limitação, que diríamos circunstancial, não nos impeça de implementar na máquina funcionalidades que, no passado, eram exclusivas da mente humana. Sucede *apenas* que as máquinas não têm consciência das tarefas que estão a realizar, nem as articulam numa sinfonia cognitiva bem composta. Podemos, por exemplo, programar a geração de contra-factuais; criar simulações de natureza estatística que nos permitam validar os efeitos de comportamentos egoístas, ou altruístas, ou de um misto de altruísmo e egoísmo devidamente doseado; ou ainda avaliar os efeitos do acto de pedir desculpa no contexto da moral dos grupos.

Tais aspectos são ganhos de investigação alcançáveis no contexto do conhecimento sem consciência, propiciado pela actual IA. Devemos reconhecer que estamos limitados não apenas pelo actual poder de computação, mas fundamentalmente porque ainda não conhecemos muito bem os processos que fazem a consciência emergir em certos organismos vivos; também não conhecemos suficientemente os recursos e procedimentos através dos quais a nossa mente constrói mapas internos e externos, por forma a interpretar o presente e antecipar o futuro, e temos um conhecimento fragmentado dos nossos mecanismos morais.

A mente humana, apoiada num entrelaçado de raciocínios e emoções, tem esse carácter antecipador, não apenas porque gostamos de construir melhores explicações, mas fundamentalmente porque sabemos que as nossas possibilidades de sobrevivência estão, como sempre, muito dependentes dessa capacidade de antecipar futuros possíveis; e, muito provavelmente, se tivermos tal poder, de influenciarmos no sentido do que mais nos convém e sabermos evitar e combater os seus maus usos.

4. Perigos e precauções

Presentemente, há um *chatbot* de conversação que sinaliza de vez a presença cada vez mais determinante da IA no nosso quotidiano; referimo-nos ao ChatGPT4, uma criação da empresa Open AI. Este Chat GPT tem a particularidade de interagir connosco numa dinâmica de perguntas ou comentários e respostas; não apenas alimentado por dados existentes na internet, como também pelas modalidades de interacção que estabelece com humanos. Num diálogo aprofundado, o Chat GPT não passa o teste de Turing e outros testes que se lhe põem no caminho; mas, quando se trata de fornecer definições de conceitos, teorias, comparações entre teorias, e muitas outras acções baseadas em conhecimento já padronizado e bastante difundido, apresenta um desempenho assinalável.

No entanto, devemos ter cuidado com os seus limites, e os dos chats seus similares, pois enquanto nos ajuda com respostas que vão sendo reforçadas pelos utilizadores, também reduz o leque de possibilidades de abordagem a um problema ao que cabe dentro dos padrões maioritariamente validados. No limite, as ferramentas que trabalham grandes massas de dados a fim de encontrar padrões acabam por reduzir as alternativas aos padrões dominantes. Gostamos da facilidade, esquecendo com demasiada ligeireza que estamos perante máquinas de colectar dados; no caso do Chat GPT, reproduzindo as expressões linguísticas mais comumente associadas na miríade de documentos *consultados*. Ainda assim, dado que na maior parte de nosso tempo lidamos com conhecimento padronizado, é possível que a máquina nos crie a ilusão de ser um interlocutor genericamente credível, nomeadamente pelo seu bom domínio gramatical, comparável ao humano.

Para aprofundarmos um pouco mais os efeitos múltiplos da tecnologia em questão, e de outras que brevemente surgirão no mercado, podemos tomar como termo de comparação a disseminação das máquinas de calcular. O efeito da massificação do uso de tal instrumento foi a externalização de uma capacidade que, anteriormente, apenas podia ser desempenhada por um cérebro humano. Concretizado este processo – actualmente, parte do menu de qualquer smartphone de gama baixa – os humanos passaram a poder fazer cálculos complexos com muito maior rapidez e segurança; mas a consequência, porventura não desejada, foi a diminuição drástica do adestramento dos cérebros humanos nesse domínio. É certo que aprendemos a calcular na escola, mas também não há dúvidas de que, sem exercícios futuros, acabamos por esquecer certas regras segundo as quais se executam essas operações. O resultado, no que concerne à faculdade específica de calcular, é a nossa dependência das máquinas.

Do ponto de vista estritamente humano, não diremos, ainda assim, que perdemos muito; pois estamos a considerar uma competência mecânica, que pouco acrescenta à flexibilidade multifuncional que se espera de um cérebro humano. Acresce que os riscos decorrentes da perda de tal faculdade são baixos, já que perto de nós existirá sempre uma máquina de calcular! Por contraposição, um chatbot como o GPT, ou outro sucedâneo mais desenvolvido, tem capacidade para substituir competências – a exemplo da do cálculo – mas num vasto espectro de funções permitidas pela nossa mente, e expressas em linguagem natural. Referimo-nos a sínteses, explanações de teorias, comparações, e muitas outras capacidades que apenas realizamos por termos uma mente.

Ora, mesmo sem ser de facto uma mente, uma ferramenta como o ChatGPT, operando sobre dados já produzidos, pode actuar como tal e apresentar resultados muito

aceitáveis. Combinando os dados existentes, pode ajudar a criar compostos químicos inovadores que sejam melhores medicamentos, novas terapêuticas, ou evidenciar padrões comportamentais que antes nos escapavam. Em suma, a capacidade de tratar e retirar informação de grandes massas de dados e as combinatórias exploradas através de estratégias diversas—a exemplo do que nós, os humanos fazemos, mas sem a capacidade de lidar com o mesmo volume—são um prometedor campo de aplicação das máquinas que usam os modelos LLM (*Large Language Models*).

Neste sentido, estas tecnologias têm capacidade para, ou gerar dependência, ou para se tornarem parceiras, em áreas muito mais significativas para os humanos do que o cálculo. Ou seja, usadas sem critério e massificadas, comprometerão a nossa disposição para construir conhecimento e desenvolver pensamento crítico, detecção de contradições e procura de fontes, por, aparentemente, nos substituírem nessas funções, já que os actuais chatbots não têm capacidade crítica de opinião nem sobre a credibilidade das suas fontes (que até podem, recursivamente, serem outros chatbots). Por outro lado, devidamente contextualizadas e integradas a título de ferramentas, serão um excelente auxiliar para agregação de informação sobre a qual a mente poderá exercer as suas superiores faculdades.

É neste contexto que realçamos o facto de parte substancial das nossas aspirações sobre o conhecimento estarem relacionadas com a necessidade de antecipar e moldar o futuro. Ora, nesse domínio uma máquina que agrega dados e encontra padrões neles; nomeadamente, identifica as palavras que, estatisticamente consideradas, melhor se articulam com as anteriores, seleccionando a mais adequada, será de pouca utilidade. Pensar fora dos padrões, ou – de modo mais ousado – pretendermos usar o conhecimento que detemos como base para a criação de novos paradigmas, exige o uso de uma mente livre e ousada; capaz de colocar e explorar hipóteses que, dadas as suas características emergentes, não encontraremos nos padrões já estabelecidos.

Uma sociedade maquinal, altamente padronizada, normalizadora dos comportamentos, apreciará infelizmente as competências de tais máquinas e de mentes humanas moldadas por elas, restringindo-lhes por habituação estatística o vocabulário e conceitos nele expressos, à maneira da língua padronizada *Newspeak* no livro "1984" de George Orwell. Mas, nesse território, temos de ter a noção de que estamos já a contextualizar o que será uma ciberditadura, usando a máquina a título de delimitação dos âmbitos, e instrumento de controlo. Como se tivéssemos chegado ao fim da História, tendo agora como desígnio gerir e optimizar os recursos cognitivos adquiridos. Cumulativamente, surge um

novo neo-colonialismo cultural, uma vez que estas tecnologias tratam em número desigual línguas e alfabetos de todo o mundo, por desigualdade na origem dos documentos rastreiam.

Reproduzidos via ensino, os fenómenos perversos chegarão rapidamente às novas gerações, as quais terão menos sentido crítico e menos aptidão para verificar as fontes de informação. O recurso à informação estatística para formulação das suas próprias opiniões nas redes sociais converterá estes espaços em diálogos entre "GPTs" configurados de forma pessoal, como avatares que nos vão substituir. E estes são ingredientes perfeitos para potenciar o crescente afunilamento conceptual, de vocabulário e de opiniões. O cruzamento de fontes será o caminho para restaurar o pensamento crítico. Caso contrário, os humanos terão mentes cada vez mais previsíveis e formatadas nas respostas que vão dar.

O lançamento do ChatGPT foi como se disponibilizasse um novo avião que não foi avaliado e certificado. Isto acontece porque a legislação actual assume que o "software" não precisa disso. Mas precisa! No actual ambiente legislativo corre-se o risco de se lançar um vírus cognitivo com consequências importantes e não avaliadas previamente. A disponibilização destes instrumentos é como um "tsunami" que, como em quase tudo na informática, se propaga de forma imediata, por todos e por toda a parte. Acresce que os investigadores não estão habituados a este nível de avaliação de grandes impactos sociais da tecnologia. Pelo contrário, entendem que há uma oportunidade para arrecadarem mais recursos, abstendo-se de opiniões negativas globais, perpetuando dessa forma a prática de se esconderem os eventuais problemas debaixo da carpete. Com este contexto, a pergunta que subsiste é: quem beneficia lucrativamente com a IA? Essa questão não é trazida para a superfície. A IA deveria originar uma maior riqueza adquirida para todos, e não concentrar o lucro apenas em alguns.

A propósito, não podemos deixar de fazer uma breve alusão ao impacto destas novas tecnologias no mercado de trabalho. Dado que as profissões humanas, na sua maioria, são rotineiras, exigindo poucas capacidades ligadas à componente criativa e crítica da mente, máquinas rotineiras estarão em condições de substituir muitos humanos, sem que exista por agora uma reflexão suficientemente articulada sobre a matéria. Resumidamente, quer do ponto de vista colectivo, quer individual, devemos ter a noção de que as tecnologias da IA são tanto uma oportunidade, se devidamente enquadradas, como um risco, se abdicarmos das nossas faculdades e as substituirmos pelos seus aparatosos, mas ao mesmo tempo fracos recursos. Reiteramos que mentes digitais são, por agora, uma miragem e que os seus simulacros não agem por conta própria, mas ao serviço de reais men-

tes cujo interesse passa por funcionalizar os seus semelhantes, desvalorizar o trabalho, e concentrar a riqueza que será cada vez mais produzida por máquinas.

A IA está já a gerar impactos sociais graves em níveis cognitivos inferiores, mas muito rapidamente evoluirá para novos patamares cognitivos. Uma IA simples pode ser mais perigosa do que uma IA complexa, por ser meramente estatística. Numa interrogação ao ChatGPT, ele vai verificar as sequências de palavras/resposta mais prováveis na documentação que se relaciona com as palavras da interrogação. Ou seja, é um sistema que mede as coisas em termos estatísticos e que acaba por assumir que o futuro será igual ao que já está fixado do passado. Este nível, o do reconhecimento de padrões, é considerado o mais baixo da IA. Esta simplificação tem um lado perigoso e ao mesmo tempo perverso. Porque o ser humano, ao tornar-se muito dependente dela, treina cada vez menos o pensamento crítico, e a capacidade de avaliar a credibilidade e consistência da informação fica diminuída. Num nível superior de IA teríamos as hipóteses e os cenários possíveis, bem como o patamar confra-factual. A característica deste é o humano perguntar-se: «Que aconteceria se eu no passado tivesse feito de outra maneira?»

5. Da moral do homem à moral da máquina

Mas o que existe é o actual contexto, e nele interessa sobremaneira continuarmos a aprofundar o conhecimento sobre a mente humana, os seus processos e potencialidades. Esse conhecimento é a base para podermos implementar funcionalidades específicas em máquinas; esperando que futuramente seja possível uma melhor integração funcional. Por agora, como já mencionámos, um dos maiores riscos associados à IA é o de estarmos a delegar demasiadas funções, com os poderes a elas associados, em máquinas cuja inteligência é rudimentar.

De facto, a interacção com máquinas e a delegação de poderes nelas implica confiança; isto é, embora certos de que elas não têm uma mente crítica como a humana, temos de esperar que, ainda assim, tomem decisões de acordo com critérios aceitáveis para os seres humanos. Eis mais uma razão para suportar a tese que temos defendido, segundo a qual um dos domínios que urge investigar e aprofundar respeita a um melhor conhecimento da moral humana, os seus princípios fundamentais e o impacto deles nos grupos. É certo que estamos ainda numa fase muito embrionária, mas o esforço justifica-se porque não é possível dispormos de maquinaria cognitiva, com graus de autonomia cada vez maior, sem nos certificarmos que as mesmas são confiáveis. Acresce que o contexto onde

as máquinas têm de tomar decisões é, amiudamente, complexo; podemos tomar a guerra como exemplo, mas também os cenários decorrentes de catástrofes naturais, ou a necessidade de decidir no momento se se compram, ou vendem, certos lotes de acções na Bolsa; tal é o espectro de aplicação da IA.

Ora, esta investigação só é possível porque deixámos de percepcionar a moral como uma dádiva divina e, dentro do paradigma das investigações já referidas, olharmos para esse fenómeno evolucionário enquanto cola agregadora dos grupos e elemento fundamental da sua coesão. É assim que, presentemente, se investigam elementos presentes em todos os sistemas morais (independentemente da sua origem cultural), como é o caso do sentimento de culpa e do pedido de desculpa, ou do dilema cooperar/competir.

Quer em termos académicos, quer em termos empresariais, quer também em termos políticos, os avanços da IA colocaram a ética no centro da discussão. Mesmo na Meca do liberalismo, os EUA, os impactos possíveis de uma ferramenta como o ChatGPT levam a que empresários e outros cidadãos clamem pelo papel regulador do Estado. De entre os temas considerados como prioritários no âmbito da moral computacional, há a destacar a questão da responsabilidade algorítmica, onde se discute quem assume as consequências de acções e decisões tomadas por algoritmos e sistemas de inteligência artificial. Questões como o enviesamento algorítmico, a transparência na recolha de dados, a explicabilidade do processo de decisão, e a responsabilidade moral e legal por agentes artificiais são de natureza crítica; daí que a questão de como garantir que os sistemas de IA sejam projectados para respeitar a integridade humana, evitar fraudes, competição desregulada, e entrada no mercado de produtos sem os devidos testes de segurança sejam urgentes.

Em suma, referimo-nos sempre a questões que, anteriormente, estavam sob a exclusiva responsabilidade de mentes humanas. Este processo de delegação de competências é tão ancestral quanto a própria humanidade. O esforço físico árduo e repetitivo desumaniza-nos; por isso, os mais fortes escravizam os mais fracos e inventamos tecnologias que amenizam essas tarefas. Enquanto isso, trabalhamos os músculos no ginásio, mas essencialmente por razões estéticas, ou, na melhor das hipóteses, de saúde. Da mesma forma, o esforço intelectual é dispendioso em matéria de concentração e dispêndio de energia, de modo que aceitamos com tranquilidade, e até ansiedade, que as máquinas nos substituam nessas tarefas. E, nesse processo, não reflectimos suficientemente sobre os seus modos de concretização sem, simultaneamente, nos desumanizarmos.

Numa outra linha de problemáticas, encontram-se as questões associadas à recolha massiva de dados, e os impactos na privacidade e segurança dos utilizadores da internet e dos recursos digitais em geral. Se os dados são a matéria-prima mais valiosa do Século XXI, cumpre perguntar porque é que os seus produtores não recebem por eles. É certo que, não existindo mentes digitais capazes de, por elas próprias, transformar dados em conhecimento, não há o perigo de a nossa mente ser imediatamente controlada por máquinas. O que não significa que não o possa ser pelos seus donos ou pelos utentes dos serviços prestados pelas máquinas cognitivas.

Dada toda esta ordem de problemas, o desafio de se desenvolverem algoritmos que incorporem critérios morais nos seus processos de decisão, e que sejam capazes de os explicitar em justificações, é de fundamental relevância. Continuaremos a não ter, ainda assim, verdadeiras mentes digitais como parceiras, mas poderemos aspirar a mais confiança no apoio que é dado ao desenvolvimento da nossa. De facto, presentemente dispomos de algoritmos para desempenhar funções delimitadas, como pilotar um drone ou colectar dados sobre matérias específicas, encontrando aí padrões que escapavam ao olhar humano.

Dada esta especialização, que é, simultaneamente, uma severa limitação, especula-se sobre o que se designa por singularidade tecnológica, ou seja, esse momento hipotético no qual a máquina superará o humano concretizando o que se designa por Inteligência Artificial Geral (IAG). Para já, não estamos próximos desse momento, dado que, embora conheçamos com bastante detalhe cada um dos nossos processos cognitivos, ou o papel das emoções, estamos muito longe de nos pronunciarmos com detalhe sobre a sua integração num todo a um tempo consciente e inconsciente, e o modo como o nosso cérebro com o seu corpo, em interacção com os outros cérebros e corpos fazem isso. Aumentar a velocidade e a escala no que diz respeito a tratamento de dados não permitirá, em princípio, obter ganhos em termos qualitativos.

A forma mais complexa de inteligência que temos ao nosso alcance para estudo e aplicação em outros suportes é a humana, daí que os desafios estruturais da IA estejam muito dependentes do desenvolvimento das Ciências Cognitivas; estas, dada a sua abordagem interdisciplinar da mente e da cognição, interceptam a Filosofia, a Psicologia, a Inteligência Artificial, as Neurociências, a Linguística, e a Antropologia. O seu desenvolvimento, nas várias áreas que a compõem, é de crucial relevância para suportar a emergência de melhores modelos de cognição artificial, ancorados num melhor conhecimento das mentes biológicas e dos seus processos. De entre toda a panóplia de problemas que

constituem a sua agenda, gostaríamos de referenciar a emergência de um novo inconsciente colectivo, de natureza digital, moldável pelas preferências dos utilizadores e com um incomensurável potencial não só para condicionar os comportamentos humanos, mas também para influenciar o domínio emergente da Ética da Máquina. E é com esse tópico que gostaríamos de terminar o presente artigo.

6. Moralidade humana e ética da máquina

A ética da máquina questiona como projectar, implementar e tratar robôs; que capacidades morais devem eles ter, e como concretizar cada uma delas. Dada a complexidade do tema, em vez de almejarmos fixar todos os critérios para a competência moral de um robô, podemos ter como objetivo identificar certos elementos da moral humana e, em seguida, investigar o desenho da máquina considerando selectivamente alguns deles.

Para concretizar tais desideratos, temos de ter em conta que os robôs poderão interagir entre si, mas sempre ao serviço de algum interesse humano, grupal ou individual. Daí que os seus padrões de moralidade tenham de estar em simbiose connosco. Ora, a moralidade desenvolveu-se durante a evolução. Somos uma espécie gregária, o que implica ter regras de convivência. Por outro lado, não existe uma teoria universal da ética, mas uma combinação de teorias éticas: Categórica; Construtivista; Utilitária; de Virtude; etc. Daí o ser problemático que não conheçamos a nossa moral suficientemente bem, e com detalhe, para que ela possa ser prontamente programada. Dado esse constrangimento, devemos começar por programar as normas que já temos bem padronizadas e definidas em contextos específicos, como sejam o hospitalar, as bibliotecas, as casas de repouso, a negociação financeira, os parques de diversões, os centros de compras, ou os teatros de guerra. Nunca podemos perder de vista que estamos apenas no início da ética de programação para máquinas.

Ainda considerando essa fase embrionária, é urgente que se conheçam melhor as facetas morais humanas como sejam o vocabulário moral, as normas morais, a cognição moral e o afecto, a tomada de decisão moral e a sua relação com a acção, ou a comunicação moral e o consentimento. O seu estudo profundo é um pré-requisito para se progredir no ADN ético da máquina. Algumas capacidades que servem de substracto ao comportamento ético nem precisam de linguagem verbal; referimo-nos ao reconhecimento de comportamentos prototipicamente pró-sociais e antissociais, ou à empatia e reciprocidade básicas. Já para outras, torna-se necessário um vocabulário padronizado a fim de que

possamos aprender, ensinar e deliberar sobre elas de modo bem definido. Não devemos deixar fora do processo de explicitação e definição todo o vocabulário para expressar práticas morais como sejam as decisões de culpar ou desculpar um comportamento, perdoar, justificar, ou negociar as prioridades duma norma.

Além disso, necessitamos de um vocabulário para as próprias normas: nomeadamente no que respeita aos termos *justo, virtuoso, recíproco, honesto, obrigatório, proibido, desejável* etc. E ainda de explicitar o que entendemos por palavras que exprimem violações de normas como sejam *errado, culposo, imprudente, ladrão, intencional, conscientemente, acidental etc*.; bem como de resposta às violações de noemas, como sejam *culpa, repreensão, desculpa, perdão etc.*

Para formar juízos dirigidos a agentes, como seja a culpa, um robô precisa de capacidades para raciocínio causal sobre eventos segmentados, Inferências sócio-cognitivas do comportamento para determinar intencionalidade e razões, além de raciocínio contra-factual para decretar a prevenção. Note-se que uma componente proeminente da competência moral humana é a tomada de decisão e respectiva articulação com a acção. Neste contexto, a culpa é pedagógica pois é o modo de fornecer ao infractor da norma razões para não repetir. A culpa regulará o comportamento do robô se ele aprender a levá-la em consideração, nas suas próximas escolhas. Já o livre-arbítrio metafísico não é, de todo, necessário.

Na projecção de robôs capazes de decisões e ações morais, a tensão entre o interesse próprio e os benefícios da comunidade deve ser evitada desde o início. Deve ainda ter-se em conta que os robôs de diferentes fabricantes irão competir entre si! O tipo de robô que imaginarmos não pode ser programado para agir moralmente em todos os futuros possíveis. Terá normas orientadoras no início, mas precisa aprender a reformulá-las e a seleccionar outras. Portanto, corre o risco de deixar de agir moralmente por ignorância. Daí que o feedback seja fundamental para ele poder fazer melhor na próxima vez. No entanto, algumas situações apresentam problemas de decisão em que nem todas as normas relevantes podem ser satisfeitas em conjunto. Tais dilemas morais exigem uma escolha genuína entre opções imperfeitas.

Noutras ocasiões, cada opção pode ser moralmente justificada por referência a normas aceitáveis. Daí que as ferramentas cognitivas para juízo moral e tomada de decisão, por si só, sejam insuficientes para a função social de regular o comportamento dos outros. Ou seja, por vezes disposições humanas como a compaixão, ou a empatia, são critérios utilizáveis em ordem à boa decisão moral. Sendo assim, os robôs precisarão ganhar

flexibilidade de capacidade de reconhecimento de disposições humanas, geradores de um nível de confiança tal que lhes permita monitorizar e fazer cumprir as normas. Devem ainda declarar a obrigação de denunciar infrações às normas, e utilizar a comunicação para alertar e lembrar os preceitos aplicáveis.

As simulações envolvendo IA são ainda um veículo privilegiado para ensinar e treinar comportamentos morais, de forma interactiva com humanos, através de jogos morais. Tais jogos de computador podem ser empregues para testar teorias éticas e melhorar a educação moral, por meio de exemplos e explicações. Esses jogos podem contribuir com ferramentas para conceber, gerar e ilustrar comportamentos morais interactivos, em jogos multi-jogador, individuais e colectivos (Pereira 2022)[8].

E se, no processo de diversificação da oferta, surgirem máquinas com moral incompatível? Este é um cenário plausível, pois diferentes fabricantes produzirão máquinas com software moral distinto. Esse cenário pode vir a revelar-se caótico, pois as máquinas precisam de cooperar entre si por meio de uma moralidade comum, em vez de competir fora da ética, havendo o risco de robôs deliberadamente programados com intenções sinistras. Um objetivo importante da moralidade é, pois, a detecção de intenções desfavoráveis, trapaceiras e aproveitadoras. Só devíamos aceitar máquinas inteligentes autónomas se sua bússola moral for semelhante à nossa. Mas, dado que a Jurisprudência e o Direito derivados da Ética da Máquina e da Moralidade Humana estão atrasadas, tão cedo não podemos esperar uma moralidade genérica para máquinas cognitivas.

Em suma, sabemos, a partir da moral humana, que podemos ser sumamente solidários com os elementos do nosso grupo, enquanto competimos e guerreamos com os grupos rivais. Sabemos também que o domínio ético é o do *dever ser* e que, por vezes, actuamos de acordo com o que podemos, e não com o que reconhecemos ser esse tal dever. A investigação no domínio da ética, orientada para a implementação de uma IA benévola, não resolverá todas as incongruências e contradições desse domínio; mas, um melhor conhecimento da moral humana, dos fundamentos que estão para além da sua concretização em preceitos, é o único caminho para irmos mitigando os efeitos nefastos

8 Luís Moniz Pereira, The Anh Han, António Barata Lopes. Employing AI to Better Understand Our Morals. *Entropy* 24(1): 10, 2022.
Luís Moniz Pereira, António Barata Lopes. Machine Ethics: From Machine Morals to the Machinery of Morality. In: *Studies in Applied Philosophy, Epistemology and Rational Ethics* (SAPERE, volume 53), Springer Nature AG, Switzerland, 2020.
Luís Moniz Pereira, António Barata Lopes, Máquinas Éticas: Da Moral da Máquina à Maquinaria Moral, Colecção: "*Outros Horizontes*", NOVA.FCT Editorial, Campus FCT-UNL, 2829-516 Caparica, Portugal, 2020.

da competição sem regras, e para uma *inter-moralidade* agregadora de agentes cognitivos autónomos e colaborativos. Esses agentes poderão e deverão, também, competir; mas de acordo com regras explicitadas.

O fantasma da super IA (AGI) é um mito que nos desvia a atenção do importante. Para nós, o risco de extinção pela AGI não é sobre a IA se apoderar de nós, humanos, mas sobre o uso que os humanos farão da IA sem a devida preocupação coordenada com outros humanos. E essas atividades cumulativas podem resultar em efeitos colaterais incontroláveis, colocando a espécie em risco. Não vamos colocar esse problema, o verdadeiro, debaixo do tapete, levantando o fantasma de uma super IA. Os fantasmas de tudo o que traz ainda mais lucro estão em abundância entre nós. Mas agora os desenvolvedores de referência têm ferramentas de IA mais poderosas para promover seus erros, e inesperados efeitos colaterais emergentes surgirão, para além do controlo de qualquer pessoa ou nação.

REPRESENTAÇÃO E COGNIÇÃO SITUADA: UMA PROPOSTA CONCILIADORA PARA AS GUERRAS REPRESENTACIONAIS

Carlos Henrique Barth (Doutorando, PPG-Filosofia, UFMG)

Felipe Nogueira de Carvalho (Professor e Pesquisador, UFLA)

Resumo:

Abordagens pós-cognitivistas mais recentes têm lançado duras críticas à noção de representação mental, procurando ao invés disso pensar a mente e a cognição em termos de ações corporificadas do organismo em seu meio. Embora concordemos com essa concepção, não está claro que ela implique necessariamente a rejeição de qualquer tipo de vocabulário representacional. O objetivo deste artigo é argumentar que representações podem nos comprar uma dimensão explicativa adicional não disponível por outros meios e sugerir que, ao menos em alguns casos, elas podem participar da explicação de performances ou capacidades cognitivas. A noção de representação apresentada, como deixaremos claro ao longo do artigo, não viola os preceitos metodológicos mais caros à cognição 4E, em geral, e ao enativismo, em particular, podendo, portanto, ser utilizada como uma ferramenta teórica útil em investigações sobre a natureza corporificada e situada da mente.

Abstract:

Recent post-cognitivist approaches have raised sharp criticisms against the notion of mental representation, proposing instead to think of the mind and cognition in terms of embodied actions of an organism in its environment. While we agree with this conception, it is not clear that it necessarily implies the rejection of any kind of representational vocabulary. The aim of this paper is to argue that representations afford an additional explanatory dimension that's unavailable through other means, and to suggest that, in at least some cases, they may participate in the explanation of cognitive performances or capacities. The presented notion of representation, as we will make clear throughout the paper, does not violate the methodological precepts most dear to 4E cognition in general and enactivism in particular, and can therefore be used as a useful theoretical tool in investigations about the embodied and situated nature of the mind.

Palavras-chave:

Representação mental; cognição situada; cognição 4E; enativismo.

Keywords:

Mental representation; situated cognition; 4E cognition; enactivism.

Introdução: breve relato das guerras representacionais

Embora a noção de 'representação' tenha desempenhado um papel fundamental no desenvolvimento das ciências cognitivas a partir dos anos 50 e 60, abordagens pós-cognitivistas mais recentes – notadamente o enativismo – têm lançado duras críticas a esta noção, redesenhando o estudo da mente de forma a dispensar qualquer tipo de vocabulário representacional. Assim, ao invés de pensar a cognição em termos de computações sobre estados internos com propriedades semânticas que (de alguma forma) representam o mundo externo, o enativismo enfatiza o caráter corporificado e situado da cognição, insistindo que o mental emerge da exploração ativa do organismo em seu meio e que não há um 'mundo' a ser representado independentemente do modo como este organismo age e modifica seu meio. Ao contrário, mente e mundo são co-determinados mutuamente a partir do histórico de enação do organismo. Nessa perspectiva, para explicar a ação bem-sucedida (e os erros) de um organismo, basta especificar seus acoplamentos habilidosos em seu nicho, não havendo necessidade de se postular algo como uma representação interna do mundo exterior. Como dizem Varela, Thompson & Rosch no clássico enativista *The Embodied Mind*:

> We propose as a name the term *enactive* to emphasize the growing conviction that cognition is not the representation of a pregiven world by a pregiven mind but is rather the enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs (Varela; Rosch; Thompson, 1991, p. 9).

Mesmo que o enativismo seja um programa de pesquisa amplo com diferentes vertentes[1], é seguro dizer que elas têm em comum uma certa concepção da mente expressa

1 Como por exemplo o enativismo autopoiético de Varela, Thompson & Rosch (1991), o enativismo radical de Hutto & Myin (2013) ou o enativismo fenomênico de Noë (2004). Para uma excelente introdução a esse programa de pesquisa em suas diferentes vertentes, ver Rolla (2021).

pelos 4 "E" da cognição – ou seja, que esta é essencialmente corporificada (*embodied*), situada (*embedded*), enativa (*enacted*) e estendida (*extended*). Neste artigo também subscrevemos a essa concepção[2], algo que nos coloca em um ambiente teórico hostil para qualquer tipo de vocabulário representacional, sob o risco de associação a um ultrapassado "cognitivismo de velha guarda" que Rolla (2021, p. 23-4) coloca da seguinte forma:

> Cognitivismo de velha guarda, ou antigo cognitivismo, é a combinação de duas teses: em primeiro lugar, que a mente é essencialmente representacional (representacionalismo) e, em segundo lugar, que processos cognitivos são computacionais (computacionalismo). Segundo essa perspectiva, seria possível entender como a mente funciona completamente à parte das características corporais do organismo e do ambiente em que ele se encontra.

Embora estejamos de acordo com as críticas enativistas ao chamado "cognitivismo de velha guarda", ainda não é claro para nós que pensar a cognição como corporificada, enativa e situada implique a rejeição de qualquer tipo de vocabulário representacional; ou, dito de outra forma, que introduzir 'representação' como um termo teórico útil em alguns contextos nos comprometa com algo remotamente semelhante ao que Rolla (2021) descreve como "cognitivismo de velha guarda". Mas como nos encontramos em um território hostil a esse tipo de vocabulário, será preciso esclarecer o que entendemos por 'representação', porque julgamos que esta noção pode nos trazer alguma vantagem explicativa em certos casos e porque aceitá-la não implica de forma alguma trair esses preceitos metodológicos.

Que fique claro, nosso objetivo neste artigo não é argumentar em favor de algum tipo de representacionalismo. Na verdade, não acreditamos que seja possível sustentar o representacionalismo ou o antirrepresentacionalismo como uma tese generalizada acerca da cognição. Ao contrário, se devemos ou não introduzir uma 'representação' como um termo teórico útil é algo que só pode ser estabelecido empiricamente a partir de uma análise caso a caso, considerando as características específicas do organismo, da tarefa e do comportamento a ser explicado.

Nossa estratégia neste artigo consistirá, portanto, em: (i) argumentar que, pelo menos em alguns casos, postular representações pode nos comprar uma dimensão explicativa adicional que não está disponível fora deste vocabulário; e (ii) argumentar que esta noção de representação não contradiz os preceitos metodológicos mais caros ao enativis-

2 Embora permaneçamos neutros em relação ao quarto "E", isto é, a mente estendida, enquanto uma tese ontológica sobre as fronteiras do mental.

mo. Em particular, (a) ela não nos compromete com um mundo pré-dado representado por uma mente pré-dada, nem (b) com um computacionalismo ou representacionalismo generalizados. Isto é, mesmo que a representação seja um recurso teórico útil em certos casos, disso não se segue que ela possa ser aplicada sem restrições a outros casos.

A estratégia de apresentação será a seguinte: na seção (1), delinearemos nossa posição entre o cognitivismo clássico e as abordagens não representacionais como o enativismo. Na seção (2), apresentaremos a distinção entre estados representacionais e estados intencionais (alvos), que será utilizada ao longo de todo o argumento. Na seção (3), apresentaremos a teoria da representação de Robert Cummins, argumentando que ela é capaz de evitar tanto os problemas clássicos quanto os desafios antirrepresentacionais conhecidos. Na seção (4), argumentaremos pela relevância explicativa das representações, sugerindo que elas nos permitem formular teses empíricas que estariam indisponíveis (ou seriam distorcidas) em sua ausência. Na seção (5), antes de proceder para a conclusão, faremos algumas considerações sobre como essas ideias podem ser adotadas em *frameworks* já existentes.

1. Batalha de *frameworks*

Como vimos, é comum caracterizar o cognitivismo clássico pelo seu comprometimento com o computacionalismo e o representacionalismo: a mente cognitiva é constituída por registros representativos do mundo sobre os quais se realizam operações exclusivamente computacionais. Nessa perspectiva, a cognição se esgota no papel de mecanismos computacionais e estados representacionais. Menos comum é enfatizar que o cognitivismo clássico adotava uma concepção específica do que são mecanismos computacionais, e uma concepção igualmente específica do que são estados representacionais. Por exemplo, representações mentais seriam simbólicas e registrariam o mundo num formato proprietário, com forma sentencial (FODOR, 1980)[3]. A ontologia subjacente a essa forma de registro (objetos, propriedades e relações) é o que tornava plausível descrever processos cognitivos por meio de formalismos lógicos. Vem daí também sua característica mitigação da importância da percepção. Processos perceptuais seriam "anteriores" a esse registro do mundo, e não pareciam ensejar dificuldades de

3 Uma vez que o foco deste trabalho são representações mentais, deixaremos a discussão sobre mecanismos computacionais de lado.

naturalização[4]. A despeito de todos os seus problemas, essas ideias harmonizavam bem com a tese de que a cognição compreende um domínio de pesquisa autônomo e independente do corpo e do ambiente.

A explicitação desse modo específico de compreender o que são representações – e qual seu papel – é fundamental para distinguir bons argumentos contra o cognitivismo clássico de bons argumentos contra representações mentais em geral. Quando a autonomia do domínio cognitivo começou a ser questionada por uma nascente concepção 4E, nem todos se atentaram a isso. Via de regra, adotou-se um apressado pessimismo quanto à possibilidade de convivência pacífica entre estados representacionais e os preceitos da cognição 4E. Isso se mostra ainda hoje na resistência de alguns enativistas em aceitar qualquer concepção de representação mental. O temor parece ser o de que, ao permitir o uso de recursos explicativos de berço cognitivista, abra-se espaço para a adesão sub-reptícia de princípios metodológicos indesejados.

Um exemplo concreto de como esse receio se manifesta está na relação entre cognição, ação e percepção. A cognição 4E entende que a relação organismo-ambiente deve ser tomada como unidade de análise elementar. Resulta daí um abandono do que Hurley (2001) caracterizou como modelo "sanduíche" da relação entre percepção, mente e ação. A mente cognitiva não seria um espelho do mundo posicionado entre *inputs* perceptuais e *outputs* comportamentais, mas algo que emerge da interação contínua e direta entre ação e percepção. Se reintroduzirmos alguma concepção de representação, não poderia isso nos forçar a uma retomada do modelo sanduíche?

Trabalhos como o de Piccinini (2021) e Clark (2016) respondem a essas preocupações acomodando preceitos caros à cognição 4E sem abrir mão de representações. Porém, isso tem se mostrado insuficiente. Parte do problema é que persiste o clima de embate entre *frameworks* que almejam autossuficiência. Isso gera uma luta por espaço que só acabará se e quando uma engolir a outra. Esse embate é alimentado por diagnósticos equivocados acerca do que representações são e de como são usadas. A preocupação com um possível retorno ao modelo sanduíche, por exemplo, não diz respeito à existência de conteúdo representacional. Ela concerne se, quando e como esse conteúdo é explorado pelo sistema. No decorrer do artigo, buscaremos mostrar como essa distinção pode ser feita. Por ora, importa notar que, se estivermos no caminho certo, talvez seja possível mudar a relação entre os *frameworks* atuais: em vez de inflamar a disputa por espaço, será possível ampliar o conjunto de preceitos partilhados.

4 Até onde se sabe, capacidades perceptuais nunca motivaram nenhum tipo de dualismo ao longo da história.

Mas por que proponentes do enativismo, da psicologia ecológica e/ou de quaisquer abordagens não representacionais deveriam nos dar ouvidos? Há muito trabalho sendo feito para mostrar como capacidades que parecem demandar representações mentais podem ser explicadas mesmo na sua ausência. Certamente nenhum desses trabalhos é conclusivo, mas por que abrir mão do norte que perseguem? Não é isso que queremos sugerir. O ponto é mostrar que nenhum preceito essencial à cognição 4E precisa ser abandonado, mesmo que ao fim do dia as representações se mostrem indispensáveis em alguns casos. Se formos bem sucedidos, portanto, teremos em mão uma proposta legitimamente conciliatória e (esperamos) bastante atraente.

2. Preliminares: a distinção entre representações e alvos

Há algum tempo, Cummins (1996) percebeu que os representacionalistas tendem a ignorar uma distinção entre duas dimensões presentes nos mecanismos cognitivos que exploram representações. Quando um sistema utiliza uma representação para lidar com algum elemento do mundo, essa aplicação pode ser analisada tanto pela via semântica quanto pela via funcional. De um lado, a dimensão semântica responde por aquilo que um dado estado ou processo efetivamente representa. Do outro, temos a dimensão do alvo a que a representação é aplicada, ou seja, aquilo que o sistema pretende explorar pelo uso da representação como um modelo do alvo. A noção de alvo é funcional: ela especifica aquilo que um dado mecanismo ou processo tem a função de representar em uma dada circunstância.

Essa distinção é facilmente identificável no âmbito agencial. Suponha-se um grupo de amigos a viajar por um país desconhecido nos anos 1990. Igor recebe a tarefa de providenciar um mapa do bairro em que se encontram para que todos possam localizar um museu que desejam visitar. Ele não encontra nenhum mapa impresso à venda, mas coleta informações diversas dos habitantes locais. Com base nelas, desenha a estrutura das ruas, suas intersecções, acrescenta alguns pontos de referência e, por fim, especifica a localização do museu. Ele reencontra os amigos e lhes apresenta o mapa, que será então utilizado por todos para chegar ao local desejado. Esse exemplo permite ver com clareza que, qualquer que seja o conteúdo do mapa, ele será tomado pelos amigos como um *mapa do bairro da cidade em que se encontram*. Esse era o alvo de Igor em virtude da função que lhe foi atribuída.

Contudo, o mapa produzido pode não representar adequadamente a estrutura do bairro da cidade. Um dos habitantes consultados pode ter mentido, ou talvez se confun-

dido, ainda que de boa fé. Seja como for, todas as inferências feitas pelo grupo partirão do princípio de que o mapa representa o bairro da cidade em que eles se encontram. Essa é a distinção usada por Cummins (1996) no âmbito sub-pessoal: um mecanismo pode ter como função representar um dado alvo X. A representação que ele produz será tomada pelos mecanismos consumidores como sendo de X, mesmo que ele produza algo que não represente X acuradamente.

Neste cenário, estados ou processos representacionais explorados por sistemas cognitivos se relacionam com o mundo por dois caminhos: primeiro, por terem conteúdo (i.e. representarem algo); segundo, por serem explorados como modelos de algo. Essa distinção permite que cada dimensão seja objeto de uma teoria específica. Uma teoria do conteúdo representacional trata da dimensão semântica, e uma segunda teoria é responsável por tratar do modo como um dado organismo fixa os alvos aos quais aplicará as representações produzidas. Para Cummins (1996), essas teorias podem e devem ser independentes. Perguntar pela cidade que um mapa representa é diferente de perguntar pela cidade que se está tentando explorar por meio daquele mapa. Da mesma forma, o modo como o conteúdo de um estado ou processo é fixado independe do modo como o organismo fixa seus alvos.

Com efeito, o caráter intencional de um determinado estado ou processo não é fruto do conteúdo representacional explorado, mas sim do alvo fixado. A intencionalidade é um fenômeno mais amplo que a intencionalidade representacional. Depreende-se disso que, em larga medida, o debate contemporâneo entre abordagens representacionalistas e não representacionalistas pode ser compreendido como um debate sobre o modo como um dado organismo consegue fixar em certos alvos. Um alvo pode ser fixado tanto pela especificação funcional de um mecanismo inato, quanto pela especificação de um processo que emerge num arranjo muito particular envolvendo o aparato cognitivo, o corpo e o ambiente do organismo. Por exemplo, a tese de que organismos exploram informação ecológica é uma tese sobre fixação de alvos. Ela nos diz que sistemas (organismos) podem fixar como alvos certas estruturas complexas presentes no ambiente, inclusive estruturas dinâmicas, como as que se articulam ao longo do tempo. Evidentemente, cognitivistas e enativistas podem divergir sobre quais os alvos utilizados por um dado organismo e como ele os fixa. Mas essa divergência não necessariamente diz respeito à aplicação de representações no modo como esses alvos são explorados. Representar não é um modo distinto de fixar alvos. Representações não são um meio de caracterizar a sensibilidade a características do ambiente, mas sim uma estratégia à disposição do organismo para explorá-las e armazenar informação sobre elas.

Sob a luz dessa distinção, podemos nos voltar agora para as razões que tipicamente justificam o pessimismo antirrepresentacionalista. Costuma-se explorar dois flancos: primeiro, uma aparente dificuldade de naturalização. Segundo, o papel aparentemente trivial da representação na explicação científica, geralmente em virtude da suposta inércia causal do conteúdo semântico. Ambas conduzem a uma negação da realidade do conteúdo representacional ou a abordagens deflacionadas, como a de Egan (2020). Mas elas podem ser contornadas, e é disso que nos ocuparemos agora.

3. Representações: o que são?

Nos últimos anos, a dificuldade de naturalizar estados representacionais se condensou na formulação de Hutto & Myin (2013), conhecida como *o problema duro do conteúdo*. Uma síntese do argumento pode ser encontrada em Rolla (2023, p. 213):

> […] de acordo com cognitivistas, representações mentais são portadoras de informação semanticamente carregada […]. Porém, o único tipo de informação encontrada na natureza é a covariação. […] Não podemos inferir que um dos termos em uma relação de covariação representa o outro. Números de anéis no tronco da árvore não representam a sua idade — a representação aqui é imputada por nós […]. Por si só, estados naturais são piamente quietistas e não dizem nada sobre ninguém. […] [P]or que seriam os estados cerebrais diferentes dos demais estados naturais? […] Ou o cognitivista aceita que existe cognição sem conteúdo (acarretando o enativismo) ou nos passa um cheque sem fundo com a promessa de que um dia a física do futuro vai descobrir conteúdos semânticos em estados naturais. A segunda via é naturalisticamente temerosa. Donde se segue que o cognitivista deve dar o braço a torcer […] (Rolla, 2023, p. 213)[5].

O alvo do argumento são teses ancoradas na covariação confiável entre estados e propriedades: se F covaria com G, F *significa* G. Esse é o tipo de informação a que Dretske (1981; 1986) e outros recorrem para ancorar conteúdo representacional. Mesmo antes da formulação de Hutto & Myin (2013), os desafios dessa abordagem eram bem conhecidos[6]. Porém, não é necessário aprofundarmos a discussão. Independentemente do que se conclua

5 Embora esse ponto seja tangencial à discussão, é importante salientar que mesmo o eventual sucesso do argumento não acarretaria o enativismo como se afirma. A situação resultante seria compatível com toda sorte de teorias não representacionais, a exemplo do conexionismo eliminativista ou da "teoria sintática da mente" de Stich (1983), que questiona a relevância explicativa do conteúdo representacional sem abrir mão do modelo sanduíche da ação e percepção ou de modelos computacionais clássicos.

6 Por exemplo, Cummins (1996), Perlman (2002) e Ramsey (2007).

acerca de teorias baseadas em covariação, o argumento não cumpre o prometido porque é falso que a covariação seja a única relação natural explorável. Há também a relação de isomorfismo estrutural (ou homomorfismo, entendido aqui como um isomorfismo que acomoda gradações). Para Cummins (1996) e Swoyer (1991), a dimensão semântica dos estados ou processos representacionais pode ser completamente acomodada nessa relação: representação nada mais é do que a relação matemática de isomorfismo. Isso vale para qualquer estado ou processo físico, de mapas impressos a estados cerebrais. Uma estrutura A representa uma estrutura B na medida em que A é isomórfica a B. Com efeito, elementos de A representam elementos de B, relações entre esses elementos presentes em A representam relações entre esses elementos em B, e assim por diante. Temos assim uma semântica para *representações estruturais*.

A tese é simples, mas faz emergir dúvidas sobre como esse conteúdo pode cumprir o papel que se espera dele. Uma primeira preocupação diz respeito a uma aparentemente excessiva liberalidade: há estruturas isomórficas por toda parte. Isso significa que A pode representar não apenas B, mas também C, D, E, e assim por diante, indefinidamente. Um mapa contendo a estrutura das ruas de um dado bairro de Belo Horizonte (BH) representa aquelas ruas de BH, mas pode também representar as ruas de outras cidades, bem como qualquer outra estrutura existente que calhe de ser isomórfica. Se esse é o caso, o conteúdo representacional será sempre não único. Porém, é aqui que a distinção entre alvo e conteúdo começa a render dividendos. A pressão pela determinação do conteúdo se faz sentir tão somente no âmbito da fixação de alvos. Mais precisamente, o que pode fazer parecer com que a liberalidade da relação de isomorfismo seja inadequada é a confusão entre dois sentidos diversos em que podemos compreender a afirmação "A representa B":

(1) a estrutura A é isomórfica à estrutura B;

(2) a estrutura A é explorada por um sistema como um modelo de B.

Cummins (1996) se distingue ao usar o termo "representação" no sentido (1), pois o significado usualmente aplicado na literatura é (2). Considere, por exemplo, a seguinte caracterização de Hutto & Myin (2013, p. 62):

> To qualify as representational, an inner state must play a special kind of role in a larger cognitive economy. Crudely, it must, so to speak, have the function of saying or indicating that things stand thus and so, and to be consumed by other systems because it says or indicates in that way.

Tal caracterização claramente remete a (2). Ocorre que (2) confunde a dimensão semântica e a dimensão funcional associada à fixação de alvos. Ela não nos dá aquilo em virtude de que uma dada estrutura pode ser considerada representacional (para isso, o isomorfismo é suficiente). O que ela expressa são as condições a caracterizar mecanismos que exploram um dado estado representacional. Em outras palavras, ela caracteriza casos de *aplicação* de representações a alvos funcionalmente determinados.

Se prescindirmos da distinção entre alvo e representação, contudo, seremos tentados a concluir que só poder haver conteúdo onde há aplicação desse conteúdo a um alvo determinado. Isso é um problema porque, numa economia representacional, estas dimensões exercem papéis explicativos distintos. A dimensão semântica não se ocupa de identificar o que o mecanismo está tentando fazer, mas, sim, de contribuir para uma compreensão da performance resultante: por que o organismo foi bem (ou mal) sucedido? Se um mecanismo tem a função de fornecer o mapa de um bairro das ruas de BH, então seus consumidores irão tomar as representações fornecidas como sendo de BH. Ainda que ela calhe de ser isomórfica a um sem número de outras estruturas, isso não afeta a explicação da performance obtida. Afinal, são as ruas de BH que constituem a norma contra a qual a acurácia da representação produzida deve ser mensurada.

Nessa abordagem, nenhum dos problemas tradicionais relacionados à naturalização do conteúdo precisa nos preocupar. Não é preciso, nem faria sentido, reduzir isomorfismo (ou homomorfismo) a qualquer outra concepção de informação, visto ser ela uma relação perfeitamente natural. Não há, portanto, mistério sobre como uma representação se ancora no mundo (*grounding problem*). Não há problema de coordenação entre conteúdo e veículo, e tampouco há dúvidas sobre se e como eles podem ter eficiência causal: representações guiam o processamento em virtude de seu conteúdo, e seu conteúdo é dado pela sua estrutura. Uma vez que tanto os poderes causais quanto o conteúdo são dados pela estrutura, não há necessidade de "interpretadores" desse conteúdo. Consumidores das representações são causalmente sensíveis às estruturas, da mesma forma que um cadeado é causalmente sensível a uma chave. Além disso, não há real problema relacionado à determinação do conteúdo. Como vimos, o caráter não único do conteúdo estrutural não o impede de cumprir adequadamente o papel que o cognitivismo lhe prescreve. Afinal, a pressão por determinação se dá sobre a dimensão da fixação de alvos.

No cenário articulado, o problema de como identificar mecanismos que exploram representações (i.e. que exploram propriedades de uma estrutura a fim de lidar com um alvo do mundo), emerge como empírico. Deve-se determinar caso a caso se, e quando, um dado mecanismo explora estados representacionais no exercício de sua função. Conside-

re, por exemplo, o contraste com mecanismos não representacionais que exploram sinais emitidos por detectores. Tais sinais são indicadores da presença de alguma propriedade ou estado de coisas no mundo. Exemplos clássicos são sinais de detectores de temperatura em sistemas de ar condicionado, de nível de combustível ou pressão de óleo em veículos, e de certos padrões no campo visual, a exemplo das células no córtex visual primário. Esses sinais também têm alvos, mas não os representam. Seu conteúdo é tão somente uma indicação de presença que advém da especificação funcional do mecanismo emissor. Sabemos estar com pouco combustível diante de um sinal emitido por um mecanismo com essa função. Dissociar o sinal daquele detector faz com que a informação seja perdida. Indicadores e representações são, nesse sentido, fontes distintas de tipos distintos de conteúdo.

Como podemos distinguir casos em que um organismo está efetivamente explorando uma representação dos casos em que ele está explorando outros tipos de recurso, tais como sinais de indicadores? A mera detecção de estruturas instanciadas no aparato cognitivo é insuficiente. Elas podem caracterizar o que Cummins denominou "representações não exploradas" (Cummins *et al*., 2010), isto é, estruturas isomórficas a algo, mas que não são explorados em virtude desse isomorfismo[7]. A distinção é importante porque, como nos lembra Facchin (2021), mesmo mecanismos indicadores podem apresentar estados isomórficos ao que detectam no mundo[8]. Mecanismos que consomem sinais de indicadores, contudo, não são sensíveis à essas estruturas. Portanto, é preciso determinar se a capacidade ou performance estudada envolve processos causalmente sensíveis às estruturas instanciadas. Shea (2018) sugere que o foco desse tipo de investigação deve estar na correlação entre o grau de acurácia representacional e o comportamento resultante. Na medida em que representações estruturais mais acuradas resultam em comportamentos mais adequados, é plausível inferir que a estrutura exerce papel relevante.

7 A possibilidade de conteúdo representacional não explorado é usada por Cummins (1996; 2000) para criticar abordagens teleosemânticas como a de Millikan (1987). A tese central é a de que a teleosemântica inverteria a ordem explicativa, fazendo com que a atribuição de conteúdo derive da adaptatividade, quando o que queremos é explicar como foi possível que a exploração de um dado conteúdo tenha se mostrado adaptativa. Curiosamente, Hutto & Myin (2013) citam Cummins *et al.* (2010) explicitamente ao fazer suas próprias críticas a Millikan, sem, contudo, abordar a solução oferecida por Cummins no mesmo artigo (a mesma que aqui desenvolvemos).

8 Isso leva Facchin (2021) a concluir equivocadamente que essas estruturas não representam. Elas representam, sim, aquilo a que são isomórficas. O que está em jogo é a existência de algum mecanismo sensível a essas estruturas, isto é, consumidores.

Embora seja um passo na direção certa, Shea (2018) parte de uma suposição problemática: a de que maior acurácia representacional implica melhor performance. Mas sucesso semântico não implica sucesso comportamental ou funcional. No caso de criaturas limitadas em recursos (como nós), representações pouco precisas podem ser a razão pela qual um dado comportamento resultante se mostrou adaptativo ou apto o suficiente para ser adotado de forma robusta. Acurácia representacional costuma exigir maior dispêndio de recursos do organismo na tentativa de explorá-la, e isso pode resultar em comportamento inadequado. Além disso, uma resposta correta à pergunta "qual o grau adequado de acurácia?" é claramente dependente das circunstâncias. O que seria considerado um erro grosseiro numa situação pode ser suficiente ou vantajoso em outra. Situações que acomodem tempo e memória suficientes podem se beneficiar de uma precisão maior, e situações que impõem constrangimentos de tempo e memória (o organismo está fugindo de um predador) podem mitigar o grau de acurácia considerado ideal. Pela mesma razão, mecanismos mentais podem ser considerados funcionais ou eficientes, ainda que envolvam frequentes erros semânticos (i.e. imprecisões). Com efeito, os critérios adotados para identificar casos em que um organismo explora representações não podem pressupor correlação entre sucesso funcional, comportamental e semântico.

Embora isso complique as coisas, obtemos em troca um grau de liberdade crucial. Modelos que busquem explicar como funcionam os mecanismos subjacentes a uma dada capacidade cognitiva podem articular essas diferentes dimensões. Pode-se formular diferentes hipóteses empíricas sobre o alvo adotado, o conteúdo produzido (se algum), como esse conteúdo é explorado (se o for), e qual o comportamento resultante em diferentes circunstâncias. Ainda que diferentes modelos possam explicar satisfatoriamente o mesmo tipo de comportamento, a articulação dessas dimensões amplia o conjunto de possíveis efeitos incidentais associados a cada modelo. Por "efeito incidental", queremos indicar uma espécie de efeito colateral do modelo. Trata-se de um efeito causal que não participa da explicação da capacidade que se busca modelar, mas que pode ser útil para decidir qual tese evoca o modelo mais adequado dentre os candidatos disponíveis (Cummins, 2010). Para deixar esse ponto mais claro, vamos tratar primeiro de um exemplo hipotético e depois de um exemplo mais concreto.

Suponha-se que queiramos explicar como uma calculadora é capaz de multiplicar números e tenhamos dois modelos concorrentes. Um deles supõe que a calculadora realiza somas sucessivas: o resultado de 25x15 é dado por 15 somas sucessivas (25+25+25…). O outro modelo supõe que a calculadora faça uso do algoritmo de produtos parciais que todos aprendemos na escola. Ainda que ambos predigam o mesmo *output*, o primeiro

modelo o faz por meio de uma composição simples, enquanto o segundo explora uma relação estrutural existente no esquema arábico (note a importância da posição em que os produtos parciais são armazenados, e também como esse algoritmo está indisponível para a notação romana). Isso gera efeitos incidentais distintos. O primeiro modelo é mais sensível ao tamanho do número (15, 30, 45…), resultando numa determinada trajetória a descrever o consumo de tempo e memória. Já o segundo modelo prediz uma trajetória de consumo de recursos que, embora também sensível ao tamanho dos números envolvidos, se mantém estável quando o número de dígitos é o mesmo, apresentando saltos somente quando há um incremento destes (15, 115, 1015…). Temos assim um efeito empírico que pode ser mensurado para decidir qual dos modelos descreve melhor a capacidade que se buscava explicar, mesmo em casos onde o comportamento resultante não varie.

Um exemplo do mundo real pode demonstrar a utilidade de efeitos incidentais de modo ainda mais claro. Newen & Vosgerau (2020) apontam para um caso interessante envolvendo ratos e sua capacidade de articular informações sobre a organização espacial de labirintos, tipos de comida e períodos do dia. Eles demonstram capacidade de compreender, por exemplo, que quando um certo tipo de comida (salgada) foi disponibilizada em certa posição do labirinto na parte da manhã, haverá um outro tipo de comida (doce) em um outro dado local do labirinto na parte da tarde (Crystal, 2013; Panoz-Brown *et al.*, 2016). Os autores defendem que a melhor explicação disponível envolve articulações entre representações de tipos de comida, de períodos do dia e, claro, de localização espacial. Mais do que a verdade ou falsidade dessa explicação, o que realmente nos importa notar aqui é a argumentação utilizada pelos autores contra hipóteses alternativas:

> The informational state of rats which have learned to behave according to a conditional in the maze is best characterized as structured into components of <object-type; location; time>. The alternative would be to presuppose a high number of independent, non-structured dispositions which need to include all the possible permutations of associations between a starting state of affairs and a type of behavior. And these dispositions would need to be learned independently of each other, since there would be no common component to be taken over (Newen; Vosgerau, 2020, p. 184).

Essa argumentação depende crucialmente da articulação de um efeito incidental à capacidade sendo explicada. Embora seja concebível que o tipo de comportamento exibido possa ser explicado por um conjunto razoavelmente grande de diferentes *affordances*, essa alternativa requer pelo menos uma *affordance* independente para cada

permutação da estrutura representacional. Não sendo esse o caso, o comportamento do rato permaneceria parcialmente inexplicado. O problema com esse requisito é que ele implica a manifestação de um efeito incidental incompatível com as atuais evidências empíricas (CRYSTAL, 2013; PANOZ-BROWN *et al*., 2016). Ratos apresentam uma curva de aprendizado rápida e com razoável flexibilidade, e isso é observado especialmente em casos de variantes conservadoras das situações já conhecidas. Eles podem, por exemplo, compreender que um labirinto foi parcialmente rearticulado ou que, em certas condições, o conteúdo de uma parte do labirinto pode ser considerado irrelevante para determinar a provável presença de alimento num certo período do dia. Um conjunto não estruturado de *affordances* ou disposições implicaria uma trajetória de aprendizado mais lenta e, possivelmente, mais problemática. Em particular, ela seria sujeita a falhas comportamentais distintas das observadas sob a hipótese representacional, especialmente no caso de permutações mais complexas entre diferentes elementos da tríade objeto-local-tempo.

Temos então um cenário em que a pergunta "o que são representações?" é claramente distinta da pergunta "como elas são exploradas?". Representações são isomorfismos. Representações mentais são isomorfismos que ocorrem no interior do aparato cognitivo. A pergunta sobre como elas são exploradas é, portanto, uma pergunta sobre como determinados sistemas fazem uso delas para lidar com os alvos a que têm sensibilidade. Segue-se que não é preciso formular um conjunto de condições necessárias e suficientes para estabelecer quais mecanismos podem ou não ser considerados representacionais. O que se deve fazer é buscar os casos em que modelos envolvendo representações caracterizam a melhor explicação disponível face às evidências. E o que caracteriza a melhor explicação no caso de um mecanismo pode ser diferente em outro. Como se viu, ainda que possam haver casos em que dois ou mais modelos sejam capazes de explicar uma dada capacidade, eles podem fazê-lo por caminhos distintos, e efeitos incidentais podem ser utilizados para identificar qual hipótese é melhor acomodada pelos dados.

4. Representações: modo de usar

O próximo passo é argumentar que a atribuição de conteúdo às estruturas exploradas é capaz de cumprir um papel explicativo não trivial. Precisamos mostrar que representações fazem diferença na explicação de capacidades cognitivas e que, portanto, não estão apenas nos olhos de quem vê. Representações são reais e podem ter papel explicativo relevante porque permitem a formulação de teses empíricas indisponíveis na sua ausência. Essa dimensão explicativa adicional se caracteriza pela distinção entre erro

representacional e erro de processamento/exploração. Considere novamente um exemplo de como representações não mentais podem ser exploradas em nível agencial:

(a) Igor errou o caminho porque o mapa era excessivamente impreciso; e

(b) Igor errou o caminho porque usou o mapa de cabeça pra baixo.

A tese (a) caracteriza inacurácia no modo como a informação foi representada. É um exemplo de erro representacional. Já a tese (b) caracteriza um erro no modo como a informação foi explorada. Falhas mecânicas, erros de engajamento ou acoplamento, bem como condições aquém das ideais, são exemplos típicos de erros desse tipo. Erros representacionais independem da aptidão com que a informação é explorada ou de problemas no veículo (quando um mapa físico é danificado, por exemplo). O que os torna possíveis é a relação *direta* entre a estrutura a que a representação é isomórfica (conteúdo) e a estrutura a que ela é aplicada (alvo). Assim, eles desdobram toda uma classe de hipóteses empíricas: as que envolvem a articulação da diferença entre conteúdo e alvo. A tese de que Igor errou o caminho por imprecisão do mapa é distinta da tese de que ele o explorou mal. Ainda que resultem em comportamentos idênticos, elas têm efeitos incidentais distintos.

Outro sinal de que erros representacionais desdobram uma classe particular de hipóteses empíricas está na sua independência. Por si mesma, a imprecisão semântica não implica nem sugere insucesso funcional ou comportamental. Por isso, no exemplo (a), é necessário caracterizar o erro comportamental de Igor (errar o caminho) como fruto de imprecisão excessiva, isto é, de um grau de inacurácia prejudicial aos seus objetivos. Contudo, abrir mão de acurácia em virtude de tratabilidade é frequentemente vantajoso. Quando desenhamos um mapa simples, ou usamos um modelo simplificado de um domínio complexo para compreendê-lo melhor, buscamos nos poupar do esforço de lidar com grande quantidade de detalhes que ampliam a acurácia, mas que são irrelevantes para o resultado. A mesma estratégia está disponível para a natureza. Com recursos limitados, uma representação simplificada da dinâmica dos movimentos do predador num mecanismo sub-pessoal pode ser a chave para que a presa tenha tempo de escapar.

Para um exemplo concreto, podemos retomar os experimentos anteriormente citados envolvendo ratos em labirintos. Newen & Vosgerau (2020) argumentam que o modo flexível e rápido como os ratos aprendem é melhor explicado representacionalmente. Ou seja, a trajetória desse aprendizado é melhor caracterizada nos termos da correção da imprecisão representacional (i.e. redução da diferença entre conteúdo e alvo). O ponto,

contudo, não é o de que representações mentais sempre caracterizarão a melhor explicação. Trata-se de ressaltar que, na ausência de boas razões *a priori* para descartá-las, a questão é empírica. Determinar se o exercício de uma capacidade cognitiva envolve erros do tipo (a), do tipo (b) ou de alguma articulação mais complexa entre eles, requer avaliação caso a caso. Assim, o problema de *frameworks* que rejeitam representações mentais é que descartam ou distorcem uma classe de hipóteses empíricas por meios não empíricos. A empreitada científica acaba indevidamente constrangida, fazendo parecer que a única caracterização empiricamente plausível do comportamento é a de que se cometeu um erro ao explorar a informação disponível, i.e., um erro do tipo (b).

Contudo, antirrepresentacionalistas não são os únicos tentados a dizer que o que parecem erros do tipo (a) são, no fundo, erros do tipo (b). Vários autores representacionalistas acreditam (erroneamente) que esse é o único caminho plausível para uma teoria naturalista da representação. Entender onde essa tentação começa é importante para apreciar a robustez que a distinção entre conteúdo e alvo nos compra. Considere o que John Haugeland (1998, p. 309–10) nos diz acerca do alcance explicativo de normas biologicamente sedimentadas:

> […] there is another important distinction that biological norms do not enable. That is the distinction between functioning properly (under the proper conditions) as an information carrier and getting things right (objective correctness or truth), or, equivalently, between malfunctioning and getting things wrong (mistaking them). Since there is no other determinant or constraint on the information carried than whatever properly functioning carriers carry, when there is no malfunction, it's as "right" as it can be. In other words, there can be no biological basis for understanding a system as functioning properly, but nevertheless misinforming […].

Haugeland (1998) argumenta que normas biológicas não permitem falar de organismos que "erram", no sentido de que estão operando adequadamente, em condições ideais e, mesmo assim, operam com informações inacuradas. Ele demonstra com convicção de que erros representacionais só podem ser naturalizados por meio da redução a erros de outra ordem.[9] Novamente, a análise de Cummins (1996) nos permite enxergar

9 O texto de Haugeland (1998) foi escrito em um momento ainda dominado pelo adaptacionismo ancorado na síntese evolutiva moderna em biologia. A síntese evolutiva estendida fornece novas dimensões nas quais talvez possamos ancorar erros perceptuais. Para um exemplo, vide Carvalho & Rolla (2020). Contudo, importa notar que esse tipo de erro é diferente do que se busca articular aqui: ele não é semântico e não nos compra uma dimensão representacional não trivial. Nesse sentido, ele se insere no mesmo grupo das abordagens aqui tratadas, que buscam fundamentar o erro semântico em elementos não semânticos. Além

as razões dessa dificuldade com clareza. Comecemos com a hipótese mais simples, em que o conteúdo de uma representação é dado pelo alvo a que ela é aplicada. Esse é o caso das semânticas de papel conceitual ou inferencial, quando aplicadas à determinação de conteúdo representacional. Nelas, o conteúdo atribuído a uma certa representação é fruto direto do seu padrão de uso. Tais padrões determinam os alvos a que uma representação é aplicada pelo sistema e são determinados pelas relações mantidas com as demais representações que o sistema é capaz de instanciar. Assim, tanto o alvo quanto o conteúdo de um estado representacional são dados pelo modo como ele é usado. O conteúdo é, por definição, idêntico ao seu alvo.[10] Isso impossibilita o erro representacional, que se caracteriza justamente pela distinção entre a estrutura de um alvo e a estrutura utilizada como modelo daquele alvo.

Para escapar dessa limitação, é preciso criar um vão entre representação e alvo. Isso é exatamente o que a distinção de Cummins (1996) nos compra. Mas se a ignorarmos, como fazem os teóricos mais conhecidos, seremos forçados a ancorar esse vão em algum elemento não semântico. Como Perlman (2002) sintetizou, esses elementos adicionais buscam dividir as aplicações de representações em duas classes distintas: as que fixam conteúdo, e as que aplicam conteúdo previamente fixado, sendo as últimas sujeitas a erro. Como exemplo, considere a abordagem teleológica de Millikan (1987). Ela busca acomodar a dimensão do erro representacional na adaptatividade. O que distingue aplicações corretas e incorretas é o papel adaptativo do conteúdo na história evolutiva da espécie. Segue-se que toda representação responsável por estabelecer comportamento que se mostrou adaptativo é considerada correta. Em outras palavras, a adaptatividade é responsável por distinguir os casos de uso que fixam conteúdo dos casos de uso que aplicam o conteúdo fixado, e são os últimos que acomodam a possibilidade de erro. Mantém-se, portanto, a doutrina de que o conteúdo representacional de um determinado estado é fruto do uso que é feito dele no interior do sistema.

O resultado não é um erro representacional, mas a tentativa – análoga à do antirrepresentacionalista – de reduzi-lo a outro tipo de erro. Millikan (1987), por exemplo, fundamenta a correção semântica na correção comportamental (casos que geraram comportamento adaptativo). Mas como vimos na discussão anterior, sucesso semântico, funcional e comportamental são independentes. Se aceitarmos que a dimensão semântica

disso, vale notar que o cognitivismo é compatível com a síntese evolutiva estendida, conforme demonstra o projeto de Heyes (2018).

10 Note que essas considerações dizem respeito ao uso de semânticas de papel conceitual como teorias do conteúdo representacional. Não está em discussão o uso desse tipo de teoria na determinação de atitudes.

é parasitária das demais, então seremos forçados a concordar com Haugeland (1998): a biologia não dá espaço para mecanismos que estejam exibindo comportamento apto e que não obstante operem com informações inacuradas. Afinal, só haverá erro semântico se houver também erro funcional ou comportamental. Esse é um problema grave, pois o objetivo do representacionalista é explicar performances funcionais e comportamentais com a ajuda (ainda que não exclusiva) de conteúdo representacional. Mas ele acaba invertendo a ordem e explicando a atribuição de conteúdo a partir das performances funcionais e comportamentais.[11]. Não admira o antirrepresentacionalista sentir-se à vontade para perguntar: se o erro funcional e comportamental já estão explicados, por que acrescentar representações a essa história? Em síntese: é a ausência de uma noção de erro puramente representacional que abre o flanco para que o papel explicativo da dimensão representacional seja trivializado.

Para resistir ao desafio antirrepresentacionalista, portanto, é preciso mostrar como a dimensão representacional pode ser estabelecida sem parasitar as demais. Felizmente, já temos tudo o que precisamos. Dado o modo independente como conteúdo representacional e alvo são determinados, segue-se que eles podem divergir. A pressão que nos faz tentar ancorar erro semântico em erros funcionais ou comportamentais sequer emerge, pois não é preciso fabricar um vão entre usos que fixam e usos que apenas exploram conteúdos representacionais. Assim, o vão entre alvo e conteúdo consegue cumprir a demanda de acomodar uma dimensão explicativa (i.e., um espaço para formulação de teses empíricas), indisponível para abordagens não representacionais. *Pace* Haugeland (1998), normas biológicas podem, sim, ancorar o tipo de erro em que um mecanismo (ou um agente) pode atuar em condições ideais, sem problemas de funcionamento e, mesmo assim, estar na posse de informações erradas.

Nesse cenário, ainda é possível que todas as teses empíricas envolvendo representações sejam falsas. Porém, superados os problemas de naturalização, não se pode mais rejeitá-las por princípio, pois sua presença e efeitos são agora objeto de investigação empírica. Nos casos em que elas participarem da melhor explicação disponível face às evidências, não haverá razão para prescindir delas. Por isso é importante que a *framework* adotada permita a articulação e a posterior verificação de hipóteses dessa natureza.[12] Seria

11 Note como a crítica é também aplicável a outras abordagens, como à de Dretske (1981; 1986)

12 Convém notar que este é um caso em que o princípio da parcimônia pode levar a engano. A melhor explicação para uma capacidade simples tomada isoladamente pode levar a problemas quando ela é articulada junto a capacidades mais complexas ou demandantes de recursos explicativos adicionais. É preciso encontrar o modo mais parcimonioso de acomodar a participação de certa capacidade em todo o conjunto

inadequado permitir que todo um conjunto de possibilidades empíricas seja descartado apenas em função da preferência por uma abordagem ou outra. É melhor que a discussão se dê caso a caso, mecanismo a mecanismo.

5. Revoluções incompletas

Nas ciências cognitivas, a ausência de unidade sempre foi a regra. Não apenas existem *frameworks* diversas, como cada uma delas pode operar com diferentes articulações de arquiteturas, modelos e princípios. Não falta quem busque formular princípios unificadores, mas o risco de distorção das evidências disponíveis não deve ser minimizado. A insistência em boas práticas científicas que persigam diferentes estratégias parece ser o melhor remédio. Foi assim que aprendemos, paulatinamente, que o campo das ciências cognitivas é fértil em revoluções incompletas, isto é, ideias que prometiam uma reviravolta radical no modo como compreendemos a mente, mas que sob escrutínio revelaram-se tão somente reformistas.

Considere, por exemplo, os desdobramentos do processamento preditivo (CLARK, 2016). Essa abordagem concebe o aparato cognitivo como sendo essencialmente bayesiano e advoga extenso uso de modelos neurais generativos (HINTON, 2007) como a ferramenta ideal para modelar essas capacidades. Tais modelos são frequentemente anunciados como capazes de contar uma história radicalmente nova sobre a cognição. História essa que promete unificar as explicações dos mecanismos envolvidos na ação, percepção e raciocínio. Contudo, modelos generativos compreendem apenas uma das formas possíveis de se modelar sistemas preditivos. Em um trabalho provocativo, Cao (2020) argumenta que modelos generativos e não generativos são matematicamente equivalentes (ao menos em certos casos centrais, como o da percepção), e isso a ponto de não haver efeitos incidentais claros que permitam decidir entre as duas alternativas. Além disso, há também argumentos plausíveis sugerindo que o tipo de inferência sobre o qual sistemas preditivos repousam podem ser modelados em arquiteturas clássicas (PIANTADOSI; JACOBS, 2016). Não bastasse, há também quem aponte equivalências entre os resultados obtidos pela explicação de certas capacidades a partir de modelos generativos e modelos dinâmicos (ANDERSEN; MILLER; VERVAEKE, 2022).

---

de capacidades cognitivas. Isso pode levar à escolha de teses que demandem mais recursos, mesmo no caso de capacidades relativamente simples, a exemplo do que ocorre no exemplo dos ratos.

Com efeito, temos um cenário em que se torna paulatinamente mais difícil distinguir o que há de realmente inovador no processamento preditivo, ou o que exatamente a adoção de modelos generativos nos compra que já não pudéssemos comprar com outras arquiteturas. Não se busca aqui sugerir que a resposta a essas perguntas é negativa ou que o processamento preditivo caminha inevitavelmente rumo a uma dissolução em abordagens já existentes. Antes, o objetivo é ilustrar o tipo de escrutínio e dialética a que boas ideias costumam ser submetidas no âmbito das ciências cognitivas.

Assim, ainda que a revolução prometida se mostre incompleta, esse processo pode resultar em alterações relevantes no modo como as ciências envolvidas são praticadas, e estas tendem a ser assimiladas. O resultado são *frameworks* mais ricas em recursos cognitivos à disposição das práticas científicas. Um exemplo promissor pode ser encontrado nos trabalhos de Piccinini (2021; 2022). Sua abordagem acomoda todas as características associadas à cognição 4E. Não obstante, em linha similar à adotada por Clark (1997), ele defende a tese de que o caráter situado da cognição não acarreta incompatibilidade com a existência de mecanismos computacionais ou representacionais. Processos computacionais, inclusive os que exploram conteúdo representacional, são eles mesmos situados.

Dizer de um estado representacional que ele é situado significa dizer, entre outras coisas, que a determinação do seu conteúdo não implica a necessidade de representações adicionais. Em abordagens clássicas, representar a posição de um copo sobre a mesa implicava representar a mesa, o copo, quaisquer demais objetos presentes e, não bastasse, as devidas relações entre eles. Caso contrário, não seria possível atribuir ao estado representacional o papel de explicar como o organismo é capaz de antecipar possíveis desdobramentos ("se o copo for tocado de modo indevido, ele pode cair"). Em contraste, o conteúdo de representações situadas pode ser tal que leve em conta aspectos do corpo e do ambiente, resultando em conteúdos como "o copo pode ser alcançado com um leve girar da mão direita no sentido horário".

Note-se que tal conteúdo só faz sentido para um dado organismo, com um dado corpo num dado estado e no interior de uma situação particular. Assim, o papel explicativo que esse conteúdo cumprirá na explicação de uma determinada capacidade é determinado de um modo situado. Ao contrário do que alguns sugerem, portanto, a ideia de representações situadas não mitiga o papel explicativo que representações agregam. Elas dizem respeito tão somente ao modo como representações são ou podem ser exploradas no interior de um dado sistema.

Na mesma linha, os mecanismos computacionais que encontramos no sistema neural estão longe de ser os mecanismos passivos que apenas reagem a *inputs* do ambiente a partir de seus estados internos. Pelo contrário, esses mecanismos se engajam em processos de aprendizagem ativa, acumulando informações de múltiplas fontes, tanto internas ao aparato neural (frequência e tempo de disparos neuronais, por exemplo) quanto externas (sinais visuais, auditivos, olfativos, etc.) Esse tipo de processo não apenas é compatível, mas dependente de o aparato cognitivo ser incorporado e integrado ao ambiente. Isso se dá não no sentido clássico segundo o qual corpo e ambiente são fontes de *inputs* e espaços para despejo de *outputs*, mas no sentido de que sistemas cognitivos exploram acoplamentos contínuos em tempo real.

Essa integração já acomoda uma dimensão de caráter enativista, pois ela permite entender como um organismo pode compreender o ambiente ao seu redor a partir do modo como o explora. Processos de aprendizagem ativa influenciam o corpo e ambiente ao mesmo tempo em que ambiente e corpo afetam o modo como os processos cognitivos se desenrolam. Por fim, processos de aprendizagem também podem envolver elementos afetivos, na medida em que o organismo dispõe de objetivos ou expectativas, certos *inputs* e/ou acoplamentos com o ambiente irão importar mais do que outros.

É importante salientar, contudo, que Piccinini (2021; 2022) não adota a teoria do conteúdo representacional de Cummins (1996). Ele endossa o uso de representações estruturais ancoradas no isomorfismo, e por isso sua abordagem também evita o problema duro do conteúdo. Contudo, Piccinini (2021; 2022) adota uma variante da teleosemântica que confunde fixação de alvos e determinação de conteúdo representacional.

Com efeito, seu tratamento do conteúdo representacional sofre dos problemas tratados nas seções anteriores. Em particular, o tratamento que Piccinini (2021; 2022) dá ao erro representacional ancora o que deveria ser um erro semântico em elementos não semânticos, tais como condições ambientais aquém das ideais ou mau funcionamento dos mecanismos envolvidos. Como vimos, isso torna sua abordagem mais vulnerável a argumentos pela trivialização do papel explicativo das representações. Felizmente, é possível ler o trabalho de Piccinini (2021; 2022) como tão somente uma descrição das estratégias que organismos utilizam para fixar alvos. Como a fixação de alvos é independente da determinação de conteúdo representacional, isso permite rejeitar a teoria do conteúdo representacional que Piccinini (2021; 2022) endossa em virtude da teoria de Cummins (1996) sem abrir mão daquilo que o projeto de Piccinini tem de melhor. O resultado desse amálgama é um bom exemplo do tipo de projeto que consideramos promissor: o que almeja acomodar as virtudes da cognição 4E e do cognitivismo sem carregar consigo seus vícios.

Evidentemente, é possível que essa compatibilidade repouse sobre diferentes concepções de cognição situada. Ser incorporado, integrado ou enativo pode significar coisas distintas. Mas ainda que existam, as consequências dessas divergências podem ser insuficientes para justificar uma clivagem. A distinção que Cummins (1996) faz entre alvo e conteúdo pode ser (novamente) útil para tornar esse ponto mais saliente: parte considerável do contraste entre representacionalistas e não representacionalistas sequer alcança a dimensão semântica.

Nesse sentido, o contraste é melhor compreendido como enfatizando diferentes modos pelos quais organismos são capazes de fixar em alvos, isto é., de exibirem alguma forma de intencionalidade. Se representações carregassem consigo o compromisso com abordagens não situadas da cognição, então sem dúvida o modo como sistemas representacionais conseguem fixar alvos (seja para representá-los, seja para explorá-los por meio de recursos não representacionais), seria incompatível. Essa restrição não se dá, contudo, sobre uma teoria da determinação do conteúdo representacional. Ela se dá sobre teses que associam a presença de representações ao tipo de mecanismo que as explora, uma confusão que Cummins (1996) nos ajuda a enxergar. Uma vez que sabemos a diferença entre o que é representar e o que é explorar, desdobra-se um cenário em que a revolução antirrepresentacionalista não é tão revolucionária assim, pois não implica (ainda que deseje) um completo abandono de mecanismos representacionais, mas tão somente de sua versão não situada. O poder explicativo que as *frameworks* tradicionais encontraram no conceito de representação não precisa ser descartado.

Conclusão

Neste artigo procuramos mostrar que há (pelo menos) uma noção de representação que não está atrelada ao "cognitivismo de velha guarda" e não implica o abandono de uma concepção da mente como *corporific ada*, *situada* e *enativa*. Procuramos também esclarecer que tipo de coisa pode ser uma representação e que papeis teóricos ela visa desempenhar, desfazendo os principais mal-entendidos associados a esta noção na filosofia da mente contemporânea. Finalmente, procuramos mostrar em que sentido a introdução de representações pode nos comprar uma dimensão explicativa extra que não está disponível fora deste vocabulário teórico. Se esta dimensão será de fato útil, é algo que deverá ser estabelecido empiricamente através de uma análise caso a caso, não podendo ser descartada *a priori* apenas em virtude das inclinações teóricas do pesquisador.

Nosso esforço nessa direção, mais uma vez, não é motivado por uma reabilitação do representacionalismo ou uma crítica ao enativismo. Nosso objetivo é apenas aliviar algumas preocupações dos enativistas acerca dos comprometimentos ontológicos e metodológicos associados ao vocabulário representacional, para que possam enfim reconhecer como possíveis aliados aqueles teóricos que também buscam investigar o caráter corporificado e situado da mente, ainda que o façam através de (uma certa noção de) representações mentais. Caso esse objetivo seja ao menos parcialmente realizado, poderemos ter a nosso dispor mais uma ferramenta teórica compartilhada, dissociada de uma vez por todas de suas origens computacionalistas e representacionalistas.

Referências

ANDERSEN, B. P.; MILLER, M.; VERVAEKE, J. Predictive processing and relevance realization: exploring convergent solutions to the frame problem. **Phenomenology and the Cognitive Sciences**, ago. 2022.

CAO, R. New Labels for Old Ideas: Predictive Processing and the Interpretation of Neural Signals. **Review of Philosophy and Psychology**, v. 11, n. 3, p. 517–46, ago. 2020.

CARVALHO, E. M. De; ROLLA, G. An Enactive-Ecological Approach to Information and Uncertainty. **Frontiers in Psychology**, v. 11, abr. 2020.

CLARK, A. **Being There:** Putting Brain, Body, and World Together Again. Cambridge: MIT Press, 1997.

CLARK, A. **Surfing Uncertainty**. Oxford: Oxford University Press, 2016.

CRYSTAL, J. D. Remembering the past and planning for the future in rats. **Behavioural Processes**, v. 93, p. 39–49, fev. 2013.

CUMMINS, R. **Representations, targets and attitudes**. MIT Press, 1996.

CUMMINS, R. Reply to Millikan. **Philosophy and Phenomenological Research**, v. 60, n. 1, p. 113, jan. 2000.

CUMMINS, R. et al. Representation and unexploited content. In: **The World in the Head**. Oxford, 2010. p. 120–33.

CUMMINS, R. How does it work versus "what are the laws?": Two conceptions of psychological explanation. In: **The world in the head**. Oxford, 2010. p. 283–310.

DRETSKE, F. **Knowledge and the flow of information**. Cambridge: MIT Press, 1981.

DRETSKE, F. I. Misrepresentation. In: BOGDAN, Radu (Ed.). **Belief: Form, Content, and Function**. Oxford University Press, 1986. p. 17–36.

EGAN, F. A Deflationary Account of Mental Representation. In: SMORTCHKOVA, Joulia; DOLEGA, Krzysztof; SCHLICHT, Tobias (Eds.). **Mental Representations**. New York, USA: Oxford University Press, 2020.

FACCHIN, M. Structural representations do not meet the job description challenge. **Synthese**, v. 199, n. 3-4, p. 5479–5508, jan. 2021.

FODOR, J. A. **The Language of Thought**. Harvard University Press, 1980.

HAUGELAND, J. Truth and rule-following. In: **Having thought**. Cambridge: Harvard University Press, 1998. p. 305–61.

HEYES, C. **Cognitive Gadgets**. Harvard University Press, 2018.

HINTON, G. E. Learning multiple layers of representation. **Trends in Cognitive Sciences**, v. 11, n. 10, p. 428–34, out. 2007.

HURLEY, S. Perception and action: alternative views. **Synthese**, v. 129, n. 1, p. 3–40, 2001.

HUTTO, D.; MYIN, E. **Radicalizing enactivism:** basic minds without content. Cambridge, Mass: MIT Press, 2013.

MILLIKAN, R. G. **Language, Thought, and Other Biological Categories:** New Foundations for Realism. The MIT Press, 1987.

NEWEN, A.; VOSGERAU, G. Situated mental representations: why we need mental representations and how we should understand them. In: SMORTCHKOVA, Joulia; DOLEGA, Krzysztof; SCHLICHT, Tobias (Eds.). **What are mental representations?** Oxford University Press, 2020. p. 178–212.

NOË, A. **Action in Perception**. Cambridge: MIT Press, 2004.

PANOZ-BROWN, D. et al. Rats Remember Items in Context Using Episodic Memory. **Current Biology**, v. 26, n. 20, p. 2821–826, out. 2016.

PERLMAN, M. Pagan teleology: adaptational role and the philosophy of mind. In: ARIEW, André; CUMMINS, Robert; PERLMAN, Mark (Eds.). **Functions:** new essays in the philosophy of psychology and biology. Oxford University Press, 2002. p. 263–90.

PIANTADOSI, S. T.; JACOBS, R. A. Four Problems Solved by the Probabilistic Language

of Thought. **Current Directions in Psychological Science**, v. 25, n. 1, p. 54–9, fev. 2016.

PICCININI, G. **Neurocognitive Mechanisms:** Explaining Biological Cognition. Oxford University Press, 2021.

PICCININI, G. Situated Neural Representations: Solving the Problems of Content. **Frontiers in Neurorobotics**, v. 16, abr. 2022.

RAMSEY, W. **Representation reconsidered**. Cambridge: Cambridge University Press, 2007.

ROLLA, G. **A mente enativa**. Porto Alegre: Editora Fi, 2021.

ROLLA, G. Por que não somos só o nosso cérebro: em defesa do enativismo. **TRANS/FORM/AÇÃO: Revista de Filosofia**, n. 46, p. 207–36, 2023.

SHEA, N. **Representation in Cognitive Science**. Oxford University Press, 2018.

STICH, S. **From folk-psychology to cognitive science:** the case against belief. MIT Press, 1983.

SWOYER, C. Structural representation and surrogative reasoning. **Synthese**, v. 87, n. 3, p. 449–508, jun. 1991.

VARELA, F. J.; ROSCH, E.; THOMPSON, E. T. **The Embodied Mind:** cognitive science and human experience. MIT Press, 1991.

DEBATES CONTEMPORÂNEOS EM FILOSOFIA DA MEMÓRIA: UMA BREVE INTRODUÇÃO

César Schirmer dos Santos (MemLab - Universidade Federal de Santa Maria)

André Sant'Anna (Department of Philosophy – University of Geneva[1])

Kourken Michaelian (Centre for Philosophy of Memory - Université Grenoble Alpes)

James Openshaw (Centre for Philosophy of Memory - Université Grenoble Alpes)

Denis Perrin (Centre for Philosophy of Memory - Université Grenoble Alpes)

Resumo:

Neste artigo apresentamos, de forma concisa e em português, alguns elementos-chave dos principais debates contemporâneos na filosofia da memória. Nosso principal objetivo é tornar essas discussões mais acessíveis aos leitores de língua portuguesa, fornecendo uma atualização importante para esforços anteriores (SANT'ANNA & MICHAELIAN, 2019a). Começamos introduzindo a noção de viagem no tempo mental, a qual estabelece a base empírica para a metodologia empregada em trabalhos recentes, antes de apresentar dois debates centrais. Primeiro, o debate entre causalistas e simulacionistas sobre a teoria da lembrança filosofi-

Abstract:

In this article we present, concisely and in Portuguese, some key elements of the main contemporary debates in the philosophy of memory. Our principal aim is to make these discussions more accessible to Portuguese-speaking readers, providing an important update to previous such efforts (SANT'ANNA & MICHAELIAN, 2019a). We begin by introducing the notion of mental time travel which lays the empirical basis for the methodology employed in recent work, before presenting two core debates. First, the debate between causalists and simulationists about the correct philosophical theory of remembering (§2). Second, the debate between

1 Grande parte do trabalho realizado neste artigo por André Sant'Anna ocorreu em seu período como Pesquisador de Pós-Doutorado afiliado ao *Cologne Center for Contemporary Epistemology and the Kantian Tradition* (CONCEPT), Universität zu Köln e financiado pela Alexander von Humboldt-Stiftung no programa Humboldt-Forschungsstipendium.

camente correta (§2). Segundo, o debate entre continuístas e descontinuístas sobre a relação entre lembrança episódica e certas formas de imaginação (em particular, o pensamento episódico acerca do futuro) (§3). Na segunda parte do artigo, apresentamos e exploramos dois tópicos de discussão crescente: a natureza do sentimento de passado característico da memória episódica (§4) e questões de meta-nível relativas ao próprio caráter das controvérsias exploradas na primeira parte do artigo (§5).

Palavras-Chave:

memória; memória episódica; viagem no tempo mental; causalismo e simulacionismo; mnemicidade; sentimento de passado.

continuists and discontinuists concerning the relationship between remembering and certain forms of imagining (in particular, episodic future thought) (§3). In the second part of the paper, we introduce and explore two growing topics of discussion: the nature of the feeling of pastness distinctive of episodic memory (§4), and meta-level issues concerning the very character of controversies explored in the first part of the paper (§5).

Keywords:

memory; episodic memory; mental time

1. Introdução

A filosofia da memória é uma nova (mas já bastante rica) área de investigação. É difícil apontar para um único fator que tenha disparado a explosão de estudos filosóficos contemporâneos sobre a natureza da memória, mas é incontroverso que as recentes investigações empíricas sobre a capacidade de viajar no tempo mental (ver, entre muitos outros, TULVING 1985; SUDDENDORF E CORBALLIS 1997; ADDIS *ET AL.* 2007; DE BRIGARD 2014; MICHAELIAN 2016A; ADDIS 2020; MICHAELIAN *et al.* 2022) sopraram novos ventos na filosofia, seja na maneira de se recortar naturalisticamente o fenômeno do recordar, seja no que diz respeito a métodos de investigação oriundos das ciências empíricas.

A pesquisa contemporânea em filosofia da memória deve muito à pesquisa sobre a capacidade humana de se projetar, imaginativamente, para situações do passado e do futuro pessoal (ver DE BRIGARD 2014; MICHAELIAN 2016a). Esta projeção de si mesmo para o passado real é o que se entende por *lembrança episódica*. Mas também há projeção imaginativa de si mesmo para o futuro (SUDDENDORF & CORBALLIS 2007) – o que podemos chamar de *imaginação episódica orientada ao futuro* – e para situações passadas

que não aconteceram, mas poderiam ter acontecido – fenômeno conhecido como *pensamento episódico contrafactual* (De BRIGARD & PARIKH 2019). Em todos esses casos, a capacidade de viajar no tempo mental nos permite simular, em nossas mentes, eventos que não estão ocorrendo no mundo exterior e que, por isso, não podem ser percebidos.

A filosofia da memória contemporânea é diretamente motivada por estudos empíricos sobre tais tipos de casos. Mais precisamente, uma nova visão sobre a relação entre memória e imaginação se fortaleceu a partir de resultados surpreendentes de estudos de neuro-imagem (ver OKUDA *et al.* 2003; ADDIS *et al.* 2007; HASSABIS & MAGUIRE, 2007). Que sejamos capazes de lembrar não é, por si só, curioso. Mas é surpreendente que haja, no nível da realização neural, como se descobriu empiricamente, semelhanças processuais notáveis entre o lembrar episódico e variedades de imaginação nas quais o sujeito se projeta seja para uma situação futura (antecipação ou prospecção episódica), seja para uma situação passada que não se deu, mas poderia ter acontecido (imaginação contrafactual) (ver BENOIT & SCHACTER 2015; ADDIS 2020).

Tradicionalmente, a memória foi vista como um sistema cuja função primária seria evocar o passado acuradamente. No entanto, a pesquisa empírica deu subsídios para a hipótese de que a função primária da memória episódica seria simular eventos futuros (ver ADDIS *et al.* 2007, p. 1374–5; SUDDENDORF & CORBALLIS, 2007). Isso porque, segundo a pesquisa empírica, a memória episódica não é um sistema que reproduz passivamente os eventos vivenciados no passado, mas, sim, um sistema ativo que usa elementos de experiências passadas para simular e antever futuros possíveis. Assim sendo, algumas vulnerabilidades da memória, como a suscetibilidade a distorções, não seriam defeitos, mas efeitos colaterais do fato de que o sistema de memória episódica tem, como função central, a simulação de eventos futuros (ver SCHACTER *et al.*, 2012).

Filosoficamente, a similaridade (no nível dos processos neurais) entre lembrança episódica e simulação do futuro abre espaço para que revisitemos uma questão clássica: será que lembrar é imaginar? A investigação empírica fortalece a visão da memória como uma capacidade construtiva (ver BARTLETT, 1932), e traz dificuldades para propostas que entendem o lembrar como uma maneira de preservar, fiel e passivamente, o que foi percebido no passado. Considerando a pesquisa empírica, a melhor hipótese é que o sistema de memória episódica não opera como um mecanismo que gera um *replay* de uma percepção passada, mas, sim, como um sistema que (re)constrói ativamente as lembranças (ver MICHAELIAN, 2016a; LANGLAND-HASSAN, 2023b).

Neste processo de construção, o sistema de memória episódica (re)combina elementos de experiências passadas de modo a formar seja uma lembrança coerente, seja um cenário futuro plausível (ver De BRIGARD, 2014). Na construção de cenários futuros, o sistema de memória episódica opera flexivelmente, tanto que é capaz de gerar simulações de eventos que nunca ocorreram (ver ADDIS *et al.*, 2007). Por exemplo, você pode simular na sua mente como seria apresentar um trabalho numa conferência, embora nunca tenha feito isso antes. A mesma flexibilidade, no entanto, também se manifesta na construção de lembranças episódicas. Assim, a reflexão empírica sobre a arquitetura da memória abre espaço para a conclusão (abdutiva) que talvez aquilo que lembramos não seja tão distinto do que imaginamos. A viagem no tempo mental para o passado pode ser um filme tão editado quanto nossas fantasias sobre o futuro.

A pesquisa sobre a capacidade de viajar no tempo mental se entrelaça à pesquisa sobre a memória episódica, a qual se origina da proposta de Tulving (1972) de distinguir as lembranças de eventos das lembranças de fatos. Lembranças de eventos são chamadas de *lembranças episódicas* (*episodic memories*), e lembranças de fatos são chamadas de *lembranças semânticas* (*semantic memories*)[2]. Por envolverem eventos, em vez de fatos, as lembranças episódicas se distinguem das lembranças semânticas. Considere sua lembrança daquela festa de aniversário do seu amigo. Você lembra do que aconteceu (a comemoração), de quando aconteceu (no dia tal), e onde aconteceu (na casa de um amigo). Esta é uma lembrança episódica, pois faz referência a um único evento. Por envolver representações, as lembranças episódicas se distinguem dos hábitos, os quais não envolvem a geração de representações. No entanto, o conceito inicial de lembrança episódica se mostrou problemático, visto que a distinção entre eventos e fatos é porosa. É um fato que 2+2=4, mas não é um evento que 2+2=4, pois operações aritméticas não acontecem no tempo e no espaço. Mas é um fato que seu amigo comemorou o aniversário na casa dele em tal dia, e isso também é um evento – o que é um problema para a definição inicial de memória episódica.

Dada essa situação, para que a noção de memória episódica se mostrasse valiosa, seria preciso encontrar uma base mais apropriada para sua distinção em relação a outros tipos de lembrança. É aqui que entra a noção de viagem no tempo mental. A partir de

2 Aqui estamos seguindo a distinção, cada vez mais frequente, entre o ato de lembrar e o sistema neurocognitivo que gera lembranças. O primeiro, o ato, tem sido chamado de "lembrança" *(remembering)*, enquanto o segundo tem sido chamado de "memória" *(memory)*. Ver, por exemplo, Mahr (2023). Esta não é uma distinção padrão, haja vista que os termos "lembrança" e "imaginação" são usados das mais diversas maneiras. Ainda assim, é uma distinção que se mostra esclarecedora no contexto deste artigo.

evidências relacionadas ao comportamento de certa população de amnésicos (TULVING, 1985) e de visões sobre a arquitetura cognitiva (SUDDENDORF & CORBALLIS, 1997), pareceu importante redefinir a lembrança episódica como uma forma de viagem no tempo mental. Isso porque, ao lembrar episodicamente, o sujeito tem a experiência de "reviver" algo que experienciou no passado (ver JAMES, 1890). Tal tipo de experiência não ocorre em casos de lembrança semântica, que envolvem somente consciência de fatos. As evidências empíricas indicavam que esta capacidade de viajar no tempo mental (ou sua ausência, no caso de certos amnésicos) se liga intimamente à capacidade de se projetar imaginativamente ao futuro (ver ADDIS *et al*., 2007). Indicavam também que seria plausível entender a arquitetura cognitiva como sendo tal que sua função fosse, prioritariamente, levar o sujeito a se imaginar no futuro. Desse modo, o lembrar episódico seria uma função secundária do sistema que flexivelmente gera viagens no tempo mental (ver SUDDENDORF & CORBALLIS, 1997; De BRIGARD, 2014; e MICHAELIAN, 2016a; para visões críticas, ver MAHR & CSIBRA, 2018; BOYLE, 2022; MAHR, 2023; ROBINS, 2023).

A partir desta moldura teórica, este artigo apresenta, de forma concisa e em língua portuguesa, alguns dos principais debates contemporâneos em filosofia da memória. Não partiremos do zero, contudo. Em vez disso, vamos expor alguns elementos-chave dos debates contemporâneos em filosofia da memória a partir do trabalho que publicamos, em língua portuguesa, na revista *Voluntas* (SANT'ANNA & MICHAELIAN, 2019a). Tomando-o por referência, no presente texto, propomo-nos a atualizar e aprofundar a apresentação de questões e discussões fundamentais da filosofia da memória contemporânea. Todavia, devido a restrições de espaço, selecionamos para análise uma pequena amostra de debates proeminentes. Nosso objetivo principal é tornar essas discussões mais acessíveis ao leitor de língua portuguesa interessado na filosofia da memória. Tendo em vista este objetivo, apresentamos nesta seção a noção de viagem no tempo mental, a qual fornece um pano de fundo empírico que é fundamental para a compreensão dos métodos empregados pelos filósofos da memória. Em seguida, nas próximas seções, sem a pretensão de exaustividade, apresentaremos outros debates importantes para uma introdução ao tema.

Assim, na seção 2, voltaremos a um tema abordado por Sant'Anna & Michaelian (2019a), mas que precisa ser atualizado, a saber, o debate entre causalistas e simulacionistas sobre a natureza do lembrar episódico (ver, entre muitos outros, DEBUS, 2014; De BRIGARD, 2014; MICHAELIAN, 2016a; BERNECKER, 2017a). Trata-se de um debate maduro, cujos meandros e resultados explicam porque os outros debates se dão da maneira em que se dão. Por ser um debate já consolidado, esta será a seção mais longa deste

artigo. Esperamos que se compreenda esse desbalanço proporcional com as demais seções do texto.

Na terceira seção, cobriremos um tema que não foi abordado diretamente em Sant'Anna & Michaelian (2019a), qual seja, o debate entre continuístas e descontinuístas sobre a relação entre memória e imaginação (ver, entre muitos outros, PERRIN, 2016; MICHAELIAN, 2016B; ROBINS, 2020), também bastante consolidado entre especialistas. Com isso, esperamos fornecer mais subsídios ao leitor, em língua portuguesa, para as discussões sobre a questão da mnemicidade, isto é, sobre a relação entre o lembrar e o imaginar (ver MICHAELIAN & SUTTON, 2017).

Por fim, na quarta seção, apresentaremos dois debates ainda pouco desenvolvidos, embora importantes para orientar as futuras direções da filosofia da memória. Nela abordaremos as discussões sobre a natureza do sentimento de passado que é típico das lembranças episódicas (ver FERNÁNDEZ, 2019; PERRIN, MICHAELIAN & SANT'ANNA, 2020; PERRIN & SANT'ANNA, 2022). Na quinta e última, trataremos dos meta-debates sobre a própria natureza das controvérsias na filosofia da memória (CRAVER, 2020; MCCARROLL, MICHAELIAN & NANAY, 2022; SCHIRMER DOS SANTOS, MCCARROLL & SANT'ANNA, 2023).

2. Causalismo e simulacionismo

Em Sant'Anna & Michaelian (2019a), apresentamos ao público de língua portuguesa a teoria causal da memória e a teoria da memória como simulação (simulacionismo). No entanto, o debate se manteve vivo desde aquele artigo, e há tópicos relacionados a essa contenda entre os defensores dessas duas teorias que não foram cobertos na publicação. Como se trata de um debate-chave em filosofia da memória, aproveitamos esta oportunidade para aprofundar a apresentação dessa importante discussão ao público lusófono.

2.1 O causalismo: panorama

Na segunda metade dos anos 1960, Martin & Deutscher (1966) propuseram uma teoria *causal* da memória. Várias outras teorias do mesmo estilo foram desenvolvidas na mesma época. Por exemplo, Grice (1961) propôs uma teoria causal da percepção, e Goldman (1967) elaborou uma teoria causal do conhecimento. Na filosofia, uma teoria causal de algum tipo de coisa estabelece que a existência, as propriedades ou a identidade deste

tipo de coisa são determinadas por certas relações causais. Por exemplo, teorias causais do conhecimento estabelecem que um sujeito conhece um fato se sua crença foi apropriadamente causada pelo fato. Neste caso, o que faz com que a crença que p seja um caso de conhecimento é "uma conexão causal entre o fato que torna p verdadeira […] e a crença de [S] que p" (GOLDMAN 1967, p. 358; nossa tradução).

De maneira semelhante, numa teoria causal da memória, S lembra de um evento E não por simplesmente estar visualizando E na sua mente, mas porque a representação mental de E foi causada da maneira apropriada pelo próprio evento E. "Da maneira apropriada", no caso das lembranças, significa causação através de traços de memória que foram codificados na ocasião da experiência de E e são evocados na ocasião da representação mental de E. Esses exemplos mostram que teorias causais não consideram que uma mera representação mental (uma crença no caso do conhecimento, uma imagem mental no caso da memória) seja suficiente para estabelecer a ocorrência de um certo tipo de cognição (conhecimento e lembrança, respectivamente), pois também é preciso que a representação mental tenha sido causada da maneira requerida.

Em resumo, há teorias causais que se aplicam a diversos domínios das representações mentais, como a percepção, o conhecimento e a memória, sendo que em cada domínio certos tipos de relações causais (apropriadas) são propostas como explicadores dos assuntos de cada domínio. Estes explicadores, as relações causais, são, em cada caso, condições necessárias para que o tipo de representação mental que está em foco seja explicado. Juntamente com outras condições propostas por cada teoria, obtêm-se um conjunto de condições suficientes para que o fenômeno mental em tela em cada caso seja explicado.

Quando falamos numa teoria causal da memória, estamos falando de uma *metafísica* da memória, isto é, de uma teoria que busca estabelecer o que é a memória, o que, na abordagem de Martin & Deutscher (1966), acarreta buscar as condições necessárias (a essência, alguns diriam) do lembrar[3]. Uma metafísica da memória tem amplos usos, sendo o mais importante, contemporaneamente, o de responder a questões de demarcação sobre as fronteiras da memória. Vários tipos de estados cognitivos "fazem fronteira" com o lembrar. Sobre a demarcação da memória, podemos perguntar, por exemplo: Qual a diferença entre lembrar e perceber? Qual a diferença entre lembrar e crer que algo aconteceu? Qual

3 É preciso ter claro o tipo de teoria filosófica proposta por Martin & Deutscher (1966). A filosofia da memória envolve uma variedade de teorias, cada uma com sua função específica. Por exemplo, uma epistemologia da memória descreve e explica as propriedades das crenças relacionadas ao lembrar (ver SENOR, 2023), e uma ética da memória investiga questões relacionadas ao direito de esquecer (ver MATHESON, 2017) e ao dever de lembrar (ver BLUSTEIN, 2017).

a diferença entre lembrar e aprender de novo (*relearning*)? A principal questão respondida pela teoria causal da memória de Martin & Deutscher (1966), mas não a única, é a questão sobre a linha que separa o lembrar do imaginar. Esta linha se relaciona à questão acerca da mnemicidade (ver MICHAELIAN & SUTTON, 2017), isto é, à questão: Lembrar é imaginar?

Com respeito a esta questão, uma teoria causal se mostra atraente por várias razões. Em primeiro lugar, é intuitivo que a lembrança de E deve ter um vínculo com E que seja diferente da mera imaginação de E, pois podemos imaginar o que não aconteceu, mas não podemos lembrar do que não aconteceu, e é plausível que a causação a partir do acontecimento marque a diferença. Além disso, o critério causal explica por que uma mera representação acurada de E pode não ser uma lembrança de E. Seguindo um exemplo proposto por Putnam (1992 [1981]), imagine que uma formiga caminha pela areia da praia, e suas pegadas formem a efígie de Winston Churchill. O desenho na areia é semelhante a Churchill, mas dificilmente foi apropriadamente causado pela experiência visual que a formiga teve de Churchill. Algo semelhante se dá quando você imagina fielmente uma situação que você nunca vivenciou. Neste caso, a falta de vínculo causal apropriado faz a diferença entre lembrar e imaginar.

2.2 A teoria causal da lembrança

Em Sant'Anna & Michaelian (2019a), apresentamos motivações para a teoria causal, e também alguns dos seus problemas. Aqui, nosso objetivo é, antes de tudo, atualizar nossa exposição, pois o debate em torno da teoria causal da memória aprofundou-se desde nosso último artigo em língua portuguesa. Na análise causalista da memória, tal como proposta por Martin & Deutscher (1966, p. 166), um sujeito S lembra de um evento E se, e somente se, S satisfaz as seguintes condições (cada uma necessária e todas em conjunto suficientes):

(1) *Condição de Experiência Anterior.* S experienciou E.

(2) *Condição de Representação Presente.* S representa E no presente.

(3) *Condição de Conexão Causal Apropriada.* A representação presente que S tem de E está apropriadamente conectada à experiência anterior que S teve de E.

A análise causalista de Martin & Deutscher (1966) ilumina o conceito comum de memória, destacando a importância da experiência anterior, da representação presente e, crucialmente, da conexão causal apropriada[4]. Esta última condição, a qual exige uma ligação causal sustentada por traços de memória originados na experiência original, é o que distingue a memória de outros processos mentais tais como o imaginar e o inferir.

As condições de experiência anterior e de representação presente são razoavelmente intuitivas. A primeira condição diz que você só lembra daquilo que já fez parte da sua experiência, seja perceptualmente, seja introspectivamente. A segunda condição diz que se você está lembrando de E, então uma representação de E se dá na sua mente. A terceira condição é aquela que distingue o causalismo de outras teorias do lembrar[5].

2.3 O caso do pintor: lembrança ignorante e relevância epistêmica

Na análise causalista, o sujeito só pode lembrar do que de fato aconteceu (ver BERNECKER, 2017b). Isto é, os causalistas interpretam o verbo "lembrar" como sendo factivo, "de sucesso". No entanto, a teoria causalista "clássica" de Martin & Deutscher (1966) não estabelece, como condição necessária para o lembrar, que o sujeito *acredite* que aquilo que é representado na sua mente tenha ocorrido no seu passado pessoal[6]. Esta posição, contudo, abriu espaço para duas variedades de causalismos: o causalismo "neo-clássico" de Bernecker e a teoria híbrida (causal-epistêmica) de Debus (2010)[7]. A diferença entre

4 Cabe notar que frequentemente, na literatura, a descrição definida "a teoria causal da memória" é entendida como abrangendo tanto, por um lado, a proposta de Martin & Deutscher (1966) na sua integralidade quanto, por outro, todo e qualquer *causalismo* sobre a memória. No entanto, há causalistas que rejeitam elementos da proposta de Martin & Deutscher (1966) e, neste sentido, não satisfazem essa leitura comum da descrição definida "a teoria causal da memória". Werning (2020, p. 304), por exemplo, concorda com Martin & Deutscher (1966) que lembrança requer ligação causal entre experiência anterior e representação posterior, mas rejeita que deva haver "analogia estrutural" (ver MARTIN & DEUTSCHER, 1966, p. 173) entre a representação mnêmica e o evento lembrado. Além disso, Werning (2020) rejeita que deva haver transmissão de conteúdo da experiência à representação via traços de memória. Deixaremos o aprofundamento deste tópico para uma oportunidade futura.

5 Esta condição envolve a noção de uma "conexão causal apropriada". Resumindo uma longa história (ver BERNECKER, 2010; ROBINS, 2016), uma conexão causal apropriada é aquela que é sustentada por traços de memória originados na experiência que S teve de E. Para uma introdução ao papel da noção de traço (engrama) na filosofia da memória, ver De Brigard (2020).

6 Isto é, na análise causalista do lembrar, há a exigência de que E tenha acontecido para que haja lembrança de E, mas não há exigência de que haja *crença de que E aconteceu* para haver lembrança de E.

7 Com respeito à rica taxonomia de teorias da memória desenvolvidas após Martin & Deutscher (1966), ver Michaelian & Robins (2018).

estas duas propostas se torna visível a partir de um experimento mental famoso: o caso do pintor (ver MARTIN & DEUTSCHER 1966, p. 167-8).

Considere o caso de um pintor que foi comissionado para pintar um cenário realístico, mas totalmente imaginário. O pintor pinta o quadro e o mostra aos seus pais. Seus pais, no entanto, reconhecem na cena pintada um lugar ao qual levaram o pintor, na infância, uma única vez. A partir deste caso, podemos perguntar: Qual a melhor explicação para a pintura? Será que o pintor simplesmente imaginou a cena que viu na infância, ou será que o pintor está lembrando? Para Martin & Deutscher (1966, p. 167-8), o pintor lembra, mesmo que não acredite que lembra.

Essa visão, no entanto, foi desafiada em teorias causais mais recentes. São dois os conceitos que nos auxiliam a avaliar o caso do pintor: lembrança ignorante e relevância epistêmica. Por um lado, dá-se *lembrança ignorante* quando o sujeito lembra, mas não acredita que lembra (ver BERNECKER, 2010, p. 103). Esse seria o caso do pintor, pois é plausível que ele pinte a cena do passado por estar lembrando, ainda que não ache que esteja lembrando. Por outro lado, uma representação mental é *epistemicamente relevante* para um sujeito se este sujeito está disposto a levar essa representação em conta ao julgar sobre o que se deu no passado (ver DEBUS, 2010, p. 21). Ora, a cena do passado não é epistemicamente relevante para o pintor, uma vez que ele não a leva em conta ao julgar sobre o que se deu no passado.

Como se vê, os conceitos de lembrança ignorante e de relevância epistêmica nos permitem detalhar algumas variedades de teorias causais da memória. No causalismo neo-clássico (Bernecker, 2010), aceita-se a visão de Martin & Deutscher (1966) de que o pintor lembra da cena, pois, dada a riqueza de detalhes que se ajustam à vivência passada, é implausível que se trate de mera imaginação. Desse modo, para Martin & Deutscher (1966) e para Bernecker (2010), o pintor lembra, porquanto a explicação mais simples é que a representação da cena pintada foi causada pela experiência anterior da cena. Mas, como o pintor não acredita que está lembrando, o pintor tem uma "lembrança ignorante".

Debus (2010), no entanto, entende que Martin & Deutscher (1966) propuseram condições necessárias, mas não suficientes, para distinguir a lembrança da imaginação. Isto é, para Debus (2010), para haver lembrança, é preciso não somente que as três condições propostas pela análise causalista do conceito de memória sejam satisfeitas, mas também que, como uma quarta condição para o lembrar, se satisfaça a condição de relevância epistêmica. Isto é, não basta que a representação mental tenha sido causada pela experiência anterior. Além disso, o sujeito tem que acreditar que a representação mental é

relevante para o pensamento sobre o que se deu no passado.

De modo que, mesmo entre os causalistas, a teoria causal da memória "clássica" proposta por Martin & Deutscher (1966) está longe de ser a última palavra – sendo, antes, apenas o início de uma importante maneira de se investigar a natureza da memória. Como os conceitos de lembrança ignorante e de relevância epistêmica nos deixam ver, há ainda um longo caminho para que os causalistas, os quais partem de uma metafísica da memória robusta, cheguem a um consenso sobre a epistemologia da memória.

2.4 Construção

2.4.1 Preservacionismo: o modelox "xerox" da memória

Em Sant'Anna & Michaelian (2019a, seção 1.2.2), apresentamos o fato de que a memória é uma capacidade construtiva (em vez de preservativa) como um problema para a teoria causal da memória (De BRIGARD, 2014, é uma boa exposição acerca deste fato). Neste texto, queremos dar mais um passo, agora expondo soluções causalistas para este problema.

Objeta-se, contra a teoria causal, que ela seria incompatível com o caráter construtivo do processo que gera tanto as lembranças quanto as imaginações episódicas (ver MICHAELIAN, 2016a)[8]. Nesta leitura, o causalismo está comprometido com o preservacionismo (ver MCCARROLL, 2018; MICHAELIAN & ROBINS, 2018), ou "modelo xerox da memória" (BERNECKER, 2008, p. 144), segundo o qual o conteúdo da percepção anterior e o conteúdo da representação mnêmica posterior devem ser do mesmo tipo. Por exemplo, segundo o modelo xerox da memória, se você experienciou perceptualmente a visão de um gato sobre a mesa (representação do tipo G, digamos), mas você representa mnemonicamente um animal doméstico sobre a mesa (representação do tipo A, digamos), então você não lembra, pois representações de gatos e representações de animais são de tipos diferentes.

Causalistas, no entanto, não precisam se comprometer com o preservacionismo. Causalistas podem aceitar que o sistema de memória *constrói* uma representação que pode

8 Esta objeção é natural, pois a teoria causal da memória enfatiza o vínculo inalterado entre o evento anterior e a representação posterior, o que é uma tese metafísica. No entanto, é comum que se faça uma leitura epistêmica deste vínculo, entendendo-se que, para um causalista, a representação mnêmica em nada difere da representação perceptual passada.

diferir da representação perceptual. Mas nem todo tipo de construção gera memórias genuínas.

Uma maneira pela qual o caráter construtivo da memória se compatibiliza com o causalismo, gerando lembranças genuínas, é através da *subtração* de informação (ver BERNECKER, 2008, capítulo 9). Considere novamente o caso no qual o sujeito que percebeu um gato lembra de um animal (sendo que a lembrança é mais geral do que a experiência anterior). Neste caso, aquilo que é lembrado é a informação que já estava incluída na percepção, pois (suponhamos) o sujeito concebe gatos como animais (em vez de robôs controlados por marcianos, para darmos outro exemplo proposto por Putnam, 1992 [1981]). Assim sendo, mesmo havendo transformação da informação, se há uma conexão causal apropriada entre a experiência anterior e a representação posterior, então se trata de uma lembrança (ver BERNECKER, 2008).

2.4.2. A extensão de limites

Há construções mnêmicas, no entanto, que envolvem *adição* de informação. No fenômeno da extensão dos limites *(boundary extension)*, por exemplo, o sujeito lembra do que não percebeu (ver INTRAUB, 2020). Suponha que você foi na loja, foi atendido, e você lembra das pernas do balconista, mas você nunca as viu, por causa do balcão que barrava sua visão. Este é um caso que desafia o causalismo, pois não há vínculo causal entre o conteúdo percebido anteriormente e a representação mnêmica posterior. A resposta causalista, neste caso, é que a extensão de limites é um *processo* confiável e que adições de conteúdo causadas por processos confiáveis devem ser reconhecidas como geradoras de lembranças genuínas (ver BERNECKER, 2017a, p. 9). Esta resposta, no entanto, não é satisfatória, visto que parece estar em tensão com a própria ideia de uma teoria causal da memória, a qual enfatiza a necessidade de uma conexão causal apropriada entre a experiência original e a representação posterior, pois, no fenômeno da extensão dos limites, a pessoa lembra de algo que não percebeu originalmente, o que conflita com o requisito de uma conexão causal apropriada[9].

9 É importante observar que o debate usual em filosofia da memória sobre o fenômeno da extensão dos limite parte de pressupostos que vêm sendo questionados. Para Nanay (2022), a memória não é mera cópia da percepção, pois é simplesmente usual que a mente use imagens mentais para preencher lacunas da percepção. Assim, não se dá nada fora do usual no fenômeno mnêmico da extensão dos limites.

2.4.3. A mudança de perspectiva

Outra maneira pela qual a memória pode ser construtiva é pela mudança de perspectiva (McCARROLL, 2018). A perspectiva, em uma memória, é o ponto de vista a partir do qual a memória é reconstruída. Em alguns casos, uma pessoa pode lembrar de um evento como se estivesse observando a si mesma de fora – isto é, como se estivesse assistindo a uma cena em terceira pessoa. Isso é conhecido como *perspectiva do observador* (ver NIGRO & NEISSER, 1983). Em outros casos, a pessoa pode lembrar do evento a partir de sua própria perspectiva, como se estivesse revivendo a experiência exatamente como ocorreu, com o campo de visão correspondendo ao da situação original. Isso é conhecido como *perspectiva de campo* (ver NIGRO & NEISSER, 1983). Por exemplo, considere a lembrança do seu primeiro dia na escola. Talvez você se lembre deste dia "de dentro", revivendo a sensação de segurar a mochila e a visão do rosto do professor te dando boas-vindas. Mas quem sabe você, da perspectiva de um observador externo, se veja, "de fora", entrando na sala de aula pela primeira vez, curioso e ansioso.

Será que lembranças "de fora" são, genuinamente, memórias? Não para Von Leyden (1961), um filósofo da memória dos anos 1960. Mas, desde os anos 1980, os psicólogos têm entendido que é possível lembrar "de fora": "Em algumas memórias, a pessoa tem a perspectiva de um observador, vendo a si mesmo 'de fora'. Em outras memórias, a pessoa vê a cena de sua própria perspectiva; o campo de visão nessas memórias corresponde ao da situação original" (NIGRO & NEISSER, 1983, p. 467; tradução nossa). Assim sendo, a memória constrói até mesmo representações de experiências que não poderíamos ter tido, posto que nos vemos de um ponto de vista diferente do ponto de vista da experiência inicial.

Os causalistas são capazes de explicar o fenômeno da lembrança "de fora". De acordo com o causalismo de McCarroll (2018), um sujeito é capaz de se lembrar "de fora" porque a experiência perceptual é, ela mesma, construtiva. Mais precisamente, a experiência perceptual é rica, envolvendo informação sensorial, mas também informação emocional e conceitual. Ora, de tal riqueza de informação codificada, várias são as permutações possíveis na evocação, sendo que o lembrar-se "de fora" é uma dessas permutações.

O caso das "lembranças de fora" é desafiador para todos os filósofos da memória, pois há razão para se suspeitar que se trate de mera imaginação (novamente, indicamos a leitura de VON LEYDEN, 1961). No entanto, há boas razões para seguir de perto a pesquisa psicológica contemporânea, a qual aceita a possibilidade de lembranças do ponto de vista do observador. Assim sendo, a proposta de McCarroll (2018) se mostra valiosa não

só por atualizar a compreensão filosófica deste fenômeno, como também por mostrar a vitalidade da teoria causal da memória ante um caso tão desafiador.

2.4.4 O efeito DRM

Outra transformação da memória é conhecida como o *efeito DRM* (DEESE, 1959; ROEDIGER & MCDERMOTT, 1995). O nome do efeito faz referência aos pesquisadores que o estudaram: James Deese, nos anos 1950, Henry Roediger e Kathleen McDermott, nos anos 1990. Trata-se de um fenômeno psicológico comum, no qual o sujeito lembra de palavras que não foram escutadas ou lidas antes, mas que estão semanticamente associadas às palavras que foram ouvidas ou lidas antes.

O procedimento para observar o efeito DRM é apresentar ao sujeito várias palavras semanticamente relacionadas a uma "isca" que não é apresentada, mas tem grandes chances de ser "lembrada" após a apresentação das outras palavras. Por exemplo, considerando a isca "sono", apresenta-se ao sujeito as palavras "cama", "acordar", "cansado", "sonho", "soneca", "cobertor", "ronco" e "cochilo". Após ouvir a lista de palavras apresentadas, há boa chance de que o sujeito "lembre" erroneamente de ter ouvido a palavra "sono", a qual não estava na lista, mas está fortemente associada às palavras presentes na lista.

Os causalistas são capazes de explicar o efeito DRM. Robins (2016), uma importante filósofa causalista, entende que o efeito DRM não se dá porque o sujeito busca pelo item errado, dado que o sujeito reconhece os outros itens da lista apresentada. O efeito DRM também não se explica por ser um caso de raciocínio, pois o sujeito acredita *lembrar* da "isca". Ou seja, a experiência se apresenta, para o sujeito, como sendo a de uma memória, e não uma de um raciocínio. Também não se pode explicar o efeito DRM por decaimento do traço de memória, visto que o resultado de "lembrar" da "isca" envolve a lembrança, em vez do esquecimento, dos outros itens da lista apresentada. Por fim, o efeito DRM não se explica por aleatoriedade, pois se trata de um efeito psicológico comum. A explicação do efeito DRM é, antes, que o sujeito erra ao "lembrar" da "isca" por lembrar dos outros itens antes apresentados. Mas, assim, o efeito DRM prova que a experiência anterior é causa da lembrança, seja nos casos de lembrança acurada, seja nos casos de "lembrança" inacurada.

Em suma, o caráter construtivo da memória é um desafio enorme para a teoria causal da memória. No entanto, o causalismo vem buscando maneiras de superar este obstáculo, seja através da noção semântica de subtração de informação, seja através de maneiras de explicar situações nas quais há adição de informação, como se vê no caso de

fenômenos psicológicos como a extensão dos limites, da mudança de perspectiva e do efeito DRM. Em quase todos esses casos, a teoria causal da memória propõe explicações compatíveis com a premissa fundamental de uma conexão causal entre a experiência original e a representação posterior.

2.5 A proposta simulacionista

2.5.1. A análise simulacionista do lembrar

Quais são as condições que precisam ser satisfeitas para que um sujeito lembre, segundo o simulacionismo? Partindo de elementos da análise causalista do lembrar, o simulacionista defende que um sujeito S lembra de um evento E se, e somente se (ver MICHAELIAN, 2016a):

> (2) *Condição de Representação Presente.* S representa E.
>
> (4) *Condição de Confiabilidade.* O sistema de construção episódica que gera a representação de E (4a) opera confiavelmente e (4b) tem como objetivo representar um evento do passado pessoal de S.

A Condição de Representação Presente adotada pelo simulacionista é a mesma adotada pelo causalista. Mas o simulacionista não aceita nem a Condição de Experiência Anterior, nem a Condição de Conexão Causal Apropriada. Em vez disso, o simulacionista propõe uma Condição de Confiabilidade que busca descrever os produtos do sistema de construção episódica que merecem ser considerados como casos de lembranças genuínas. A ideia é que se este sistema, operando confiavelmente (4a), tem como objetivo representar um evento específico do seu passado pessoal (4b), então você lembra deste evento.

Os simulacionistas rejeitam a terceira condição proposta pelos causalistas, isto é, a Condição de Conexão Causal Apropriada. No entanto, é intuitivo que S lembra de E porque S vivenciou E anteriormente (ver MICHAELIAN & ROBINS, 2018). Se lemos este "porque" causalmente, uma explicação causal do lembrar é intuitiva. Por que o simulacionista rejeita tal tese? A principal razão para tal rejeição é a visão, motivada pelos resultados empíricos discutidos na Seção 1, de que o sistema de construção de episódios deve operar da mesma maneira em todos os casos. Isto é, nos casos de lembrança episódica, de pensamento episódico contrafactual e de imaginação orientada ao futuro.

Mais especificamente, dado que imaginar eventos futuros ou contrafactuais não exige que os tenhamos experienciado, os simulacionistas sustentam que o sistema pode representar eventos do passado pessoal de um sujeito mesmo que este sujeito não os tenham vivenciado[10]. Ou, mesmo que a vivência do passado seja, circunstancialmente, a explicação do lembrar, esta explicação não envolve nenhuma necessidade, pois o mecanismo pode gerar resultados independentemente de tais vínculos causais (ver MICHAELIAN, 2016a). Portanto, segundo o simulacionismo, você pode lembrar do que não experienciou. E, nas situações nas quais você lembra do que vivenciou, não é necessário que aquilo que você vivenciou seja a explicação do fato de você lembrar em vez de meramente imaginar. Em suma, se você representa uma situação do seu passado, e o sistema que gera tal representação opera confiavelmente, então você lembra dessa situação.

2.5.2 Críticas ao simulacionismo

A proposta simulacionista, no entanto, enfrenta duras críticas. Mais adiante, focaremos na crítica de McCarroll (2020) ao simulacionismo. Nesta subseção, gostaríamos de simplesmente reconhecer, brevemente, algumas críticas recentes ao simulacionismo. Como são muitas e muito variadas, nesta seção nos limitamos a apresentar, sucintamente, algumas das principais objeções ao simulacionismo que encontramos na literatura recente – em um debate ainda em curso, diga-se de passagem. Para dar alguma ordem a esta literatura, organizamos nossa resenha da seguinte maneira: questões conceituais, questões fenomenológicas, questões empíricas, questões epistemológicas e questões sobre referência.

Questões conceituais. Uma família de críticas ao simulacionismo diz respeito aos conceitos centrais da teoria, como, por exemplo, os conceitos de simulação e de imaginação. Andonovski (2019) aponta para uma dificuldade relacionada ao emprego do termo "simulação" na teoria proposta por Michaelian (2016a). Por um lado, Michaelian (2016a) parece seguir a proposta de Schacter *et al.* (2008, p. 42), para quem "simular" é construir imaginativamente cenários e eventos hipotéticos. Por outro lado, esta visão parece trivialí-

10 Note que, por "experiência" ou "vivência", entende-se a ideia de que o sujeito tem experiências conscientes que resultaram no registro de informação no sistema por meio de traços mnêmicos. Desse modo, existem eventos que pertencem ao seu passado pessoal, mas que não foram experienciados ou vivenciados – por exemplo, o evento do seu nascimento ou da primeira vez que você dormiu em seu berço. O que o simulacionista nega é, portanto, a afirmação de que a informação que constitui uma lembrança no presente precisa ter origem na experiência ou vivência passada. Veja, no entanto, a discussão do simulacionismo radical abaixo, que "radicaliza" a teoria e abandona até mesmo a exigência de que o sistema represente eventos do passado pessoal.

zar a proposta de se caracterizar o lembrar episódico como uma maneira de simular imaginativamente. Pois, se "imaginar episodicamente" for construir tais cenários hipotéticos, não há dúvida que lembrar é imaginar.

Mas a pergunta permanece aberta: lembrar episodicamente é simplesmente construir cenários? Esta crítica é importante por colocar pressão sobre a questão acerca da substantividade das discussões acerca da proposta simulacionista. Usualmente, presume-se que os debates acerca do simulacionismo envolvem discordâncias substantivas entre os participantes. No entanto, como busca mostrar Langland-Hassan (2021), talvez haja mais concordância entre continuístas e descontinuístas do que usualmente se suspeita. Voltaremos ao tema das questões conceituais em filosofia da memória mais adiante.

Ainda nas críticas aos conceitos empregados pelos simulacionistas, Schwartz (2020) foca na noção de *função* empregada pelos simulacionistas. Segundo Schwartz (2020), os simulacionistas entendem por "função" de um sistema o seu valor para a sobrevivência e o sucesso reprodutivo. Nesta visão, a capacidade de simular o passado e o futuro, isto é, de viajar no tempo mental, é vista como uma vantagem evolutiva. No entanto, a visão tradicional da memória como um armazém, a qual foi imortalizada nas metáforas de Platão (2007), no diálogo *Teeteto*, do bloco de cera e do aviário, também é sobre a função da memória, mas não sobre o mesmo sentido de "função".

No caso tradicional, ao se pensar sobre a função da memória, o que se tem em mente é como a memória contribui para a realização de tarefas. Não se trata, neste caso, do valor da memória para a sobrevivência, mas, sim, do que a memória permite que um animal faça. A comparação relevante, neste caso, é entre a memória e as capacidades sensoriais, as quais sem dúvida têm valor para a sobrevivência, e realizam uma função biológica, mas também realizam uma função de mecanismo de processamento de informação sensorial que modifica e modula o comportamento do animal. Nada impede, nota Schwartz (2020), que a memória realize a função de sobrevivência indicada pelos simulacionistas e também a função sistêmica indicada pela tradição. Assim, o simulacionismo pode, quem sabe, ser compatível com a visão tradicional da memória como um armazém de informações.

Simulacionismo e fenomenologia. Rivadulla-Duró (2022) apresenta críticas acerca da capacidade da teoria da simulação de explicar o sentimento de lembrança que é típico da experiência de lembrar. Toda teoria do lembrar episódico precisa explicar este sentimento, e este requisito é ainda mais crítico no caso do simulacionismo, pois se trata de uma teoria que, primeiro, não diferencia em nível profundo o lembrar do imaginar; segundo, reconhece que há tal sentimento; e, terceiro, dá a este sentimento o papel de marcador, para o sujeito, da diferença entre lembrar e imaginar.

Mas, pergunta Rivadulla-Duró (2022), como o simulacionista explica a capacidade que o sujeito tem de distinguir uma lembrança genuína de um pensamento contrafactual episódico? Para Rivadulla-Duró (2022), a realização de tal distinção depende de um mecanismo involuntário que é capaz de distinguir o atual do meramente possível. No entanto, os simulacionistas não explicam como este mecanismo de detecção da realidade opera. Assim sendo, o simulacionismo tem uma séria lacuna explicativa.

Simulacionismo e ciências da memória. Não raro, o simulacionismo é visto como uma proposta preferível ao causalismo por ser imediatamente motivado pela psicologia cognitiva da memória, o que dá ao simulacionismo as vantagens de ser uma teoria compatível com as melhores evidências disponíveis e de não ser uma mera reflexão filosófica acerca do conceito de lembrança do senso comum. No entanto, recentemente, Perrin (2021) reavaliou as evidências empíricas em favor do simulacionismo e concluiu que as mesmas evidências podem ser usadas em favor de uma versão procedural do causalismo[11]. Mais especificamente, a partir de dados empíricos sobre movimentos oculares correlacionados à experiência de imagens mentais mnemônicas, Perrin (2021) defende uma abordagem "corporificada" da memória episódica. Michaelian (2022a) responde a Perrin (2021) apontando para o caráter inconclusivo das evidências que favoreceriam o causalismo e o caráter demasiado restritivo da caracterização da memória pela causação apropriada entre experiência original e lembrança subsequente.

Simulacionismo e epistemologia. Hoerl (2022) interpreta o simulacionismo como uma teoria que leva ao eliminativismo com respeito às lembranças episódicas. O argumento de Hoerl (2022) se apoia numa analogia. Assim como, na teoria do conhecimento, a melhor estratégia para definir conhecimento seria uma abordagem do tipo primeiro-o-conhecimento (*knowledge-first*, ver WILLIAMSON, 2000), a melhor estratégia em filosofia da memória seria partir de uma visão da lembrança episódica como a capacidade de reter conhecimento obtido em experiências oriundas do passado pessoal.

Desse ponto de vista, o problema para o simulacionismo seria que, ao negar que a memória seja uma capacidade diacrônica – isto é, uma capacidade que preserva informação obtida em uma representação passada e que transmite essa informação para uma representação presente –, a teoria não teria recursos para explicar como a memória preserva conhecimento. Como resultado, ao invés de ser vista como uma explicação da memória episódica, a proposta simulacionista seria melhor interpretada como uma proposta de eliminação desse conceito.

11 Note que o causalismo procedural é uma teoria processualista acerca das lembranças episódicas. Não se trata de uma teoria acerca da memória procedural.

Outra crítica de caráter epistemológico ao simulacionismo foi proposta por Robins (2019). Robins (2019) entende o simulacionismo como uma teoria que distingue lembrança de imaginação em parte segundo a confiabilidade do sistema de construção episódica, sendo que, primeiro, lembranças são representações do passado geradas por sistemas confiáveis que buscam representar o passado, e, segundo, a confiabilidade do sistema de construção episódica é determinada pela frequência na qual o sistema gera representações acuradas. No entanto, objeta Robins (2019), o sistema de construção episódica de uma pessoa com desordens psiquiátricas pode, infelizmente, produzir mais representações inacuradas do que lembranças genuínas, mas, ainda assim, produzir lembranças genuínas. Assim sendo, Robins (2019) propõe, contra os simulacionistas, que falsas lembranças sejam tipificadas pelo tipo de erro envolvido em vez de pela confiabilidade do sistema de construção episódica.

Ainda no terreno das críticas epistemológicas ao simulacionismo, Werning (2020) entende que o simulacionismo não é capaz de explicar a confiabilidade do sistema de memória episódica. Para o simulacionista, o sistema de memória episódica constrói representações a partir de informações armazenadas que não necessariamente se originam das experiências e eventos que estão sendo representados. No entanto, o uso de traços de memória oriundos das experiências e eventos representados seria a explicação mais simples para a confiabilidade do sistema de memória episódica.

É claro, o sistema de memória episódica é dinâmico e flexível, operando sob a influência de diversos fatores externos e internos. Ainda assim, se o sistema opera sem sofrer a influência causal da experiência anterior, a suposta confiabilidade do sistema é duvidosa[12]. Uma opção, como indica Rivadulla-Duró (2022), seria o abandono, da parte do simulacionista, da confiabilidade do sistema de memória episódica. Este, no entanto, pode ser um preço que o simulacionista não está disposto a pagar.

Simulacionismo e referência. Alguns críticos apontam para dificuldades na proposta simulacionista que dizem respeito à referência a objetos e eventos no conteúdo de uma lembrança episódica (ver SANT'ANNA, 2021a; OPENSHAW, no prelo). Digamos que no passado S tenha experienciado E, e agora seu sistema de construção episódica, operando confiavelmente, gere uma representação de E. É suficiente que o sistema gere a representação de um evento singular que tem similaridade com o evento vivenciado para que esta representação denote o evento experienciado? Se este for o caso, no que o lembrar se

12 Para uma crítica ao argumento de Werning (2020), e portanto um modo em que o simulacionista poderia responder a essa objeção, ver Andonovski (2022).

diferencia do aprender de novo? Não há resposta clara, da parte dos simulacionistas, para esta e outras questões relacionadas à referência[13]. Quem sabe, o simulacionista defenderia que a mera acurácia (o mero ajuste, ao menos) entre o que foi experienciado e o que é representado seja suficiente para que o sujeito esteja lembrando, desde que a lembrança seja gerada por um sistema de construção episódica que esteja operando de maneira confiável. Esta, ao menos, é a interpretação do simulacionismo proposta por Aranyosi (2020).

Pois suponha que você imagine, fielmente, o que se deu na última reunião do seu departamento, e que esta imaginação seja gerada pelo seu sistema de construção episódica, o qual, vamos supor, é confiável. Neste caso, parece que há elementos suficientes para que o simulacionista defenda que você está lembrando desta reunião. Aqui o simulacionismo mostra sua força, pois desafia a tese que lembrança episódica requer experiência direta. Michaelian (2016a) propõe que lembranças episódicas são mais parecidas com pinturas do que com fotografias. Eis um notável trecho do livro *Mental Time Travel*, no qual o autor explica o simulacionismo a partir de uma das suas memórias de infância:

> Quando eu era uma criança no Canadá, eu viajei com minha família por parte dos Territórios do Noroeste. Em algum momento, paramos perto de um rolo de búfalo – uma área onde os búfalos rolavam no chão – e minha mãe me disse para ficar de pé nele enquanto tirava uma foto. Naturalmente, eu estava assustado, imaginando que um búfalo poderia aparecer a qualquer momento, começar a rolar, e assim me esmagar. Ou assim parece que eu me lembro. Na realidade, dada a minha idade na época, e dado que meus pais repetiram a divertida história para mim várias vezes depois, não posso ter certeza de que muito – ou mesmo algum – do conteúdo da minha aparente memória do episódio realmente se origine em minha experiência, ao contrário dos relatos subsequentes fornecidos por meus pais e minhas próprias imaginações subsequentes do episódio. Como disse von Leyden (1961), as memórias são mais como pinturas do que fotografias, e as peças que faltam de um episódio podem ser preenchidas por informações de outras fontes. O argumento do livro é que, em última análise, isso não importa. Minha memória do incidente do rolo de búfalo é uma memória tanto quanto qualquer outra, independentemente de o quanto de minha experiência original tenha sido preservada. E é provável que seja razoavelmente precisa, mais uma vez, independentemente de que alguma coisa da experiência original tenha sido preservada. A memória pode ser mais como uma pintura do que uma fotografia, mas uma pintura pode, afinal de contas, ser bastante precisa. (MICHAELIAN 2016a, p. 238-9; tradução nossa).

13 Ainda não é claro que os simulacionistas são capazes, ou não, de fornecer uma teoria da referência que seja satisfatória. Este é mais um debate em curso (ver Openshaw & Michaelian, em avaliação).

A lembrança do rolo de búfalo é, provavelmente, uma mescla das próprias experiências do jovem Kourken Michaelian com as narrativas dos pais que bem exemplifica a natureza das memórias episódicas segundo o simulacionismo. Mesmo que as lembranças sejam mais parecidas com pinturas do que com fotos, o pintor – o sistema de construção episódica – é confiável. Dado o talento do artista, o resultado costuma ser acurado.

Mas é aqui, também, que o simulacionismo revela seu ponto fraco, pois parece que o lembrar requer mais do que a mera representação fiel do passado. Este é, novamente, o desafio, que se apresenta ao simulacionista, de explicar a referência de uma maneira que vá além da mera criatividade do sistema de construção episódica.

Aranyosi (2020) entende que uma maneira de se lidar com o desafio de conectar a experiência à representação é através de um realismo direto que tome os objetos e eventos percebidos no passado como sendo constituintes das respectivas representações construídas pelo sistema de construção episódica (ver, também, a discussão de propostas similares em DEBUS, 2008, e SANT'ANNA, 2020).

Esta proposta, no entanto, enfrenta seus próprios desafios, uma vez que leva à complicação teórica do disjuntivismo[14]. Considere-se o caso no qual S acha que vê, no horizonte, um oásis, mas se trata apenas de uma miragem. Para o disjuntivista, uma miragem de oásis não constitui uma representação mnêmica da mesma maneira que um oásis genuíno constitui. Ainda que este possa ser o caso, não é claro como esta visão ajudaria a explicar o modo como o simulacionista deveria lidar com a questão da referência.

Em suma, são diversas as críticas ao simulacionismo, nos seus mais diversos aspectos. Aqui, focamos em críticas à maneira como os simulacionistas empregam conceitos filosóficos, à relação entre o simulacionismo e as evidências empíricas, à explicação simulacionista da fenomenologia do lembrar, à epistemologia e à teoria da referência simulacionistas. Em todos esses casos, estamos narrando debates em curso e sugerimos ao leitor acompanhar tais debates para ter informações mais atualizadas.

2.5.3 O simulacionismo radical

Como vimos na seção anterior, há um amplo debate em torno do simulacionismo. No escopo deste artigo, no entanto, pudemos apenas dar uma breve notícia de algumas

14 Ver Sant'Anna & Michaelian (2019b) para uma discussão sobre os problemas levantados pelo disjuntivismo na filosofia da memória. Ver Moran (2022) para uma defesa detalhada do disjuntivismo.

dessas discussões. Nesta seção, no entanto, gostaríamos de ver, um pouco mais a fundo, um desses debates, o qual abriu espaço para um simulacionismo radicalizado.

Comecemos pelo contexto. Sob certas circunstâncias, os simulacionistas entendem que é possível lembrar mesmo do que não foi experienciado. Teoreticamente, a rejeição da tese da necessidade da experiência anterior levou a um importante debate sobre os limites do lembrar segundo o simulacionismo, pois parece que um sujeito pode lembrar mesmo de situações que não poderiam ser lembradas, dados os fatos sobre o funcionamento normal da amnésia infantil (ver McCARROLL, 2020). Além disso, será que é possível lembrar episodicamente de eventos que pertencem às vidas de outras pessoas? Essa não é visão padrão, mas simulacionistas radicais vêm defendendo esta proposta (ver MICHAELIAN, 2022b).

A possibilidade de lembrar episodicamente de situações que não pertencem ao passado *pessoal* levanta questões sobre o lugar do "eu" na lembrança. Em sua autobiografia, o pintor surrealista Salvador Dalí declara que, diferentemente de outras pessoas, ele tem lembranças de sua vida intrauterina. Isto é, Dalí, com seu humor afiado, alega que se lembra de quando ainda estava dentro da barriga de sua mãe. Quer este seja o caso, quer não, levando em conta o relato de Salvador Dalí, McCarroll (2020) lança uma provocação dirigida à teoria simulacionista da memória. De acordo com o simulacionismo, lembrar é simplesmente imaginar. Mas então, pergunta McCarroll (2020), se lembrar é apenas imaginar, as fantasias de Dalí são exemplos de memórias? Ao propor essa pergunta, McCarroll (2020) explora a coerência e as consequências do simulacionismo. Se as fantasias de Dalí são memórias, então o simulacionismo parece abrir a porta para uma gama quase ilimitada de "memórias". Isso poderia levar a uma reavaliação radical de como concebemos a memória e seu papel em nossa cognição.

O simulacionista afirma que "[...] é possível, em princípio, lembrar mesmo quando não se vivenciou de fato o episódio relevante, para começo de conversa" (MICHAELIAN, 2016a, p. 118; tradução nossa). Agora considere Emily, a sonâmbula. Ela tem o comportamento involuntário de caminhar durante o sono. Ontem à noite, enquanto dormia, Emily foi até a cozinha e preparou um sanduíche. Sua colega de quarto, Aline, viu tudo. De manhã, Aline contou detalhadamente a Emily o que viu. Ela disse que garrafas tilintavam quando ela abriu a porta da geladeira, que ela pegou o queijo, que ela se sentou na mesa em frente ao sanduíche, mas não o comeu, que ela sentiu o cheiro de algo temperado com alho, que Emily voltou para a cama, sempre dormindo. Emily nunca teve experiência consciente do que fez nessa noite.

Ainda assim, com base no relato de Aline, Emily criou, em sua mente, uma imagem sensorialmente rica do que fez. Suponhamos que essa imagem mental tenha sido criada por um sistema de construção episódica confiável que esteja operando com o objetivo de representar o que Emily fez na noite passada. Nesse caso, de acordo com o simulacionista, Emily se lembra episodicamente do que fez na noite passada, na medida em que a representação de Emily satisfaz uma condição de *internalidade*, segundo o qual o processo de geração de uma lembrança deve ser interno ao corpo do agente (MICHAELIAN, 2016b).

Considerando a internalidade, uma diferença importante entre lembrar e imaginar para o simulacionista seria que, quando você reaprende algo x, sua representação de x não começa em seu sistema de construção episódica. Por esse motivo, você não se lembra porque outro sistema cognitivo (que não é o seu sistema de construção episódica) é a fonte da representação do seu passado. Mas nada impede a internalização das informações reaprendidas. Por exemplo, depois que Alice conta a Emily sobre seu episódio de sonambulismo, Emily pode internalizar a informação e satisfazer a condição de internalidade.

Mas qual o papel do "eu" na lembrança? A principal resposta simulacionista ao desafio de McCarroll é o *simulacionismo radical*, o qual é a visão que a memória é uma forma de imaginação e que lembrar é simplesmente imaginar um evento do passado, independentemente de esse evento pertencer (ou não) ao passado pessoal (ver MICHAELIAN, 2022b). O simulacionista radical, tal como o simulacionista tradicional, considera que a memória episódica e o pensamento episódico futuro são sustentados por um sistema neurocognitivo comum, o sistema de construção episódica, resposnável por produzir representações de eventos passados e futuros com base em informações armazenadas que são derivadas das experiências do sujeito. E o simulacionismo radical, tal como o simulacionismo tradicional, também rejeita a condição de causação apropriada da teoria causal, segundo a qual a lembrança genuína se distingue da lembrança meramente aparente pela presença de uma conexão causal apropriada. O simulacionismo radical desafia as teorias usuais do lembrar, pois implica a ideia de que alguém pode se lembrar de eventos que não fazem parte de seu passado pessoal, como a chegada de Napoleão a Grenoble em 1815 (MICHAELIAN, 2022b, p. 16). Isso desafia a ideia de que a memória está necessariamente ligada à experiência pessoal.

3. O debate sobre a (des)continuidade entre lembrar e simular o futuro

O problema da mnemicidade diz respeito à relação de identidade ou diferença entre estados mentais que são considerados lembranças e estados mentais que são considerados imaginações (ver MICHAELIAN & SUTTON, 2017). Este problema é, por um lado, epistemológico, pois somos capazes de rotular os nossos estados mentais e os estados mentais de outros como sendo ou lembranças, ou imaginações (ver ROBINS, 2020; MAHR, 2023). Por outro, este é um problema metafísico, pois faz sentido investigar em virtude do que um estado mental é uma lembrança em vez de ser mera imaginação. No debate contemporâneo em filosofia da memória, esta discussão diz respeito à (des)continuidade entre o lembrar e o imaginar, e as principais posições são o *continuísmo*, segundo o qual lembrar é "contínuo" ao imaginar, e o *descontinuísmo*, segundo o qual há ao menos uma diferença fundamental entre memória e imaginação (ver PERRIN, 2016, MICHAELIAN, 2016C; PERRIN & MICHAELIAN, 2017; MICHAELIAN *et al*., 2022). Como este é um tema que não abordamos em Sant'Anna & Michaelian (2019a), e é, também, um tema que recebeu atualizações recentes, resenhamos este debate nesta seção.

A investigação sobre a (des)continuidade entre memória e imaginação tem se mostrado fértil em filosofia da memória. Tudo começa com o debate corrente entre causalistas e simulacionistas, o qual vimos acima. Por um lado, a motivação para a proposta causalista de Martin & Deutscher (1966) é, exatamente, distinguir as lembranças dos outros tipos de estados mentais, incluindo estados mentais de imaginação e estados mentais de reaprendizado. Assim sendo, é natural que os causalistas se mostrem, ao menos inicialmente, como descontinuístas[15]. Por outro, a pesquisa sobre a capacidade de viajar no tempo mental leva Michaelian (2016a) a propor que, fundamentalmente, lembrar é simular; portanto, lembrar é imaginar. Desse modo, ao menos inicialmente, é natural que os simulacionistas se apresentem como continuístas. Assim, surge o debate entre os filósofos da memória que veem o lembrar como sendo uma maneira de imaginar – os *continuístas* – e os filósofos da memória que defendem que há algo que faz com que lembranças sejam diferentes, fundamentalmente, de imaginações – os *descontinuístas*.

Mas (des)continuidade em relação ao quê? Esta questão já é, por si só, tema de debate. Há duas visões. Em primeiro lugar, há quem veja a memória e a imaginação como (des)contínuas com respeito aos processos neurocognitivos envolvidos. Este é o debate so-

15 Ver, entre outros, Bernecker (2008; 2010), Debus (2008; 2010; 2014), Perrin (2018), Werning (2020), De Brigard (2020). É preciso que registremos, no entanto, que há especulação sobre causalismos "continuístas" (ver LANGLAND-HASSAN, 2023b).

bre o (des)continuísmo processual (ver PERRIN, 2016; MICHAELIAN, 2016C; PERRIN & MICHAELIAN, 2017; MICHAELIAN *et al.*, 2022). No entanto, recentemente, Robins (2020), Sant'Anna (2021b) e McCarroll (2023) propuseram que o debate diz respeito não aos processos psico-funcionais, mas, sim, às atitudes de lembrar e de imaginar. Segundo o (des)continuísmo atitudinal, proposto recentemente por Robins (2020), Sant'Anna (2021b) e McCarroll (2023), o foco do debate é se a atitude que temos ao lembrar de um evento passado é a mesma que temos ao imaginar um evento futuro (ver também LANGLAND-HASSAN, 2023a). Abaixo, resenhamos sucintamente essas duas variedades de (des)continuísmo.

3.1 (Des)continuísmo processual

Com respeito ao (des)continuísmo processual, será que as evidências empíricas podem nos ajudar a decidir entre continuísmo e descontinuísmo? Alguns acreditam que sim. Uma defesa vigorosa do continuísmo processual é proposta por Donna Rose Addis:

> Proponho que a memória e a imaginação são fundamentalmente o mesmo processo – simulação episódica construtiva – […] o "sistema de simulação" atende aos três critérios de um sistema neurocognitivo. Independentemente de estarmos lembrando ou imaginando, o sistema de simulação: (1) age sobre as mesmas informações, baseando-se em elementos da experiência que vão desde detalhes perceptuais de granulação fina até informações conceituais de granulação mais grossa e esquemas sobre o mundo; (2) é governado pelas mesmas regras de operação, incluindo processos associativos que facilitam a construção de um andaime esquemático […]; e (3) é servido pelo mesmo sistema cerebral (ADDIS, 2020, p. 233; tradução nossa).

Em suma, Addis (2020) defende que o mesmo sistema opera da mesma maneira ao gerar lembranças episódicas e ao gerar antecipações imaginativas do futuro. Assim sendo, a investigadora defende o continuísmo processual.

Em oposição, os descontinuístas processuais argumentam (1) que agir sobre o mesmo tipo de informação não é suficiente para se tomar duas atividades como constituindo o mesmo tipo de processo (ver ROBINS, 2023, p. 172); (2) que as regras de operação do lembrar são diferentes daquelas do imaginar prospectivo (ver PERRIN, 2016; 2018); e (3) que as evidências empíricas propostas são insuficientes para se concluir que o sistema de construção episódica tem uma única função bem-definida (ver ROBINS, 2023, p. 172). Assim, há dúvidas sobre as bases empíricas do continuísmo processual.

Outras propostas de solução empírica para a questão (des)continuísta processual também se mostraram inconclusivas. Por um lado, há semelhanças entre as representações mnêmicas e as representações prospectivas, pois ambas são sobre eventos construídos (ver MICHAELIAN, 2016a); por outro, há diferença na maneira como o sujeito lida com cenas mnêmicas *vis-à-vis* cenas antecipadas (ver McCARROLL, 2023). Por um lado, as crianças desenvolvem a capacidade de lembrar e de imaginar o futuro mais ou menos ao mesmo tempo (ver SUDDENDORF & CORBALLIS, 1997). Por outro, elas também aprendem a distinguir lembranças de fantasias mais ou menos ao mesmo tempo (ver FIVUSH, 2011; MAHR, 2023). De forma semelhante, por um lado, há evidência de que há uma "rede neural *default*" (ver BUCKNER *et al*., 2008) que gera tanto lembrança quanto imaginação. Por outro, há evidência de que esta rede é mais demandada na geração de imaginação prospectiva do que na geração de lembranças episódicas (ver SCHACTER & ADDIS, 2007). Há ainda outras evidências relevantes, mas, até o momento, ao menos, não se pode dizer que há algum fato observado que mostre, claramente, se lembrar e imaginar são, ou não, (des)contínuos. Como veremos na Seção 5, esta situação com respeito aos estudos e teorias é relevante para se entender os meta-debates acerca dos debates de primeira ordem em filosofia da memória.

3.2 (Des)continuísmo atitudinal

Ainda com respeito à questão da (des)continuidade entre o lembrar e o imaginar, recentemente, partindo de uma proposta de Robins (2020), Sant'Anna (2021b) propôs que o debate está mal-orientado, pois ainda que o processo em virtude do qual se gera lembrança e prospecção fosse o mesmo, haveria razão para se disputar se há ou pode haver diferenças de atitude do sujeito com respeito a um certo conteúdo (ver Sant'Anna, 2021b, p. 78).

Por exemplo, digamos que o sistema de construção episódica gera a cena de uma festa de aniversário. O que tipicamente acontece numa situação como esta? Ora, a resposta a esta questão depende do tipo de atitude cognitiva na qual a simulação está encaixada. Há uma pluralidade de comportamentos e ações que podem envolver um mesmo tipo de simulação, na medida em que a atitude de sujeito é muito diferente se ele *lembra* da festa, se ele *antecipa* imaginativamente a festa, ou se ele imagina a festa que *poderia* ter acontecido. Em todos esses casos, o conteúdo simulado é o mesmo, mas há diferença na atitude cognitiva do sujeito. Ou seja, há diferentes atitudes, ainda que a simulação construída (e talvez o processo de construção) seja o mesmo.

Para um continuísta atitudinal, o sujeito tem o mesmo tipo de postura ante um conteúdo quando lembra e quando antecipa imaginativamente o futuro, mudando apenas a orientação temporal, para o passado no caso da lembrança, para o futuro no caso da antecipação (ver MICHAELIAN, 2016c, p. 63). Esta visão, no entanto, poderia ser desafiada por um defensor da teoria causal-epistêmica que apontasse para a diferença na relevância epistêmica (ver DEBUS, 2010) entre a atitude de lembrar e a atitude de antecipar, posto que a lembrança seria acompanhada de crença sobre o que aconteceu, enquanto a antecipação não precisa gerar crença sobre o que acontecerá.

A relevância epistêmica é, portanto, uma razão para se adotar o descontinuísmo atitudinal. Mas esta é apenas uma das maneiras nas quais se pode aprofundar a discussão sobre o (des)continuísmo a partir da abordagem atitudinal. De modo que a proposta de Sant'Anna (2021b) de redirecionar o debate (des)continuísta para a questão acerca das atitudes se mostra fértil, pois, se de um lado, é relativamente simples motivar um continuísmo acerca dos processos baseados em viagem no tempo mental, de outro, é muito mais difícil defender que as atitudes de lembrar e de imaginar são "contínuas". Como conclui McCarroll (2023, p. 47):

> […] ao menos algumas formas de imaginação envolvem diferentes tipos de atitudes em relação à lembrança. Muitas formas de pensamento episódico futuro são imaginações conativas que têm uma direção de ajuste do-mundo-para-a-mente, e desempenham um papel motivacional em nossas economias cognitivas. Nesse sentido, a lembrança e essas formas de imaginação são descontínuas. A memória e a imaginação geralmente manifestam ajustes muito diferentes entre mentes e mundos.

4. O debate sobre o sentimento de passado

4.1 Sentimentos epistêmicos

Há vários debates, na filosofia da memória contemporânea, sobre os sentimentos epistêmicos (*epistemic feelings*) que acompanham o lembrar episódico. Um sentimento epistêmico (ou metacognitivo) é uma ocorrência mental que ocorre espontaneamente e diz respeito aos próprios processos mentais (ver ARANGO-MUÑOZ & MICHAELIAN, 2014). Sentimentos epistêmicos têm conteúdo intencional e caráter fenomenal, sendo não-conceituais em animais não-humanos e em crianças pequenas, mas podendo ser conceituais em adultos (ver ARANGO-MUÑOZ, 2014). Sentimentos epistêmicos são parte da vida cognitiva usual.

Considere o *sentimento de saber* (*feeling of knowing*). Às vezes sentimos que sabemos de algo. Quando perguntado sobre qual é a capital do Peru, você pode sentir que sabe que a resposta é "Lima". Um exemplo mais simples é a *certeza* que você sente sobre a verdade de um pensamento ou opinião. Outro exemplo igualmente simples é o sentimento de *dúvida* com respeito a uma opinião. Há ainda o sentimento que a informação está na *ponta da língua*, ainda que você não consiga lembrar, e o sentimento de esquecimento, o qual se manifesta por haver indícios de lacunas, ou por haver dificuldade para acessar alguma informação. Sobre esses e outros tipos de sentimentos epistêmicos, ver Arango-Muñoz (2014).

Na história da filosofia da memória, encontramos várias propostas de "sentimentos" que hoje classificamos como metacognitivos e que estão relacionados ao lembrar. Aristóteles (2012 [séc. 4 a.C.], 450a) e Locke (2010 [1689], 2.10.2) descrevem o lembrar como uma experiência na qual algo é "percebido" com a "percepção" adicional de que isto que foi percebido já foi percebido antes. William James (1890) descreve o lembrar como uma experiência acompanhada dos sentimentos de calor e intimidade. Dorothy Wrinch (1920) relaciona o lembrar ao sentimento de familiaridade; Russell (1976 [1921]), na *Análise da Mente*, ao sentimento de passado; Mahr (2023) ao sentimento de lembrar. Mas para que servem tais sentimentos? Uma resposta é que o sentimento de passado que acompanha o lembrar episódico serve para indicar ao sujeito lembrante que a cena, a qual é por si mesma atemporal (ver RUSSELL, 1976 [1921]), é acerca do passado. Ou seja, o passado não seria parte daquilo que é representado na memória, mas, sim, algo que se sente ao se ter em mente uma cena simulada (sobre a representação do tempo no lembrar, ver De Brigard & Gessell, 2016).

4.2 A relação entre sentimentos metacognitivos e o lembrar

Daqui para diante, por uma questão de simplicidade, consideraremos que o lembrar é acompanhado de um sentimento de passado. Nesta seção, nos ocupamos da relação entre o sentimento de passado e o lembrar.

Dokic (2014; 2022) propôs uma explicação da memória episódica em duas camadas, a qual tem uma representação do passado numa camada e um sentimento de saber episódico na outra. A relação entre as duas camadas, no entanto, é complexa. Por um lado, a representação do passado se relaciona ao passado pessoal do sujeito. Por outro, o sentimento de passado relaciona dois tipos de "metadados" a essa representação: que se

trata da representação de algo familiar e que se trata da representação de algo que se origina da experiência pessoal de primeira mão[16]. Ou seja, o sentimento de passado indica que a representação não é fruto de raciocínio ou testemunho. Assim sendo, o sentimento de passado não é mero adorno afetivo da representação do passado. Muito ao contrário, ele fornece ao sujeito informação sobre a origem, no passado pessoal, da representação do passado. De modo que o sentimento de passado que acompanha a representação do passado no lembrar episódico serve de guia para que o sujeito forme opinião sobre o que vivenciou antes, o que se relaciona, novamente, à questão da relevância epistêmica (ver DEBUS, 2010).

Assim, a hipótese central de Dokic (2014; 2022) é que o lembrar episódico tem duas camadas cognitivas, uma envolvendo uma representação de uma situação passada, outra envolvendo um sentimento epistêmico relacionado a esta representação. O sentimento de passado faz com que a representação encaixada na primeira camada pareça ser de primeira mão. Isto é, o sentimento de passado indica a origem da representação como sendo uma certa vivência passada do próprio sujeito.

Mas o que faz com que a lembrança seja *episódica*? Simplificando, há duas possibilidades. De um lado, pode ser que a episodicidade se explique pela primeira camada, a qual envolve uma representação de uma situação. De outro, pode ser que a episodicidade se explique pela segunda camada, a qual envolve um sentimento de passado. Seria simples usar o sentimento de passado como o elemento gerador de episodicidade. Dokic (2014; 2022), no entanto, defende que a representação da primeira camada já é episódica independentemente da segunda camada. Ou seja, se o sujeito representa mentalmente sua festa de aniversário de dez anos, esta representação já é episódica por se originar da experiência passada.

Nesse sentido, o sentimento de passado é epifenomenal (um "comentário"), pois meramente indica ao sujeito a origem da representação. Dito de outro modo, o sujeito não precisa experienciar o sentimento de passado para lembrar episodicamente, pois tudo o que o sentimento de passado faz é dar indicações epistêmicas sobre o fato da representação ser de primeira mão.

16 Notamos aqui que Dokic (2014; 2022) concebe sua teoria como sendo uma explicação do que ele chama de *sentimento episódico de saber* (*episodic feeling of knowing*). Perrin *et al.* (2020) criticam Dokic nesse sentido, argumentando que o sentimento em questão é melhor concebido como um sentimento de passado (*feeling of pastness*). Como não há muita clareza sobre se há uma disputa substancial ou meramente terminológica aqui, optamos por falar unicamente de um sentimento de passado para facilitar a exposição.

4.3 Intencionalismo e metacognitivismo

Qual a fonte do sentimento de passado? Por que uma imagem mental é sentida como passada, em vez de futura? Com respeito à questão acerca da fenomenologia do tempo no lembrar, uma importante posição é o *intencionalismo*, segundo o qual o sentimento de passado que o sujeito sente ao lembrar episodicamente se explica por aquilo que é representado mnemicamente. Isto é, para o intencionalista, o conteúdo da memória origina o sentimento de passado. Uma importante versão do intencionalismo é o funcionalismo (FERNÁNDEZ, 2019). De acordo com essa proposta, uma lembrança episódica é estruturada de tal modo que ela se representa como tendo sido causada por uma experiência perceptual passada. Assim sendo, quando você lembra de algo, você não está apenas lembrando do evento ele mesmo, uma vez que, além disso, você também representa a lembrança como sendo algo que foi causado pela sua experiência passada do evento. Esse aspecto auto-referencial da memória, segundo Fernández (2019), é o que origina o sentimento de passado. É porque você se representa como tendo percebido E no passado que você sente E como passado.

Contudo, para Perrin & Sant'Anna (2022), o intencionalismo enfrenta dificuldades. Por exemplo, o intencionalista não explica o grau de intensidade do sentimento de passado. Um evento pode parecer mais passado do que outros eventos. Por exemplo, seu café da manhã de hoje pode parecer menos passado do que a última vez que você molhou os dedos dos pés num riacho. Quem sabe você tenha um sentimento de passado mais forte (ou um sentimento de "mais passado") no caso do riacho. Note que, nos dois casos, o que temos é um conteúdo que, segundo o intencionalista, estaria gerando o sentimento de passado. Mas, se tudo o que é preciso para gerar o sentimento de passado é o conteúdo, como se explica a diferença de grau de intensidade do sentimento de passado nestes dois casos?

Para Perrin & Sant'Anna (2022), dificuldades como esta motivam a busca de uma maneira diferente de explicar o sentimento de passado, a qual leva em conta tanto o que é representado mnemicamente quanto o processo de construção da lembrança. Esta é a base da abordagem metacognitiva (PERRIN *et al.*, 2020; PERRIN & SANT'ANNA, 2022; SANT'ANNA, no prelo).

Na explicação metacognitiva do sentimento de passado que se dá, como fenomenologia da consciência do tempo, no lembrar episódico, o sentimento de passado resulta do monitoramento e da interpretação subpessoais – no mais das vezes, das características do processamento neurocognitivo que resulta no lembrar episódico. A *fluência* é uma

dessas características importantes que resulta no lembrar episódico. A fluência, enquanto característica do processo de construção de uma lembrança episódica, diz respeito ao grau de facilidade que o sistema neurocognitivo encontra para gerar a representação do passado. Quando lembramos de um evento, o que se dá não é um mero *replay* de um evento passado (ver LANGLAND-HASSAN, 2023b). Em vez disso, o que se dá é que o encéfalo reconstrói o evento a partir de diversos fragmentos de informação disponível.

Esse processo de reconstrução pode ser mais ou menos fluente, dependendo do que está para ser representado. O processo de reconstrução é mais fluente se os fragmentos de informação que foram utilizados para construir a lembrança estão fortemente interconectados ou foram evocados com mais frequência. Por exemplo, se você costuma rever as fotos da sua festa de aniversário de dez anos, ou costuma falar sobre essa festa com seus familiares, os diversos fragmentos de informação (detalhes visuais registrados fotograficamente, relatos dos parentes, suas próprias memórias) estão fortemente entrelaçados. Isto pode levar a uma construção mais fluente de uma lembrança episódica de algum acontecimento que se deu durante a festa.

Ou, quem sabe, seja mais fácil para você lembrar do seu primeiro dia no seu novo emprego. Vários fragmentos de informação estão muito bem conectados: seu nervosismo, o percurso na cidade e nas instalações da empresa que te contratou, as pessoas que você encontrou, as tarefas que te pediram para cumprir. Talvez você relembre deste dia frequentemente – durante as conversas com novos colegas de trabalho, nos seus encontros com outras pessoas – e, por isso, os fragmentos de informação estejam muito bem conectados uns com os outros, levando a mais fluência no processo de evocação mnêmica do evento.

Nesses exemplos, o que se dá é que seu encéfalo pode, fácil e rapidamente, reunir informações para gerar uma representação do passado. Como o processo é fácil para o cérebro, você sente que aquilo que é representado é passado. Isto é, seu cérebro traduz a facilidade que teve para gerar a representação em sentimento de passado para você. Em comparação, se as informações A e B estão desconectadas, ou não têm sido evocadas com frequência, o encéfalo tem que trabalhar para reuni-las. Por causa do trabalho que tem, o cérebro não sinaliza tais informações como passadas. Por isso, você experimenta tais reconstruções com menos grau de passadidade – ou mesmo sem sentimento de passado. Desse modo, a fluência do processo leva a um sentimento de passado.

Além disso, a fluência do processo leva à confiança na passadidade do produto (isto é, na acurácia da representação gerada), pois se a representação foi facilmente cons-

truída é porque (provavelmente) os fragmentos de informação já estiveram reunidos no passado – na percepção do evento agora lembrado, provavelmente. Inversamente, menos fluência leva a menos confiança de que se trata de uma lembrança em vez de uma fantasia (ver PERRIN & SANT'ANNA, 2022).

Assim, na explicação metacognitiva do elemento autonoético (isto é, do sentimento de passado) que faz parte do lembrar episódico, a fluência é considerada um elemento crucial. Consequentemente, contra a abordagem intencionalista, não é apenas o produto (isto é, o conteúdo lembrado) que importa para explicar o sentimento de passado, pois o processo de construção do conteúdo é parte da explicação. Se apenas o conteúdo fosse responsável por originar o sentimento de passado, não seria de se esperar correlação deste sentimento com a facilidade relativa do processo de construção da representação do passado. Mas há tal correlação. Desse modo, a explicação intencionalista precisa ser revista.

A abordagem metacognitiva da fenomenologia do lembrar episódico também responde ao problema da mnemicidade (isto é, à questão acerca da demarcação do lembrar episódico *vis-à-vis* o imaginar episódico). Segundo Perrin & Sant'Anna (2022), a fenomenologia da consciência do tempo no lembrar episódico desempenha um papel crucial na distinção subjetiva do lembrar em constraste com o imaginar, dado que é justamente por sentir que uma certa representação é do passado que o agente a toma por uma lembrança em vez de uma mera imaginação (ver também MICHAELIAN, 2016a; PERRIN *et al.*, 2020).

Em suma, Perrin & Sant'Anna (2022) defendem que, para explicar o sentimento de passado, a abordagem metacognitiva é preferível à abordagem intencionalista, porquanto a abordagem metacognitiva explica o problema da intensidade do sentimento de passado e explica a correlação entre facilidade do processo de construção e sentimento de passadidade.

5. O meta-debate sobre a natureza dos debates de primeira-ordem

Diante da complexidade e diversidade dos debates na filosofia da memória, surgiu uma reflexão de segunda-ordem: o que realmente está em jogo nesses debates de primeira-ordem? Os filósofos da memória, envolvidos em tantas discussões relacionadas à natureza da memória, foram levados, em alguns casos ao menos, a pensar sobre o que se dá nesses debates. E esses debates de segunda-ordem (que, no momento, estão em anda-

mento) sobre os debates de primeira-ordem chegaram a algumas hipóteses importantes. Focaremos aqui em dois meta-debates, um entre causalistas e simulacionistas, e o outro entre continuístas e descontinuístas.

O debate entre causalistas e simulacionistas envolve acordos e desacordos. Por um lado, causalistas e simulacionistas concordam que dá para se investigar a natureza da lembrança episódica investigando, por exemplo, os erros de memória, os quais compreendem as falsas memórias e as confabulações (ver MICHAELIAN, 2023; BERNECKER, 2023). No entanto, causalistas e simulacionistas estão em desacordo sobre como caracterizar a natureza do lembrar episódico, porque, de um lado, os filósofos causalistas explicam o lembrar a partir da noção de causação, enquanto que, de outro, os filósofos simulacionistas explicam a noção de lembrar episodicamente através da noção de confiabilidade. Este desacordo gera preocupação, pois abre espaço para o risco de haver um debate meramente verbal (ver CHALMERS, 2017).

No outro debate, entre continuístas e descontinuístas, há um acordo geral de que as novas evidências encontradas por psicólogos acerca de haver um sistema único gerando lembranças episódicas e imaginação de si mesmo no futuro devem ser levadas em conta na caracterização do que é o lembrar. Ambos os lados concordam com isso. Ainda assim, há desacordo sobre como interpretar esses dados. Por um lado, os continuístas processuais defendem que encontrar essa continuidade é o suficiente para que se infira que se trata de uma única e mesma espécie natural (*natural kind*). A espécie natural, no caso, é a espécie natural do imaginar episódico – sendo que, dentro dessa espécie, temos uma subespécie que seria o lembrar episodicamente.

Os descontinuístas atitudinais, em contraste, defendem que mesmo havendo a mesma base na natureza para a geração do lembrar episodicamente e do imaginar episodicamente, deveríamos considerar importantes diferenças epistêmicas e metafísicas entre lembrar e imaginar. Epistemicamente, o lembrar tem papéis importantes que não são dados ao imaginar. Nós temos compromissos que dependem de lembrarmos do que aconteceu no passado, não de imaginarmos o que aconteceu no passado. E metafisicamente, nós exigimos do lembrar a factividade. Nós pedimos do lembrar que se represente algo que de fato aconteceu, enquanto que não há problema em se imaginar o que poderia ter acontecido.

Sem dúvida nenhuma, esses debates trouxeram muitos avanços. Além dos ganhos teóricos apresentados acima, há a teoria do minimalismo de traços (WERNING, 2020) e o debate sobre a função do lembrar (SCHWARTZ, 2020; MAHR, 2023; ROBINS, 2023), por exemplo. Ainda assim, os debates estão empatados.

Essa situação levou alguns filósofos a tentarem avaliar se os debates em curso estão bem estruturados. Veio à tona a suspeita de que talvez não se chegue a uma solução, porque o debate, em si, tem certas características. Quem sabe, os embates diretos entre causalistas e simulacionistas e entre continuístas e descontinuístas não resultem em algo conclusivo porque cada lado está pressupondo a própria visão do que é o lembrar, o que poderia nos levar a uma situação de disputas verbais (ver CHALMERS, 2017), na medida em que cada lado estaria falando de algo diferente. Seriam disputas vazias, nesse sentido.

Como estão estruturados, então, os debates em filosofia da memória? Para responder a essa questão, Craver (2020) distingue entre o lembrar epistêmico e o lembrar empírico. Há também o contextualismo explanatório (McCARROLL *et al.*, 2022), a leitura dos debates como negociações metalinguísticas (SCHIRMER DOS SANTOS, 2020; SCHIRMER DOS SANTOS *et al.*, 2023), e a recente proposta de Openshaw (no prelo) de ver as situações que encontramos nesses debates como conflitos de integração.

Craver (2020) propõe que os debates não progridem porque alguns filósofos estão falando de um tipo de lembrança enquanto outros filósofos estão falando de outro tipo. O que se dá é que, por um lado, os simulacionistas e os continuístas estão falando daquilo que Craver (2020) chama de "lembrar empírico". E o que é o lembrar empírico? O lembrar empírico é a produção de uma representação do passado pelo sistema neurocognitivo que gera tanto lembrança episódica quanto a imaginação de si mesmo no futuro. Assim, há uma única espécie natural envolvida na produção de lembranças e imaginações episódicas. Nesse sentido, há uma continuidade entre o lembrar e o imaginar.

Por outro lado, é bem provável que, ainda que aceitem essa noção de lembrar empírico, os descontinuístas estejam pensando em questões relacionadas ao que Craver (2020) chama de "lembrar epistêmico", que é o lembrar enquanto algo que é socialmente importante (ver MAHR & CSIBRA, 2018). O lembrar epistêmico tem um certo papel social. Por exemplo, numa situação de cooperação social, se uma pessoa diz que lembra que ontem choveu, a pessoa está comunicando que ela mesmo percebeu isso, e está dizendo também que isso é verdade. Agora, se essa mesma pessoa diz que imagina que ontem choveu, ela está dizendo que não percebeu isto que ela relata.

Outra visão é o *contextualismo explanatório* (McCARROLL *et al.*, 2022). Nesse caso, a ideia, em parte antecipada por Bernecker (2008), é que a palavra "lembrar" tem diferentes significados em diferentes contextos. Alguns contextos são normativos. Nesses contextos, as condições de sucesso para que se dê o lembrar são de um tipo. Noutros, as condições de sucesso do uso de "lembrar" seriam descritivas, pois estaria se falando das

espécies naturais que se espera individuar como fundamentos objetivos do lembrar. Então, temos novamente aqui um diagnóstico já proposto por Craver (2020): um lado da discussão pode não estar falando da mesma coisa que o outro. É possível, então, que ambos estejam certos, ao mesmo tempo, mas sobre diferentes sentidos de "lembrar".

Schirmer dos Santos e colaboradores (2023) concordam que há mais de um sentido de lembrar envolvido nesse debate, mas discordam do pressuposto tácito de que esses filósofos não estejam percebendo que estejam usando significados diferentes. Existe uma equivocidade no lembrar no debate, mas Schirmer dos Santos e colaboradores (2023) dizem que isso não é um mero acidente. Tal equivocidade do lembrar envolve efeitos pragmáticos, que no caso seriam especificamente negociações metalinguísticas (PLUNKETT & SUNDELL, 2013; PLUNKETT *et al*., 2023) nas quais cada lado, conhecendo muito bem o que o outro lado quer dizer com a palavra "lembrar", usa da sua própria maneira a palavra "lembrar" para mostrar para o outro lado como essa palavra *deveria* ser usada. De modo que a equivocidade não é um acidente. Cada lado está, de fato, sabendo que o outro usa de uma maneira diferente a palavra "lembrar", e usa a "mesma" palavra da sua maneira para mostrar como o lado oposto deveria usar a palavra "lembrar", levando em conta tudo o que se sabe sobre o lembrar. Então é uma disputa substantiva, em vez de verbal, sobre o que a palavra "lembrar" deveria significar.

Por fim, dentre as propostas de meta-debates, nós temos a contribuição de Openshaw (no prelo). O artigo de Openshaw desenvolve e busca esclarecer a proposta de Craver (2020). Enquanto Craver (2020) enfatiza dois sentidos distintos da lembrança e seus respectivos contextos intelectuais, Openshaw enfatiza *três* níveis distintos de investigação sobre a lembrança, sendo que há diferentes questões em cada um desses diferentes níveis de investigação da memória. Como há diferentes níveis de investigação da memória e cada pesquisador está trabalhando com um nível diferente, o que acontece nesse caso é que o que se deve buscar é uma integração desses níveis, numa teoria mais compreensiva.

Para Openshaw, os níveis da investigação são: o nível dos processos psico-funcionais, o nível da referência (que é o nível no qual se investiga a identidade de referência que se espera encontrar entre a percepção passada e a lembrança posterior), e também o nível da acurácia, no qual se avalia a similaridade entre aquilo que foi percebido no passado e aquilo que está sendo lembrado agora.

Para Openshaw, o que nós temos que buscar é a melhor resposta para cada nível. Dessa maneira, os filósofos da memória, cooperativamente, podem buscar criar uma teo-

ria da memória compreensiva, que seja mais completa e que abranja adequadamente cada nível.

Nesse percurso, Openshaw defende que alguns princípios do lembrar que haviam sido rejeitados por quem trabalha no nível dos sistemas psico-funcionais podem ser resgatados. Por exemplo, o princípio de que você só pode lembrar do que foi percebido no passado pode ser resgatado (e deve ser resgatado, de acordo com Openshaw), ainda que os simulacionistas digam que você pode lembrar daquilo que você não percebeu no passado.

Mais especificamente, para Openshaw, embora o simulacionismo explique a natureza dos processos psico-funcionais que ocorrem nos sistemas neurocognitivos, a questão sobre a referência exige uma resposta que pressupõe princípios explicativos diferentes, dentre os quais a relação entre lembrança e experiência passada se coloca como um princípio central.

Sobre a acurácia, alguns filósofos, como Fernández (2019) acreditam que a primeira coisa (a mais importante) para caracterizar o lembrar é se a experiência é de lembrar ou de imaginar; só depois é importante determinar se você está lembrando acuradamente ou não. Ou seja, Fernández (2019) vai dizer que você pode lembrar do que não aconteceu e do que você não vivenciou. Isto é, para Fernández (2019), você pode ter memórias que são ostensivas em vez de factivas. Com respeito a isso, se seguirmos o plano de Openshaw (no prelo), o que temos é uma situação na qual poderíamos responder a Fernández (2019) dizendo que ainda que o sistema de construção episódica gere imaginações que não são factivas, quando falamos do nível da acurácia, deveríamos seguir um princípio preservacionista, e dizer que é lembrança apenas aquilo que preserva uma similaridade com os conteúdos que foram percebidos no passado.

Essas são, portanto, propostas de avaliar o que ocorre nos debates de primeira-ordem para que haja uma melhor coordenação entre os filósofos da memória.

Conclusão

Neste artigo, buscamos enriquecer o leque de publicações sobre os debates centrais da filosofia contemporânea da memória que estão disponíveis ao leitor de língua portuguesa. Na seção 1, apresentamos o conceito de viagem no tempo mental, o qual serve de pano de fundo para boa parte da pesquisa filosófica atual sobre o lembrar episódico. Na seção 2, detalhamos o debate entre causalistas e simulacionistas sobre a metafísica do lembrar episódico. Vimos como o causalismo enfrenta desafios relacionados ao caráter construtivo da memória, e como o simulacionismo propõe um revisionismo sobre a natureza da memória.

Na seção 3, exploramos o debate entre continuístas e descontinuístas sobre a relação entre lembrar e imaginar. Vimos que há discordância sobre se a (des)continuidade entre lembrar e imaginar diz respeito aos processos neurocognitivos ou às atitudes dos agentes em relação aos conteúdos simulados. Na seção 4, abordamos discussões recentes sobre a natureza do sentimento de passado típico da memória episódica, e também os desafios à explicação intencionalista. O intencionalismo explica esse sentimento pelo conteúdo representado na memória. A abordagem metacognitiva, por sua vez, explica o sentimento de passado pelo monitoramento de características do processo neurocognitivo, especialmente sua fluência. Assim, intencionalistas e metacognitivistas discordam sobre a fonte do sentimento de passado na memória episódica.

Por fim, na seção 5, resenhamos propostas de meta-debates que buscam esclarecer a própria natureza das controvérsias nos debates de primeira-ordem da filosofia da memória. Uma proposta distingue entre lembrar epistêmico e lembrar empírico. Outra proposta distingue entre abordagens descritivas e abordagens normativas do lembrar. Uma terceira visão entende os debates como negociações metalinguísticas sobre o significado correto de "lembrar". Por fim, uma quarta visão diagnostica os debates como envolvendo uma falta de integração entre diferentes níveis de investigação da memória.

Embora concisa, esperamos que esta amostra de discussões encoraje mais pesquisas e reflexões filosóficas em português neste fértil campo de pesquisa.

Reconhecimento

Este estudo foi financiado pelo Programa Capes-PrInt da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES), código financeiro 001, processo 8881.310246/2018-1, e por Le Programme Cofecub, processo 88887.468340/2019-00. André Sant'Anna foi financiado pela Alexander von Humboldt-Stiftung no programa Humboldt-Forschungsstipendium. James Openshaw foi financiado pelo programa de pesquisa e inovação Horizon 2020 da União Europeia sob o contrato de concessão Marie Skłodowska-Curie nº 101032391.

Referências

ADDIS, Donna Rose. Mental Time Travel? A Neurocognitive Model of Event Simulation. **Review of Philosophy and Psychology** 11 (2): 233–59. 2020. Acesso: https://doi.org/10.1007/s13164-020-00470-0.

ADDIS, Donna Rose, Alana T. WONG, & Daniel L. SCHACTER. Remembering the Past and Imagining the Future: Common and Distinct Neural Substrates during Event Construction and Elaboration. **Neuropsychologia** 45 (7): 1363–77. 2007. Acesso: https://doi.org/10.1016/j.neuropsychologia.2006.10.016.

ANDONOVSKI, Nikola. Is the Simulation Theory of Memory about Simulation? **Voluntas** 10 (3): 37. 2019. Acesso: https://doi.org/10.5902/2179378640399.

ANDONOVSKI, Nikola. Causation in Memory: Necessity, Reliability and Probability**.** **Acta Scientiarum** 43 (3): e61493, 2022. Acesso: https://doi.org/10.4025/actascihumansoc.v43i3.61493.

ARANGO-MUÑOZ, Santiago. The Nature of Epistemic Feelings. **Philosophical Psychology** 27 (2): 193–211, 2014. Acesso: https://doi.org/10.1080/09515089.2012.732002.

ARANGO-MUÑOZ, Santiago, & MICHAELIAN, Kourken. 2014. Epistemic Feelings, Epistemic Emotions: Review and Introduction to the Focus Section. **Philosophical Inquiries** 2 (1). 2014. Acesso: https://doi.org/10.4454/philinq.v2i1.79.

ARANYOSI, István. Mental Time Travel and Disjunctivism. **Review of Philosophy and Psychology** 11 (2): 367–84, 2020. https://doi.org/10.1007/s13164-020-00467-9.

ARISTÓTELES. Da Memória e da Revocação. In: ARISTÓTELES, **Parva naturalia**. Trad. Edson Bini, São Paulo: Edipro, 2012. p. 75–87.

BARTLETT, Frederic C. **Remembering:** A study in experimental and social psychology. Cambridge: Cambridge University Press, 1932.

BENOIT, Roland G. & SCHACTER, Daniel L. Specifying the Core Network Supporting Episodic Simulation and Episodic Memory by Activation Likelihood Estimation. **Neuropsychologia** 75 (August): 450–57, 2015. Acesso: https://doi.org/10.1016/j.neuropsychologia.2015.06.034.

BERNECKER, Sven. **The metaphysics of memory**. New York: Springer, 2008.

BERNECKER, Sven. **Memory:** A philosophical study. Oxford: Oxford University Press, 2010.

BERNECKER, Sven. A Causal Theory of Mnemonic Confabulation. **Frontiers in Psychology** 8 (July): 1207. 2017a. Acesso: https://doi.org/10.3389/fpsyg.2017.01207.

BERNECKER, Sven. Memory and Truth. In: BERNECKER, Sven & MICHAELIAN, Kourken (Eds.). **The Routledge handbook of philosophy of memory**. London: Routledge, 2017b. p. 51–62.

BERNECKER, Sven. An Explanationist Model of (False) Memory. In: SANT'ANNA, André, McCARROLL, Christopher Jude e MICHAELIAN, Kourken (Eds.). **Current controversies in philosophy of memory**. London: Routledge, 2023. p. 109–26.

BLUSTEIN, Jeffrey. A Duty to Remember. In: SANT'ANNA, André, McCARROLL , Christopher Jude e MICHAELIAN, Kourken (eds.). **The Routledge handbook of philosophy of memory**. London: Routledge, 2017. p. 351–63.

BOYLE, Alexandria. The Mnemonic Functions of Episodic Memory. **Philosophical Psychology** 35 (3): 327–49, 2022. Acesso: https://doi.org/10.1080/09515089.2021.1980520.

BUCKNER, Randy L., ANDREWS-HANNA, Jessica R. & SCHACTER, Daniel L. The Brain's Default Network: Anatomy, Function, and Relevance to Disease. **Annals of the New York Academy of Sciences** 1124 (1): 1–38, 2008. Acesso: https://doi.org/10.1196/annals.1440.011.

CHALMERS, David J. Disputas verbais. Trad. Gregory Gaboardi. **Sképsis** 15: 57, 2017.

CRAVER, Carl F. Remembering: Epistemic and Empirical. **Review of Philosophy and Psychology** 11 (2): 261–81, 2020. Acesso: https://doi.org/10.1007/s13164-020-00469-7.

DE BRIGARD, Felipe. Is Memory for Remembering? Recollection as a Form of Episodic Hypothetical Thinking. **Synthese** 191 (2): 155–85, 2014. Acesso: https://doi.org/10.1007/s11229-013-0247-7.

DE BRIGARD, Felipe. The Explanatory Indispensability of Memory Traces. **The Harvard Review of Philosophy** 27: 23–47, 2020. Acesso: https://doi.org/10.5840/harvardreview202072328.

DE BRIGARD, Felipe, & Bryce S. GESSELL. 2016. Time Is Not of the Essence: Understanding the Neural Correlates of Mental Time Travel. In: MICHAELIAN, Kourken, KLEIN, Stanley B. e SZPUNAR, Karl K (Eds.). **Seeing the Future**. Oxford: Oxford University Press, 2016. p. 153–79.

DE BRIGARD, Felipe, & Natasha PARIKH. Episodic Counterfactual Thinking. **Current Directions in Psychological Science** 28 (1): 59–66, 2019. Acesso: https://doi.org/10.1177/0963721418806512.

DEBUS, Dorothea. Experiencing the Past: A Relational Account of Recollective Memory. **Dialectica** 62 (4): 405–32, 2008. Acesso: https://doi.org/10.1111/j.1746-8361.2008.01165.x.

DEBUS, Dorothea. Accounting for Epistemic Relevance: A New Problem for the Causal Theory of Memory. **American Philosophical Quarterly** 47 (1): 17–29, 2010.

DEBUS, Dorothea. «Mental Time Trave": Remembering the Past, Imagining the Future, and the Particularity of Events. **Review of Philosophy and Psychology** 5 (3): 333–50, 2014. Acesso: https://doi.org/10.1007/s13164-014-0182-7.

DEESE, James. "Influence of Inter-Item Associative Strength upon Immediate Free Recall." **Psychological Reports** 5 (3): 305–12, 1959. Acesso: https://doi.org/10.2466/pr0.1959.5.3.305.

DOKIC, Jérôme. Feeling the Past: A Two-Tiered Account of Episodic Memory. **Review of Philosophy and Psychology** 5 (3): 413–26, 2014. Acesso: https://doi.org/10.1007/s13164-014-0183-6.

DOKIC, Jérôme. Episodic Remembering and Affective Metacognition. **Acta Scientiarum** 43 (3): e61022, 2022. Acesso: https://doi.org/10.4025/actascihumansoc.v43i3.61022.

FERNÁNDEZ, Jordi. **Memory**: A self-referential account. New York: Oxford University Press, 2019.

FIVUSH, Robyn. The Development of Autobiographical Memory. **Annual Review of Psychology** 62 (1): 559–82, 2011. Acesso: https://doi.org/10.1146/annurev.psych.121208.131702.

GOLDMAN, Alvin I. A Causal Theory of Knowing. **The Journal of Philosophy** 64 (12): 357–72, 1967.

GRICE, H. P. The Causal Theory of Perception. **Proceedings of the Aristotelian Society** 35: 121–52, 1961.

HASSABIS, Demis & Eleanor A. MAGUIRE. Deconstructing Episodic Memory with Construction. **Trends in Cognitive Sciences** 11 (7): 299–306, 2007. Acesso: https://doi.org/10.1016/j.tics.2007.05.001.

HOBBES, Thomas. **Os elementos da lei natural e política**. Trad. Bruno Simões. São Paulo: WMF Martins Fontes, 2010.

HOERL, Christoph. A Knowledge-First Approach to Episodic Memory. **Synthese** 200 (5): 376, 2022. Acesso: https://doi.org/10.1007/s11229-022-03702-1.

INTRAUB, Helene. Searching for Boundary Extension. **Current Biology** 30 (24): R1463–64, 2020. Acesso: https://doi.org/10.1016/j.cub.2020.10.031.

JAMES, William. **The principles of psychology**. Vol. 1. New York: Henry Holt and Company, 1890.

Goldman, Alvin I. A Causal Theory of Knowing. **The Journal of Philosophy** 64 (12): 357–72, 1967.

LANGLAND-HASSAN, Peter. 2021. What Sort of Imagining Might Remembering Be? **Journal of the American Philosophical Association** 7 (2): 231–51, 2021. Acesso: https://doi.org/10.1017/apa.2020.28.

LANGLAND-HASSAN, Peter. Remembering and Imagining: The Attitudinal Continuity. In: BERNINGER, Anja & FERRAN, Íngrid Vendrell (Eds.). **Philosophical Perspectives on Memory and Imagination.** New York: Routledge, 2023a. p. 11–33.

LANGLAND-HASSAN, Peter. Remembering, Imagining, and Memory Traces: Toward a Continuist Causal Theory. In: SANT'ANNA, André, McCARROLL, Christopher Jude & MICHAELIAN, Kourken (Eds.). **Current controversies in philosophy of memory**. London: Routledge, 2023b. p. 19–37.

LOCKE, John. Um Ensaio Sobre o Entendimento Humano. In: BONJOUR, Laurence & BAKER, Ann (Eds.). **Filosofia:** Textos fundamentais comentados. Trad. André Nilo Klaudat, Porto Alegre: Artmed, 2010, p. 93–106.

MAHR, Johannes B. Episodic Memory: And What Is It For? In: SANT'ANNA, André, McCARROLL, Christopher Jude & MICHAELIAN, Kourken (Eds.). **Current controversies in philosophy of memory**. London: Routledge, 2023. p. 166–84.

MAHR, Johannes B. & Gergely Csibra. Why Do We Remember? The Communicative Function of Episodic Memory. **Behavioral and Brain Sciences** 41: e1, 2018. Acesso: https://doi.org/10.1017/S0140525X17000012.

MARTIN, C. B. & DEUTSCHER, Max. Remembering. **The Philosophical Review** 75 (2): 161, 1966. Acesso: https://doi.org/10.2307/2183082.

MATHESON, David. An Obligation to Forget. In: BERNECKER, Sven & MICHAELIAN, Kourken (Eds.). **The Routledge handbook of philosophy of memory**. London: Routledge, 2017. p. 364–72.
McCARROLL, Christopher Jude. **Remembering from the outside**: Personal memory and the perspectival mind. New York: Oxford University Press, 2018.

McCARROLL, Christopher Jude. Remembering the Personal Past: Beyond the Boundaries of Imagination. *Frontiers in Psychology* 11 (September): 585352, 2020. Acesso: https://doi.org/10.3389/fpsyg.2020.585352.

MCCARROLL, Christopher Jude. 2023. Memory and Imagination, Minds and Worlds. In: BERNINGER, Anja & FERRAN, Íngrid Vendrell (Eds.). **Philosophical Perspectives on Memory and Imagination.** New York: Routledge, 2023. p. 202335–53. Acesso. https://doi.org/10.4324/9781003153429-4.

MCCARROLL, Christopher Jude, MICHAELIAN, Kourken & NANAY, Bence. Explanatory Contextualism about Episodic Memory: Towards a Diagnosis of the Causalist-Simulationist Debate. *Erkenntnis*, November, 2022. Acesso: https://doi.org/10.1007/s10670-022-00629-4.

MICHAELIAN, Kourken. **Mental time travel**: Episodic memory and our knowledge of the personal past. Cambridge: MIT Press, 2016a.

MICHAELIAN, Kourken. Against Discontinuism: Mental Time Travel and Our Knowledge of Past and Future Events. In: MICHAELIAN, Kourken, KLEIN, Stanley B. & SZPUNAR, Karl K (Eds.). **Seeing the Future**. Oxford: Oxford University Press, 2016b. p. 62–92. Acesso: https://doi.org/10.1093/acprof:oso/9780190241537.003.0004.

MICHAELIAN, Kourken. Against Perrin's Embodied Causalism: Still No Evidence for the Necessity of Appropriate Causation. **Intellectica** 76: 175–91, 2022a.

MICHAELIAN, Kourken. Radicalizing Simulationism: Remembering as Imagining the (Nonpersonal) Past. **Philosophical Psychology**, May, 1–27, 2022b. Acesso: https://doi.org/10.1080/09515089.2022.2082934.

MICHAELIAN, Kourken. Towards a Virtue-Theoretic Account of Confabulation In: SANT'ANNA, André, McCARROLL, Christopher Jude & MICHAELIAN, Kourken (Eds.). **Current controversies in philosophy of memory**. London: Routledge, 2023. p. 127–44.

MICHAELIAN, Kourken, PERRIN, Denis, SANT'ANNA André & SCHIRMER DOS SANTOS, César. Mental Time Travel. In: **The Palgrave Encyclopedia of the Possible**, 1–8. Cham: Springer, 2022. Acesso: https://doi.org/10.1007/978-3-319-98390-5_222-1.

MICHAELIAN, Kourken & ROBINS, Sarah K. Beyond the Causal Theory? Fifty Years after Martin and Deutscher. In: MICHAELIAN, Kourken, DEBUS, Dorothea & PERRIN, Denis (Eds.). **New directions in the philosophy of memory.** New York: Routledge, 2018. p. 13–32. Acesso: https://doi.org/10.4324/9781315159591.

MICHAELIAN, Kourken & SUTTON, John. Memory. **Stanford Encyclopedia of Philosophy**. 2017. Acesso: https://plato.stanford.edu/entries/memory/.

MORAN, Alex. Memory Disjunctivism: A Causal Theory. **Review of Philosophy and Psychology** 13 (4): 1097–1117, 2022. Acesso: https://doi.org/10.1007/s13164-021-00569-y.

NANAY, Bence. Boundary Extension as Mental Imagery. **Analysis**, 81 (4): 647–56, 2022. Acesso: https://doi.org/10.1093/analys/anab023.

NIGRO, Georgia & NEISSER, Ulric. Point of View in Personal Memories. **Cognitive Psychology** 15 (4): 467–82, 1983. https://doi.org/10.1016/0010-0285(83)90016-6.

OKUDA, Jiro, Toshikatsu FUJII, Hiroya OHTAKE, Takashi TSUKIURA, Kazuyo TANJI, Kyoko SUZUKI, Ryuta KAWASHIMA, Hiroshi FUKUDA, Masatoshi ITOH, & Atsushi YAMADORI. Thinking of the Future and Past: The Roles of the Frontal Pole and the Medial Temporal Lobes. **NeuroImage** 19 (4): 1369–80, 2003. Acesso: https://doi.org/10.1016/S1053-8119(03)00179-4.

OPENSHAW, James. (No prelo). (In Defence of) Preservationism and the Previous Awareness Condition: What Is a Theory of Remembering, Anyway? **Philosophical Perspectives**.

OPENSHAW, James, & MICHAELIAN, Kourken. (Em avaliação). "Reference in remembering: Towards a simulationist account."

PERRIN, Denis. Asymmetries in Subjective Time. In: MICHAELIAN, Kourken, KLEIN, Stanley B. e SZPUNAR, Karl K (Eds.). **Seeing the Future**. Oxford: Oxford University Press, 2016. p. 39–61. Acesso: https://doi.org/10.1093/acprof:oso/9780190241537.003.0003.

PERRIN, Denis. 2018. A Case for Procedural Causality in Episodic Recollection. In: MICHAELIAN, Kourken, DEBUS, Dorothea & PERRIN, Denis (Eds.). **New directions in the philosophy of memory.** New York: Routledge, 2018. p. 33–51. Acesso: https://doi.org/10.4324/9781315159591.

PERRIN, Denis. Embodied Episodic Memory: A New Case for Causalism? **Intellectica** 74: 229–52, 2021. https://intellectica.org/en/embodied-episodic-memory-new-case-causalism.

PERRIN, Denis & MICHAELIAN, Kourken. Memory as Mental Time Travel. In: BERNECKER, Sven & MICHAELIAN, Kourken (Eds.) **The Routledge handbook of philosophy of memory**. London: Routledge, 2017. p. 228–39. New York: Routledge. Acesso: https://doi.org/10.4324/9781315687315-19.

PERRIN, Denis, MICHAELIAN, Kourken & SANT'ANNA, André. The Phenomenology of Remembering Is an Epistemic Feeling. **Frontiers in Psychology** 11 (July): 1531, 2020. Acesso: https://doi.org/10.3389/fpsyg.2020.01531.

PERRIN, Denis, & André Sant'Anna. Episodic Memory and the Feeling of Pastness: From Intentionalism to Metacognition. **Synthese** 200 (2): 109, 2022. Acesso: https://doi.org/10.1007/s11229-022-03567-4.

PLATÃO. Teeteto. In: PLATÃO, **Di*á*logos I**. Trad. Edson Bini. São Paulo: Edipro, s/d. p. 41–156.

PUTNAM, Hilary. **Raz*ã*o, verdade e hist*ó*ria**. Lisboa: Dom Quixote, 1992.

PLUNKETT, David & Timothy SUNDELL. Disagreement and the Semantics of Normative and Evaluative Terms. **Philosophers' Imprint** 13 (23): 1–37, 2013.

PLUNKETT, David, STERKEN, Rachel Katharine & SUNDELL, Timothy. Generics and Metalinguistic Negotiation. **Synthese** 201 (2): 50. Acesso: 2023. https://doi.org/10.1007/s11229-022-03862-0.

RIVADULLA-DURÓ, Andrea. The Simulation Theory of Memory and the Phenomenology of Remembering. **Phenomenology and the Cognitive Sciences**, December, 2022. Acesso: https://doi.org/10.1007/s11097-022-09881-z.

ROBINS, Sarah K."Misremembering. **Philosophical Psychology** 29 (3): 432–47, 2016. Acesso: https://doi.org/10.1080/09515089.2015.1113245.

ROBINS, Sarah K."Confabulation and Constructive Memory. **Synthese** 196 (6): 2135–51, 2019. Acesso: https://doi.org/10.1007/s11229-017-1315-1.

ROBINS, Sarah K. Defending Discontinuism, Naturally. **Review of Philosophy and Psychology** 11 (2): 469–86, 2020. Acesso: https://doi.org/10.1007/s13164-020-00462-0.

ROBINS, Sarah K. Episodic Memory Is Not for the Future. In: SANT'ANNA, André, McCARROLL, Christopher Jude & MICHAELIAN, Kourken (Eds.). **Current controversies in philosophy of memory**. London: Routledge, 2023. p. 149–65.

ROEDIGER III, Henry L. & MCDERMOTT, Kathleen B."Creating False Memories: Remembering Words Not Presented in Lists. **Journal of Experimental Psychology: Learning, Memory, and Cognition** 21 (4): 803–14, 1995. Acesso: https://doi.org/10.1037/0278-7393.21.4.803.

RUSSELL, Bertrand. **A análise da mente**. Rio de Janeiro: Zahar, 1976.

SANT'ANNA, André. The Hybrid Contents of Memory. **Synthese** 197 (3): 1263–90, 2020. Acesso https://doi.org/10.1007/s11229-018-1753-4.

SANT'ANNA, André. Mnemonic causation, construction, and the particularity of episodic memory. **Aufklarung**, 8: 57-70, 2021a. Acesso: https://doi.org/10.18012/arf.v8iesp.60017

SANT'ANNA, André. Attitudes and the (dis)continuity between memory and imagination. **Estudios de Filosofía**, 64: 73-93, 2021b. Acesso: https://doi.org/doi.org/10.17533/udea.ef.n64a04.

SANT'ANNA, André. (No prelo). Metacognition and the puzzle of alethic memory. **Philosophy and the Mind Sciences.**

SANT'ANNA, André & MICHAELIAN, Kourken. Teorias sobre o lembrar: causalismo, simulacionismo e funcionalismo. **Voluntas** 10 (3): 8, 2019a. Acesso: https://doi.org/10.5902/2179378640445.

SANT'ANNA, André, & Kourken MICHAELIAN. Thinking about events: A pragmatic account of the objects of episodic hypothetical thought. **Review of Philosophy and Psychology**, 10(1):187–217, 2019b. Acesso: https://doi.org/10.1007/s13164-018-0391-6.

SCHACTER, Daniel L. & ADDIS, Donna Rose. The Cognitive Neuroscience of Constructive Memory: Remembering the Past and Imagining the Future. **Philosophical Transactions of the Royal Society B: Biological Sciences** 362 (1481): 773–86, 2007. Acesso: https://doi.org/10.1098/rstb.2007.2087.

SCHACTER, Daniel L. ADDIS, Donna Rose & BUCKNER, Randy L."Episodic Simulation of Future Events: Concepts, Data, and Applications. **Annals of the New York Academy of Sciences** 1124 (1): 39–60, 2008. Acesso: https://doi.org/10.1196/annals.1440.001.

SCHACTER, Daniel L. ADDIS, Donna R., HASSABIS, Demis, MARTIN, Victoria C. R., SPRENG Nathan & SZPUNAR, Karl K. The Future of Memory: Remembering, Imagining, and the Brain. **Neuron** 76 (4): 677–94, 2012. Acesso: https://doi.org/10.1016/j.neuron.2012.11.001.

SCHIRMER DOS SANTOS, César. O debate causalismo versus simulacionismo em filosofia da memória como negociação metalinguística. **Perspectiva Filosófica** 46 (2), 2019. Acesso: https://doi.org/10.51359/2357-9986.2019.248088.

SCHIRMER DOS SANTOS, César, McCARROLL, Christopher Jude, & SANT'ANNA, André. The Relation between Memory and Imagination: A Debate about the Right Concepts. In: SANT'ANNA, André, McCARROLL, Christopher Jude e MICHAELIAN, Kourken (eds.). **Current controversies in philosophy of memory**. London: Routledge, 2023. p. 38–56.

SCHWARTZ, Arieh."Simulationism and the Function(s) of Episodic Memory. **Review of Philosophy and Psychology** 11 (2): 487–505, 2020. Acesso: https://doi.org/10.1007/s13164-020-00461-1.

SENOR, Thomas D."The Epistemology of Episodic Memory. In: SANT'ANNA, André, McCARROLL, Christopher Jude & MICHAELIAN, Kourken (Eds.). **Current controversies in philosophy of memory**. London: Routledge, 2023. p. 227–43. Acesso: https://doi.org/10.4324/9781003002277-18.

SUDDENDORF, Thomas & CORBALLIS, Michael C. Mental Time Travel and the Evolution of the Human Mind. **Genetic, Social & General Psychology Monographs** 123 (2): 133–67, 1997. Acesso: https://www.researchgate.net/publication/292514522_Mental_Time_Travel_and_the_Evolution_of_the_Human_Mind.

SUDDENDORF, Thomas & CORBALLIS, Michael C. The Evolution of Foresight: What Is Mental Time Travel, and Is It Unique to Humans? **Behavioral and Brain Sciences** 30 (3): 299–313, 2007. Acesso: https://doi.org/10.1017/S0140525X07001975.

TULVING, Endel."Episodic and Semantic Memory. In: TULVING, Endel & DONALDSON, Wayne (Eds.). **Organization of Memory**. New York: Academic Press, 1972. p. 381–402.

TULVING, Endel. Memory and Consciousness. **Canadian Psychology / Psychologie Canadienne** 26 (1): 1–12, 1985. Acesso : https://doi.org/10.1037/h0080017.

VON LEYDEN, W. **Remembering**: A philosophical problem. New York: Philosophical Library, 1961.

WERNING, Markus. Predicting the Past from Minimal Traces: Episodic Memory and Its Distinction from Imagination and Preservation. **Review of Philosophy and Psychology** 11 (2): 301–33, 2020. Acesso: https://doi.org/10.1007/s13164-020-00471-z.

WILLIAMSON, Timothy. **Knowledge and its limits**. Oxford: Oxford University Press, 2000.

WRINCH, Dorothy. "On the Nature of Memory." **Mind** 29 (113): 46–61, 1920.

A MEMÓRIA NA PSICOLOGIA DE RIBOT: A NECESSIDADE DE MUDANÇA E RENOVAÇÃO

Wilson Antonio Frezzatti Jr.[1]

Resumo:

O filósofo e psicólogo francês Théodule Ribot, na segunda metade do século XIX, rejeitava a psicologia metafísica (o "estudo da alma") e propunha uma nova psicologia: científica, fisiológica, experimental e evolucionista. A verdadeira causalidade dos fenômenos psíquicos seria fisiológica. Este artigo, tendo como pano de fundo os processos mnemônicos, pretende mostrar a importância, para Ribot, da mudança e da renovação nas funções psicológicas. A memória é fundamentalmente biológica, múltipla e inconsciente, e está estreitamente ligada à hereditariedade. Se não forem contrabalançadas pelo esquecimento, as memórias se associam firmemente às estruturas nervosas, se consolidam e, em casos em que o indivíduo esteja submetido sempre aos mesmos estímulos, a consciência desaparece e o automatismo

Abstract:

In the second half of the 19th century, the French philosopher and psychologist Théodule Ribot rejected metaphysical psychology (the "study of the soul") and proposed a new psychology: a scientific, physiological, experimental and evolutionary psychology. The real causality of psychic phenomena is physiological. This article, against the background of mnemonic processes, aims to show the importance of change and renewal in psychological functions according Ribot. Memory is fundamentally biological, multiple and unconscious, and is closely linked to heredity. If the memories are not counterbalanced by forgetfullness, they become firmly associated with nervous structures, are consolidated, and, in cases where the individual is always subjected to the same stimuli, the

1 Professor associado dos cursos de graduação e pós-graduação (mestrado e doutorado) em Filosofia da Universidade Estadual do Oeste do Paraná (UNIOESTE). Professor do Mestrado em Filosofia da Universidade Estadual de Maringá (UEM). Coordenador do GEN (Grupo de Estudos Nietzsche). Membro do Grupo Internacional *HyperNietzsche* e do GT-Nietzsche (ANPOF). Autor dos livros: *Nietzsche contra Darwin* (2001; 2014; 2022); *A Fisiologia de Nietzsche: a Superação da Dualidade Cultura/Biologia* (2006; 2022); e *Nietzsche e a psicofisiologia francesa do século XIX* (2019). ORCID: https://orcid.org/0000-0002-7519-3789. E-mail: wfrezzatti@uol.com.br

completo se instala. Para o ser humano não se tornar uma máquina, são necessárias mudança e renovação em suas atividades.

Palavras-chave:

Ribot. Memória. Evolução. Psicofisiologia.

consciousness disappears and the complete automatism is established. In order for the human being not to become a machine, change and renewal in his activities are necessary.

Keywords:

Ribot. Memory. Evolution. Psychophysiology.

Introdução

A luta da psicologia, enquanto disciplina científica, por sua autonomia em relação à filosofia tem, na França da segunda metade do século XIX, o filósofo Théodule Ribot como uma figura central. O embate principal ocorria entre a filosofia espiritualista, encabeçada por Victor Cousin e Théodore-Simon Jouffroy, e a proposta de uma nova psicologia: científica, experimental e fisiológica[2]. A psicologia era considerada pelos filósofos espiritualistas o principal ramo da filosofia, sendo que seu objeto por excelência seria a alma. Seu método, que constituiria a consciência como o único campo de observação do filósofo, seria a observação interior, isto é, a inspeção da consciência. Segundo Cousin (1833, p. 12): "Entrar na consciência e estudar escrupulosamente todos os seus fenômenos, suas diferenças e suas relações, esse é o primeiro estudo do filósofo; seu nome científico é *psicologia*. A psicologia é, portanto, a condição e como o vestíbulo da filosofia".

Contra a psicologia metafísica – a velha psicologia – e sua noção de alma e de faculdades, Ribot propõe uma psicologia científica, experimental e fisiológica – a nova psicologia. A alma não é causa e as suas faculdades não são sedes invariáveis das funções psicológicas.

Dentro desse contexto, o objetivo deste artigo é mostrar a importância da mudança e renovação nos processos fisiológicos em Ribot, tendo como pano de fundo as funções mnemônicas sob uma perspectiva fisiológica e evolucionista. Embora, para o psicólogo

2 Sobre esse tema, cf. NICOLAS, 2002, p. 57-139 e FREZZATTI, 2019, p. 45-72.

francês, memória seja hereditariedade, ela não é infalível, devendo haver espaço para as mudanças nos seres vivos, especialmente no ser humano. Caso contrário, a consciência desaparecerá e nos transformaremos em autômatos.

Ribot e a nova psicologia

Ribot foi aluno de Cousin e, posteriormente, tornar-se-ia positivista dissidente. Ele realizou um enorme movimento contra a psicologia metafísica, declarando a necessidade de uma psicologia científica, fisiológica e evolucionista. O positivismo dissidente diferencia-se do pensamento de Auguste Comte, entre outras características, pelo papel atribuído à psicologia. Também ferrenho opositor à psicologia metafísica, Comte (1973, p. 15), em *Curso de filosofia positiva* (*Cours de philosophie positive*, 1830-1842), classifica as ciências por seu valor experimental, e as cinco principais, em ordem crescente de importância e em sequência histórica, são: a astronomia, a física, a química, a fisiologia e a física social (sociologia). A psicologia é rejeitada pelo caráter estritamente abstrato, ou seja, pelo caráter metafísico de seu objeto (a alma) e de seu método (observação interior). Na escola positivista dissidente[3], uma psicologia fundamentada na fisiologia, a *physiologie psychique*, e um uso limitado do método de observação interior constroem uma "nova psicologia", científica, experimental e autônoma.

A tese de doutorado de Ribot, defendida na Universidade da Sorbonne, é a primeira a tratar da psicologia científica: *A hereditariedade: Estudo psicológico sobre seus fenômenos, suas leis, suas causas, suas consequências* (*L'Hérédité: Étude psychologique sur ses phénomènes, ses lois, ses causes, ses conséquences*, 1873). Em 1885, ministra um curso de psicologia experimental na mesma universidade, mas não consegue mais repeti-lo. Nesse mesmo ano, Ribot e Jean-Martin Charcot instituem a *Société de Psychologie Physiologique*. Em 1888, assume a cátedra de *Psychologie expérimentale et comparée* no *Collège de France*[4].

Entre os livros de Ribot, temos aqueles que criticam a "velha psicologia" metafísica e propagam a necessidade de uma "nova psicologia" fisiológica e aqueles que desenvolvem

3 Os principais positivistas dissidentes eram Émile Littré, Hippolyte Taine e Ribot.

4 Ribot é considerado o pai da psicologia científica francesa sem ter feito um único experimento nem uma única prática clínica. O primeiro laboratório de psicologia foi estabelecido em 1889, na Sorbonne, por Étienne-Henry Beaunis, também membro fundador da *Société de Psychologie Physiologique*. Apesar disso, nos textos de Ribot, casos ou relatos clínicos e resultados de experimentos neurofisiológicos têm uma grande presença.

um procedimento psicológico objetivo, o método patológico. Em todos eles, Ribot articula suas ideias em torno do associacionismo inglês, da fisiologia experimental alemã e do evolucionismo de Herbert Spencer. *A psicologia inglesa contemporânea* (*La psychologie anglaise contemporaine*, 1870) e *A psicologia alemã contemporânea* (*La psychologie allemande contemporaine*, 1879) são os principais livros do primeiro grupo.

A metafísica, Ribot (1870, p. 1-30) afirma, não deve ser a base da psicologia, pois a alma não pode ser o objeto de uma ciência, e o método da observação interior não pode ser o único caminho de investigação. A psicologia científica não deve se ocupar de causas primeiras, mas apenas dos fenômenos psicológicos[5], conscientes e inconscientes; seu principal método é o da experimentação fisiológica. Enfim, a psicologia deve separar-se da filosofia, a qual se tornará cada vez mais abstrata, resumindo-se a especulações metafísicas e afastando-se totalmente da realidade.

A autonomia da psicologia seria garantida pela fisiologia experimental (cf. RIBOT, 1879, p. I-XXXIII). O método experimental rigoroso investiga as variações dos fenômenos psicológicos e não a consciência ou a alma, as quais, segundo o psicólogo francês, seriam essências abstratas com faculdades imaginárias. Ribot, assim, determina o princípio básico da *psychologie physiologique*: "todo estado psicológico determinado está ligado a um ou vários acontecimentos físicos determinados" (RIBOT, 1879, p. XI). Dessa forma, o psicólogo francês pensa em expandir, de modo objetivo, os limites da "antiga psicologia": ao estudar a alma humana, ela abordava apenas o homem – e não os animais –, o adulto, o branco e o "civilizado".

O método patológico de Ribot

Sua trilogia sobre doenças psicológicas tem também o objetivo de estabelecer um método na psicologia, o método patológico (cf. FREZZATTI, 2019, p. 121-34). A falta de um método experimental na antiga psicologia impediu que houvesse uma psicologia comparada que utilizasse a noção de progresso. A patologia, segundo Ribot (1881, p. 51), deve completar e confirmar os resultados da fisiologia normal[6]. As anomalias psicológicas

5 Para Ribot, os fenômenos psicológicos não diferem qualitativamente dos fenômenos fisiológicos e dos físico-químicos. A realidade seria constituída por uma sequência contínua, sem limites definidos, e de grau de complexidade crescente: fenômenos físico-químicos, fisiológicos, psicológicos e culturais (incluindo a moral) (cf. FREZZATTI, 2019, p. 103-6).

6 Ribot propõe uma psicologia composta de três partes interdependentes: 1. Psicologia geral: produz os

são muito preciosas para o psicólogo francês, pois são experimentos refinados propiciados pela natureza (cf. RIBOT, 1870, p. 31). Isso ocorre porque, enquanto a evolução progride de estruturas mais simples para as mais complexas, as morbidades promovem a dissolução destas últimas.

Em outras palavras, as doenças nos revelam os elementos fisiológicos mais simples: elas agem no sentido inverso ao da evolução. Esses são os princípios norteadores de *As doenças da memória* (*Les maladies de la mémoire*, 1881), *As doenças da vontade* (*Les maladies de la volonté*, 1883) e *As doenças da personalidade* (*Les maladies de la personnalité*, 1885)[7]. Ao abordar a memória, a vontade e a personalidade por meio do método patológico, Ribot rejeita as perspectivas metafísicas sobre elas[8], que passam a ser fenômenos surgidos historicamente e não são mais consideradas causas primeiras: são multiplicidades e não unidades, são relações e não essências.

Assim, por exemplo, a noção metafísica de eu é vista como obscura. O eu teria sua origem nas formas inferiores de vida, o que exige uma investigação histórica e evolutiva, e não uma inspeção da própria consciência, o método de observação interna, como propunham os filósofos espiritualistas. Quanto mais elevado um organismo, mais elevada será a forma de sua individualidade. O Sujeito não é causa una das ações humanas, mas o resultado de inúmeros processos nervosos, cujo elemento mais simples é o arco reflexo, constituído, por sua vez, de fenômenos físico-químicos. A consciência não é o fundamento dos fenômenos psicológicos, ela é apenas um complemento sobreveniente, pois, em sua imensa maioria, os processos nervosos são inconscientes. Afirma Ribot (1885, p. 18): "É verossímil que a consciência tenha sido produzida como qualquer outra manifestação vital, no início sob uma forma rudimentar e aparentemente sem grande eficácia". Portanto, a consciência não é algo exclusivo do ser humano nem transcendente.

Em outras palavras, a consciência não é a essência, a propriedade fundamental da alma, mas um acontecimento complexo que supõe um estado particular do sistema nervoso (cf. RIBOT, 1881, p. 22-5). É a ação nervosa que é a condição fundamental e

---

fundamentos da psicologia por meio da investigação empírica dos fenômenos psicológicos; 2. Psicologia comparada: investigação das estruturas psicológicas pela perspectiva do progresso; e 3. Teratologia científica ou Psicologia das morbidades: a investigação das anomalias (cf. RIBOT, 1870, p. 31-7).

7 Cada uma dessas obras teve inúmeras edições desde o seu lançamento. Encontramos até 1921 26 edições para *As doenças da memória*; 32, para *As doenças da vontade*; e 18, para *As doenças da personalidade*. A editora L'Harmattan vem publicando fac-símiles das obras de Ribot.

8 Com a memória, Ribot aborda a consciência, e com a personalidade, a noção de sujeito.

não a consciência, ou seja, não é a consciência que constitui o fenômeno psíquico: sem a consciência, o orgânico permanece – a recíproca não é verdadeira. Toda ação psíquica pressupõe uma ação nervosa, isto é, a causalidade é efetivamente física (inconsciente) e não psicológica (consciente).

Para Ribot (1881, p. 22-3), há duas condições para a existência da consciência: 1. A intensidade: nossos estados de consciência lutam sem cessar entre si, e um estado intenso pode decair até atingir o limiar da consciência e, assim, tornar-se inconsciente[9]; e 2. A duração: cada ação psíquica consciente requer uma duração de tempo apreciável. Se ela tiver duração inferior à requerida para a consciência, a ação permanece inconsciente.[10]

Se considerarmos, segundo o psicólogo francês, a base inconsciente fisiológica de todo o fenômeno psicológico, acontecimentos tais como lembranças repentinas, soluções que nos aparecem subitamente, invenções poéticas e científicas, simpatias e antipatias secretas deixam de ser misteriosos.

A memória é entendida por Ribot no contexto teórico descrito acima.

A memória como fato biológico

O livro *As doenças da memória* está dividido em quatro capítulos mais a conclusão. O primeiro capítulo, "A memória como fato biológico", trata da memória em suas questões gerais: sua condição fundamentalmente biológica e inconsciente; seu mecanismo; suas condições; seu caráter ilusório; o papel axial do esquecimento; sua tendência a produzir automatismo; etc. Os outros capítulos abordam efetivamente as morbidades, seja por falta ou excesso: o segundo capítulo, "As amnésias gerais", no qual Ribot apresenta a importância metodológica das doenças, conforme mencionado mais acima, e a lei da regressão, que governa a destruição da memória; o terceiro, "As amnésias parciais"; e o último, "As exaltações da memória ou hipermnésias".

9 Nessa concepção de luta entre ações psíquicas, Ribot segue Herbart: cf. RIBOT, 1879, p. 1-34.

10 Ribot rejeita a ideia de que a velocidade do pensamento é infinita e que há vários pensamentos ao mesmo tempo na mente. A velocidade é determinada, e os pensamentos ocorrem sequencialmente. Ele nos fornece alguns tempos de percepção, advertindo que esses valores variam muito de acordo com as condições: som, 0,14- 0,16 s; tato, 0,18-0,21 s; luz, 0,20-0,22 s (cf. RIBOT, 1881, p. 23).

Na apresentação da obra, Ribot (1881, n. p.) declara a importância de sua investigação: apesar de haver muitos textos sobre a memória, há muito poucos sobre sua patologia. Ele considera seu próprio texto um ensaio de psicologia descritiva ou mesmo de história natural, o que justifica seu posicionamento no início do primeiro capítulo: a utilização do "novo método", a saber, a fisiologia experimental e o evolucionismo, para conhecer a natureza da memória (cf. RIBOT, 1881, p. 1). Enquanto a antiga psicologia, segundo Ribot, toma a memória como uma faculdade da alma, como algo inteiramente consciente, a nova psicologia a vê principalmente como uma função orgânica, portanto inconsciente: "A memória é, por essência, um fato biológico; por acidente, um fato psicológico" (RIBOT, 1881, p. 1). Como função biológica, ela tem uma história.

Mas a partir de qual ser – biológico ou psicológico – começar a investigação histórica? Embora haja nos fenômenos orgânicos processos análogos à memória, como, por exemplo, a fotografia, eles são passivos, dependentes de agentes externos e muito distantes dos seres orgânicos (cf. RIBOT, 1881, p. 3-5)[11]. Entre os seres orgânicos, Ribot descarta também os vegetais, mas aponta o tecido muscular como um primeiro exemplo da aquisição, conservação e reprodução automática de novas propriedades: quanto mais se exercita, mais forte o músculo fica; a cada ação, ele está mais disposto à repetição do mesmo trabalho[12]. Aquisição, conservação e reprodução automática são, para Ribot, as propriedades fundamentais da memória.

Do músculo, Ribot (1881, p. 5-11) passa para o que chama de "o tecido mais elevado do organismo", o tecido nervoso. Para o psicólogo francês, não basta pesquisar o arco reflexo, apesar de ele ser um tipo de memória fixada por hereditariedade, pois é uma ação geral. O que se procura são fenômenos mais específicos: os movimentos automáticos preencheriam esse critério da especificidade da ação. Esses movimentos seriam de dois tipos: as ações automáticas primitivas, inatas, e as ações automáticas secundárias, adquiridas. Os atos primitivos de hoje foram os adquiridos ontem, tendo sido fixados pela formação de associações com reflexos primitivos, nas quais houve incorporação de uns e exclusão de outros reflexos. O exercício é o responsável pela fixação. Por exemplo, quando aprendemos a andar, nos elementos nervosos dos órgãos motores, formam-se associações dinâmicas, secundárias mais ou menos estáveis, as quais se juntam às associações anatô-

11 Nesse comentário, Ribot já nos mostra o caráter ativo de sua concepção de memória.

12 Na abordagem sobre o músculo, Ribot baseou-se em *Sobre a memória como função geral da matéria organizada (Über das Gedächtniss als Allgemeine Function der organisirten Materie*, 2ª ed., 1876), do fisiologista alemão Ewald Hering.

micas, primitivas e permanentes. No início desse processo de aprendizagem, a consciência acompanhava a atividade motora, mas, com a fixação pelo exercício, o movimento torna-se inconsciente. Desse modo, temos uma memória orgânica que nos permite andar, semelhante à memória psicológica, cuja característica específica é a consciência. Ambas as memórias têm o mesmo mecanismo de aquisição, conservação e reprodução.

Além disso, há outro aspecto central na concepção de memória de Ribot: não há uma memória una, mas memórias[13]. Ainda considerando o exemplo do desenvolvimento do andar, é necessária a fixação de várias habilidades específicas, e cada parte do corpo envolvida tem sua memória particular. Assim, não há uma sede única da memória, ao contrário, existem sedes específicas para cada tipo de memória. As lembranças não estão na alma, mas fixadas em seu lugar de surgimento em alguma estrutura do sistema nervoso. Dessa maneira, no mesmo indivíduo, o desenvolvimento desigual dos diversos sentidos e diversos órgãos produz modificações desiguais nas partes do sistema nervoso e, em consequência, condições desiguais de recordação e, portanto, variações da memória (cf. RIBOT, 1881, p. 109-10).

Uma boa memória visual, por exemplo, tem como condição uma boa estrutura do olho, do nervo óptico e das partes do encéfalo (protuberância, pedúnculos ópticos e hemisférios cerebrais) envolvidas no ato da visão. Quanto melhores forem essas condições, melhores serão as memórias visuais. Em suma, as condições fisiológicas da memória são as seguintes: 1. uma modificação particular impressa nos elementos nervosos; e 2. uma associação ou conexão particular estabelecida entre um certo número de elementos.

Memória e hereditariedade: leis físico-químicas

Ao rejeitar a explicação dos fenômenos da memória por uma faculdade hipotética como a consciência, Ribot lança mão de leis físico-químicas. Em *A hereditariedade*,

13 Ribot (1881, p. 107) cita, do filósofo inglês George Henry Lewes, *Problemas da vida e da mente* (*Problems of Life and Mind*, 1879): "O antigo e ainda não refutado erro que trata a memória como uma função ou faculdade independente, para a qual se procura um órgão ou uma sede, tem origem na tendência constantemente presente de personificar uma abstração. Ao invés de reconhecê-la como uma expressão abreviada para o que é comum a todos os fatos da lembrança ou para a soma de tais fatos, muitos autores supuseram que ela possuía uma existência independente" (LEWES, 1879, p. 119)*. Ainda segundo Ribot (1881, p. 111), embora na filosofia ainda se considere a memória como uma unidade, a distinção entre as memórias é corrente na fisiologia.

(*) Essa citação foi traduzida por nós diretamente do texto inglês original.

o psicólogo francês associa as leis de memória às leis da indestrutibilidade da força e da conservação de energia – "as mais gerais que regem o universo" (RIBOT, 1873, p. 68). Isso é possível pelo fato de os fenômenos psicológicos, incluindo os morais e culturais, não ocorrerem ao acaso e sem leis (cf. RIBOT, 1873, p. 69). Nada surge do nada, e o que existe não pode se tornar nada: tanto na ordem física quanto na psicológica. Assim, nossas percepções e ideias transformam-se, mas são indestrutíveis, relacionando-se por meio de uma dinâmica de forças. Fortemente apoiado no filósofo alemão Johann Friedrich Herbart, Ribot afirma:

> Toda ideia que ocupa a consciência só pode ser deslocada por uma ideia mais forte. Se duas forças mentais que lutam para ocupar a consciência são semelhantes e agem na mesma direção, seus resultados se combinam, produzindo um estado de consciência muito intenso. Se duas forças são iguais e contrárias, ocorre o equilíbrio. Se duas forças são diferentes e contrárias, uma restringe a outra; e, ocorrendo isso, perdem a parte de sua própria força equivalente ao que desloca (RIBOT, 1873, p. 73).

Essa luta de forças para a emergência na consciência produz uma interessante propriedade para o estado inconsciente: "a existência de ideias no inconsciente poderia [...] ser considerada um estado de equilíbrio perfeito" (RIBOT, 1873, p. 74). O surgimento de uma nova ideia na consciência representaria a quebra desse equilíbrio[14].

Assim, uma recordação remete à grande lei universal da conservação da força. Em um domínio menos geral, o da vida, essa lei assume um aspecto mais específico: a lei biológica do hábito (cf. RIBOT, 1873, p. 75-6), isto é, a repetição de uma ideia torna-a mais fixada no organismo, tendendo ao automatismo. Para Ribot: "a memória é apenas uma forma do hábito" (RIBOT, 1873, p. 75). No entanto, há uma ressalva nessa proposição: hábito e memória não são completamente coincidentes, pois o primeiro é completamente inconsciente, enquanto a segunda pode ser consciente ou inconsciente.

A memória, portanto, na concepção de Ribot (1881, p. 46-8), é um processo de organização fisiológica em graus variados, compreendido entre dois extremos: um estado novo e o registro orgânico. Esse desenvolvimento tem como modelo os movimentos automáticos. Ele começa por uma aquisição nova na mente que é reavivada uma ou duas vezes[15]. Essas lembranças são instáveis e podem desaparecer se não forem reativadas. A

14 Nessa abordagem, Ribot cita os seguintes autores: Herbert Spencer (*Princípios de psicologia*), Hippolyte Taine (*Da inteligência*) e Johannes Müller (*Manual de fisiologia*).

15 Baseado em Wilhelm Wundt, Ribot aponta que, tanto na percepção quanto na lembrança, a operação nervosa é a mesma.

imensa maioria dos fatos que nos acontecem acabam desaparecendo da memória, a não ser que eles sejam reativados, voluntária ou involuntariamente, com certa frequência, o que resulta numa maior organização, ou seja, maior associação com os outros estados nervosos e num aumento de sua estabilidade. Com a repetição, a lembrança torna-se mais impessoal, mais objetiva e a localização no tempo torna-se tênue até desaparecer. Cada vez mais, ela sai da esfera psíquica (consciente) para se transformar em memória orgânica (inconsciente). Enfim, uma memória completamente organizada, inconsciente e hereditária se constitui[16]. Tal transformação é a que ocorreria durante o aprendizado de uma língua ou de um instrumento musical.

A hereditariedade, para Ribot, é uma memória da espécie. Ela é para a espécie o que a memória propriamente dita é para o indivíduo. Nos dois casos, a base é sempre biológica (cf. RIBOT, 1873, p. 77). Assim, a memória orgânica é altamente organizada, pois é fortemente associada às estruturas nervosas; inconsciente, pois puramente fisiológica, sem consciência; e hereditária, porque adquirida e incorporada ao organismo.

Memória e esquecimento: condições da vida saudável

Sendo a memória um fato biológico fundamentado, em última instância, em fenômenos físico-químicos, a sua capacidade de armazenamento não é infinita, mas limitada (cf. RIBOT, 1881, p. 46). Como já vimos, há uma luta entre os estados da consciência para permanecerem conscientes, e mesmo entre os estados inconscientes para se tornarem conscientes. Assim, o esquecimento tem um papel importantíssimo nos processos mnemônicos: ele é a própria condição da memória. Sem o esquecimento total de um imenso número de estados de consciência e o esquecimento momentâneo de um grande número deles, não poderíamos nos lembrar. Para Ribot, o esquecimento, com exceção de alguns casos, não é uma doença da memória, mas condição de saúde e da própria vida.

16 Os reflexos nervosos organizados que compõem a memória orgânica são, por sua vez, complexos formados por reflexos simples. Estes últimos, exatamente por serem reflexos simples, são anatomicamente inatos, sendo eles próprios anteriormente adquiridos e fixados pelas inúmeras experiências na evolução das espécies. Desse modo, a memória individual transforma-se em memória da espécie, que é transmitida hereditariamente (cf. RIBOT, 1881, p. 46). Ribot, como muitos em sua época, pensava como plenamente efetiva a transmissão dos caracteres adquiridos.

A memória está ligada às condições fundamentais da vida[17]. E vida, segundo o psicólogo francês, é adquirir e perder, isto é, assimilação e desassimilação (cf. RIBOT. 1881, p. 46-7, 50-1 e 101). E esquecer é desassimilar. Toda forma de memória pressupõe associações dinâmicas entre os elementos nervosos e as suas modificações particulares, as quais não ocorrem em matéria inerte, mas em matéria viva, que se renova continuamente. Para que a modificação persista, é necessário que o arranjo das novas moléculas reproduza exatamente aquele que é substituído. O fluxo de renovação das moléculas no organismo é determinado pela nutrição. Além disso, as células também se reproduzem, e a reprodução ou geração é uma forma de nutrição[18]. Portanto, a memória depende diretamente da nutrição.

Não obstante, há aparentemente um problema aqui: se a substituição é, a princípio exata, como explicar as transformações progressivas ou decadentes dos seres vivos? No âmbito científico, a hereditariedade é uma lei, ou seja, é constante e se repete – semelhante produz semelhante: "A hereditariedade é uma lei da natureza viva, uma lei biológica, fatal e necessária, como as leis físicas, um princípio de conservação e estabilidade" (Ribot, 1873, p. 513). Todavia, de fato, no processo vital, as substituições nem sempre são exatas. Por ser uma lei do mundo vivo, a hereditariedade, apenas em condições ideais, realiza uma repetição constante das mesmas características[19]. Em outras palavras, a lei da heredi-

17 Segundo Ribot (1881, p. 47), a memória, por sua íntima relação com as funções fundamentais da vida, seria um dos melhores testemunhos em favor da teoria da evolução. Seu estudo, assim, não deveria ser apenas uma fisiologia, mas também uma morfologia, ou seja, uma história das transformações.

18 Ribot considera a nutrição o processo vital por excelência. Ela não se faz em um instante, o que significa que a fixação da memória necessita de tempo para ocorrer (cf. RIBOT, 1881, p. 157-9). A memória é, afinal, uma impregnação biológica, e a fadiga é fatal a ela. Alguns biólogos do século XIX consideravam que a reprodução seria uma forma de nutrição, pois, se a nutrição promove o crescimento do indivíduo, a reprodução é uma forma de crescimento em um organismo que já atingiu o tamanho determinado por sua espécie. Sobre isso, cf. HAECKEL, 1924, p. 269; ROUX, 1881, p. 213-6 e 223-30; e também o filósofo SPENCER, 1864, p. 224-37.

19 Ribot reafirma essa concepção de que, por princípio, a substituição de moléculas pela assimilação ocorre de modo exato, ao citar o patologista inglês James Paget (cf. RIBOT, 1881, p. 159), que também tem essa ideia: "Como pode o cérebro ser o órgão da memória, se se supõe que sua substância sempre muda? Ou como é que essa presumida mudança nutritiva de todas as partículas do cérebro não destrói toda a memória e todo o conhecimento das coisas sensíveis, como ocorre com uma súbita destruição provocada por alguma grande injúria? A resposta é, - devido à exatidão da assimilação efetuada no processo formativo: o efeito uma vez produzido por uma impressão no cérebro, seja uma percepção ou um ato intelectual, é fixado e aí retido, porque a parte, seja ela qual for, que foi assim modificada é exatamente representada pela parte que a sucede no curso da nutrição" (PAGET, 1853, p. 53)*. O psicólogo francês reafirma a analogia de Paget entre a memória e uma doença infecciosa: "Tão paradoxal que possa parecer uma aproximação entre uma doença infecciosa e a memória, ela é, portanto, perfeitamente exata do ponto de vista biológico" (RIBOT, 1881, p. 159).
(*) Essa citação foi traduzida por nós diretamente do texto inglês original.

tariedade não é invariável, certa e absolutamente necessária, já que, se todas suas condições não estiverem presentes, ela não se efetiva, produzindo variações. As relações vitais são extremamente complexas e mutáveis, e várias leis se sobrepõem e atuam umas sobre as outras. Enfim, a semelhança torna-se apenas aproximada.

Portanto, a substituição de moléculas e células não é sempre feita de modo perfeito. A dinâmica de assimilação e desassimilação, com sua seleção de certas percepções e associações, faz com que a recordação do passado não seja exata. O que retemos na memória não é o que ocorreu exatamente, uma vez que não retemos todos os detalhes e o que retemos dá lugar a outras lembranças. A memória, assim sendo, tem um caráter ilusório.

Ao estudar as amnésias completas, Ribot (1881, p. 94-5) propõe que a destruição total da memória começa pelas lembranças recentes, mal fixadas nos elementos nervosos e raramente repetidas, isto é, por aquelas fracamente associadas a outras e muito pouco organizadas[20]. O processo mórbido termina na memória sensorial, instintiva, fixada no organismo e que se torna o próprio corpo. O psicólogo francês propõe uma lei para essa dinâmica, a lei da dissolução da memória: a destruição progressiva da memória "desce progressivamente do instável para o estável" (RIBOT, 1881, p. 94). Essa lei é um caso particular de uma outra ainda mais universal referente à vida: a lei da regressão ou reversão[21].

No caso do ser humano, o esquecimento é importantíssimo, pois, em uma marcha contínua em direção à organização, uma simplificação torna possível uma forma de pensamento mais elevada. Se a memória apenas crescesse em organização, sem nenhum contrabalanceamento, haveria a aniquilação progressiva da consciência: o homem tornar-se-ia um autômato. Se formos obrigados a permanecer em uma situação em que falte qualquer estado novo de consciência, isto é, percepções, ideias, imagens, sentimentos, desejos, etc., perderíamos nossa consciência. Segundo Ribot (1881, p. 50), mesmo aqueles que caem em uma rotina realizam isso de certa maneira: descartando o novo e o imprevisto, eles tendem à estabilidade perfeita – tornam-se máquinas.

---

20 As lembranças não se depositariam nos tecidos cerebrais como camadas geológicas, as quais a doença vai extraindo (cf. RIBOT, 1881, p. 100-101). Elas ocupam o mesmo lugar anatômico que as impressões primitivas, exigindo a atividade das células nervosas, ou seja, podem ocupar desde o córtex cerebral até a medula espinal.

21 Na seção "A hereditariedade como causa de decadência" de *A hereditariedade*, Ribot considera que, ao lado do progresso, pelo qual os seres vivos melhoram e se tornam superiores, temos também o enfraquecimento e o declínio (cf. RIBOT, 1873, p. 420-426). Assim, a hereditariedade, por ser uma tendência conservadora, pode fixar tanto o progresso quanto a decadência adquiridos. Trata-se do próprio processo vital: tudo que vive declina e morre, seja o indivíduo, o povo ou a própria humanidade, e suas causas são sempre fisiológicas ou orgânicas.

Considerações finais

Podemos, sem dúvida, colocar Ribot como um precursor do que chamaríamos hoje de filosofia da mente. Os temas principais dessa disciplina, a saber, a relação corpo-mente, a natureza do eu e a identidade são aspectos importantes de seu pensamento. O médico e filósofo belga Missa (1993, p. 85-6, 133 e 138-19) considera que as ideias do psicólogo francês ainda teriam uma certa validade, como, por exemplo, a multiplicidade e não a unidade da consciência, pois o seu método baseado nas neurociências o afastou da filosofia espiritualista, o que não teria ocorrido com Henri Bergson. Missa faz sua análise no contexto da filosofia da mente[22], a qual, segundo o autor, faz parte das ciências cognitivas, que, por sua vez, englobam as neurociências (cf. MISSA, 1993, p. 15).

Para o autor, as ciências cognitivas, que teriam um caráter fortemente filosófico, seriam compostas pelas seguintes disciplinas: psicologia cognitiva, linguística, inteligência artificial, neurociências e filosofia da mente. As neurociências, no sentido do conjunto de ciências que investigam o sistema nervoso, têm como constituintes principais a neurologia clínica, a neurofisiologia, a neuroanatomia, a neuroquímica, a neuropsicologia e a neurofarmacologia. Essas ciências, cada vez mais, segundo Missa (1993), atraem filósofos, tais como Patricia e Paul Churchland, Edward Hundert, John Searle e Daniel Dennett.

Assim, Ribot faria parte de uma perspectiva que defende a fundamentação da questão axial mente-cérebro (*esprit-cerveau*) nas ciências experimentais e que se contrapõe ao que seriam as três principais teses metafísicas de *Matéria e memória* (*Matière et mémoire*, 1896), de Bergson: 1. O cérebro é o órgão de ação, não de representação; 2. A memória é de natureza espiritual, e o cérebro não é um depósito de lembranças; e 3. O eu (*moi*) é uma entidade única, indivisível (cf. MISSA, 1993, p. 139).

Missa (1993, p. 39-41) inclui Ribot entre aqueles que propõem a teoria do duplo aspecto: o espírito é a face subjetiva e o cérebro é a face objetiva da mesma entidade, a mente-cérebro[23]. Os outros autores seriam: Gustav Theodor Fechner, em *Elementos da psicofísica* (*Elemente der Psychophysik*, 1860); Hippolyte Taine, em *Da inteligência* (*De l'In-*

22 Missa utiliza a expressão *philosophie de l'esprit* (filosofia do espírito). No entanto, faz uma nota esclarecendo que, apesar de na língua francesa *esprit* ter conotação espiritualista, ou seja, correspondente à alma, ele utiliza essa palavra no sentido do termo inglês *mind* (mente), que seria mais "neutro" (cf. MISSA, 1993, p. 18). Por isso, preferimos traduzir, no contexto das ideias de Missa, *philosophie de l'esprit* por filosofia da mente.

23 Ribot, de fato, defende, em *A hereditariedade*, que a diferença entre o físico e o moral (ou psicológico) não se refere à natureza de cada um, mas ao modo pelo qual os conhecemos (cf. RIBOT, 1873, p. 355-6).

telligence, 1870); Alexander Bain, em *O espírito e o corpo* (*L'Esprit et le corps*, 1873); e Thomas Nagel, em *O que isso tudo significa?* (*What does it All Mean?: A very short Introducion to Philosophy*, 1987). Embora essa teoria tenha o grave defeito de não explicar como se produz a passagem da face objetiva (a atividade cerebral) à face subjetiva (a experiência interior), Missa acredita ser ela aquela postura que melhor permite a investigação do problema *sprit-cerveau*.

À teoria do duplo aspecto, Missa (1993, p. 18-9) associa seu objetivo, qual seja: investigar os aportes consideráveis que as neurociências podem oferecer à filosofia da mente, a qual deve abandonar as considerações *a priori* e abarcar os dados das ciências cognitivas. Sua meta é construir uma filosofia natural, ou seja, uma reflexão filosófica enriquecida pelo método analítico e pelas descobertas das ciências experimentais. Ele arrola alguns problemas a serem abordados por essa disciplina: a relação corpo-mente; a percepção; a memória; as emoções; a consciência; a intencionalidade; a noção de localização cerebral; a relação entre o inato e o adquirido; o papel da evolução darwiniana no funcionamento da mente-cérebro; a noção de unidade do eu; a terminologia da disciplina; e sua metodologia.

Essa proposta parece ser próxima daquela de Patricia Churchland, em *Neurofilosofia* (*Neurophilosophy*, 1986), na qual a filosofia da mente deve se naturalizar por meio das neurociências e da psicologia cognitiva[24]. Para ela, uma nova abordagem mente-cérebro propiciaria o conhecimento de nós mesmos. No final de *Neurofilosofia*, lemos:

> as descobertas na neurociência indubitavelmente substituirão uma série de ortodoxias estabelecidas e queridas da filosofia. Exceto por um milagre ou por uma teimosia calcificada, isso transfigurará particularmente a epistemologia, quando descobrirmos o que realmente significa para o cérebro aprender, teorizar, conhecer e representar. A neurociência pode mesmo nos ensinar uma ou duas coisas fundamentais sobre como a ciência e a matemática são elas próprias possíveis para a nossa espécie. Isso é, então, o cérebro investigando o cérebro, teorizando sobre o que os cérebros fazem quando eles teorizam, descobrindo o que os cérebros fazem quando eles descobrem, e ser mudado para sempre pelo conhecimento (CHURCHLAND, 1986, p. 482).

24 Churchland, segundo Missa (1993, p. 203), quase iguala a atividade filosófica com a atividade científica, sendo que a diferença estaria somente na visão panorâmica da filosofia sobre as coisas e em assumir questões desprezadas pelas ciências. Lembremos que o subtítulo de *Neurofilosofia* é "Em direção a uma ciência da mente/cérebro unificada".

Embora Ribot também propusesse a investigação fisiológica acerca dos fenômenos psicológicos ou mentais, o que valeria uma aproximação às ideias imediatamente acima apresentadas, não podemos esquecer que seu esforço teórico sempre esteve ligado à autonomia da psicologia em relação à filosofia. Como positivista, embora dissidente, rejeitava toda abordagem metafísica: o futuro da filosofia, com a independência científica de seus vários ramos, é tornar-se uma metafísica cada vez mais abstrata, ou seja, especulações gerais do espírito humano sobre as primeiras e últimas causas (cf. RIBOT, 1870, p. 11-4). Para ele, a filosofia se esvaziará, pois estará tão afastada dos fatos, extremamente abstrata, que se tornará arte: "Diz-se que os metafísicos são poetas que lhes falta a vocação" (RIBOT, 1870, p. 15). E complementa: poesia má escrita para uns, divina para outros.

Portanto, o psicólogo francês não pretendia erigir uma filosofia da mente ou uma filosofia natural ou ainda uma filosofia naturalista. Tratava-se de uma psicologia como ciência experimental. A produção do conhecimento científico corre ao lado da rejeição de conceitos abstratos puros da metafísica, particularmente essências imutáveis que seriam responsáveis pelos fenômenos psicológicos. Ribot afirmava o movimento, a multiplicidade, a diversidade e a história evolutiva, e buscava explicar o comportamento humano por meio dessa perspectiva.

No caso da memória, ela é uma função geral do sistema nervoso, estreitamente relacionada à hereditariedade, e tem por base a propriedade dos elementos nervosos mais simples de conservar uma modificação recebida e de formar associações dinâmicas. Conservar e reproduzir, as principais operações da memória, são condições fundamentais da vida. Não há uma dualidade corpo-alma, mas, sim, uma continuidade entre o físico-químico, o fisiológico e o psicológico (que inclui o moral e o cultural). A diferença entre essas instâncias não é qualitativa, mas de graus de complexidade do agrupamento dos fenômenos físico-químicos mais elementares.

Assim, a memória psíquica é apenas uma forma mais complexa de memória. Vida é também movimento, transformação. E sendo a memória um tipo de hereditariedade, é justamente essa característica que permite a mudança e a renovação dos estados mentais. Como as condições naturais são mutáveis, a conservação da estrutura não se mantém perfeita, e isso abre a possibilidade de transformações: a memória não é algo absoluto. A quebra do equilíbrio no inconsciente fisiológico, provocada por novos estímulos, faz sur-

gir novas representações na consciência. Como consequência importante, temos que nós, seres humanos, não devemos estar sempre submetidos às mesmas circunstâncias, pois a conservação reinaria absoluta sobre a renovação de nossos estados mentais e nos tornaríamos máquinas, seres automáticos, sem reflexão, e de funcionamento apenas inconsciente.

Referências

CHURCHLAND, Patricia Smith. **Neurophilosophy:** toward a unified science of the mind/brain. Cambridge: The Massachusets Institute of Tecnology, 1986.

COMTE, Auguste. **Curso de filosofia positiva**. Tradução: J. A. Giannotti. São Paulo: Abril Cultural, 1973 (Os pensadores).

COUSIN, Victor. **Fragments philosophiques**. 2ª ed. Paris: Ladrange, 1833.

FREZZATTI Jr., Wilson Antonio. **Nietzsche e a psicofisiología francesa do século XIX**. São Paulo: Humanitas, 2019.

HAECKEL, Ernst. **Die Lebenswunder:** Gemeinverständliche Studien ***über*** biologische Philosophie. Leipzig/Berlin: Alfred Kröner Verlag/Carl Henschel Verlag, 1924.

LEWES, George Henry. **Problems of life and mind**. Third series. vol. 2. London: Trübner & Co., 1879. Disponível em: https://books.google.com.br/books?id=fokZAAAAYAAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false. Acesso em: 12/05/2023.

MISSA, Jean-Noël. **L´ esprit-cerveau:** la philosophie de l´ esprit à la lumière des neurosciences. Paris: Vrin, 1993.

NICOLAS, Serge. **Histoire de la psychologie française:** naissance ***d´ une nouvelle*** science. Paris: In Press, 2002.

PAGET, James. **Lectures on Surgical Pathology**. vol. I. London: Brown, Green, and Longmans, 1853.

RIBOT, Théodule. **La psychologie anglaise contemporaine (*école* expérimentale)**. Paris: Librairie Philosophique de Ladrange, 1870.

RIBOT, Théodule. **L´ hérédité:** ***Étude*** psychologique sur ses phénomènes, ses lois, ses causes, ses conséquences. Paris: Librairie Philosophique de Ladrange, 1873.

RIBOT, Théodule. **La psychologie allemande contemporaine (*école* expérimentale)**. Paris: Librairie Germer Baillière, 1879.

RIBOT, Théodule. **Les maladies de la mémoire**. Paris: Germer Baillière, 1881.

ROUX, Wilhelm. **Der Kampf der Theile im Organismus:** ein Beitrag zur Vervollständigung der mechanischen Zweckmässigkeitlehre. Leipzig: Verlag von Wilhelm Engelmann, 1881.

SPENCER, Herbert. **The principles of biology**. v. I. London: Williams and Norgate, 1864.

LADY LOVELACE: A CONDESSA DOS ALGORITMOS

Rozenilda Luz Oliveira de Matos

Ourides Santin Filho

Resumo:

Augusta Ada Byron (1815-1852) foi uma das mais importantes figuras no campo da história da computação. Ainda que a história da ciência tenha sido escrita nos moldes da cultura patriarcal, onde as mulheres foram silenciadas ou "ocultadas" da história, algumas delas não foram sujeitadas pelas omissões da história. Quem foi Ada Lovelace e quais desafios ela enfrentou perante a sociedade londrina, por ser filha do poeta mais famoso da época? Emblemática, enérgica e totalmente inclinada a desvendar a lógica dos cartões perfurados, foi quem abriu espaço para a primeira linguagem de programação, cem anos antes que alguma máquina conseguisse, de fato, rodar os primeiros cartões programados. Assim o presente artigo pretende trazer um pouco mais sobre a história de Ada Byron e as controvérsias que giram em torno de sua figura no cenário da grande Londres do século dezenove.

Palavras-chave:

Ada Lovelace; Computação; Mulheres na ciência.

Abstract:

Augusta Ada Byron (1815-1852) was one of the most important figures in the field of computer history. Although the history of science was written along the lines of patriarchal culture, where women were silenced or "hidden" from history, some of them were not subjected by the omissions of history. Who was Ada Lovelace and what challenges did she face in London society as the daughter of the most famous poet of the time? Emblematic, energetic and totally inclined to unravel the logic of punched cards, it was the one who made room for the first programming language, a hundred years before any machine could actually run the first programmed cards. Thus, this article intends to bring a little more about the history of Ada Byron and the controversies that revolve around her figure in the setting of the great London of the nineteenth century.

Keywords:

Ada Lovelace; Computation; Women in science.

Introdução

Quando se envereda pelos estudos da História da Ciência, as "surpresas" literárias, como seria dito sobre Ada, encantam nosso olhar. De modo algum, a mulher foi trazida nos grandes livros de história e ocupou o seu lugar no campo da pesquisa, principalmente, quando nos referimos às Ciências Exatas. Porém, basta começar a procurar que elas aparecem, algumas vezes, com nome de homens, outras, como cortesãs, outras como bruxas, curandeiras e até mesmo como encantadoras de números. Assim foi chamada Ada Lovelace.

Quando lemos sobre História da Ciência, cada vez mais encontramos referências ocultas sobre o papel das mulheres na produção de conhecimento e, aos poucos, percebemos que a interpretação da história e o seu contexto político, cultural e social foram determinantes na maneira com que se registrou a história dessas mulheres, principalmente, na relação mulher e epistemologia. "Seria simplório estudar o feminino na ciência como sendo a chave para a discriminação histórica; será também uma distorção considerar simplesmente a discriminação como sendo um reflexo da história" (MATOS, 2019), em todos os ramos do conhecimento.

O campo da computação foi durante muito tempo um espaço que se pensava ser dominado pelo mundo masculino, do qual, para o senso comum, as mulheres estão distantes. Contudo, numa leitura mais profunda, constatamos que as mulheres foram as pioneiras, sendo a maior delas Augusta Ada Byron (Lady Lovelace, 1815-1852), que, podemos afirmar, sem exagero, "abriu a fila" para uma constelação de mulheres importantes na área, das quais apontamos mais duas figuras emblemáticas: Grace Murray Hopper[1] , lembrada pela sua contribuição para o desenvolvimento da linguagem do programa *Common Business Oriented Language* (COBOL) e Mary Kenneth Keller, que desenvolveu o BASIC (*Beginner's All-purpose Symbolic Instruction Code*), esta última voltando seus interesses para que este código a auxiliasse nos processos educacionais. Grace Murray Hopper, por outro lado, foi a oficial que trabalhou durante a segunda guerra mundial desenvolvendo cálculos de trajetórias de mísseis balísticos.

1 Hopper foi para o Bureau of Ordinance Computation Project na Universidade de Harvard, onde trabalhou na programação da série de computadores Mark. Seu trabalho teve tanto sucesso, por conta da programação dos Mark I, Mark II e Mark III, que ela recebeu o prêmio Naval Ordinance Development Award. Hopper rapidamente ganhou o respeito de Aiken e dos outros membros de sua equipe, que estavam trabalhando no Mark I. O computador de 51 pés (15,5 metros) de comprimento foi ideia Aiken e tinha sido construído pela IBM (GÜRER ,2002; GOYAL, 1996). Nesse trabalho, Hopper utilizou os princípios computacionais primeiramente trabalhados por Charles Babbage e por Ada Lovelace (LUEBBERT & TROPP, 1972).

Apesar da rapidez da disseminação da informação que temos acesso, no Brasil, ainda hoje, poucas são as obras que relatam a história da computação, principalmente aquelas que abordem o papel da mulher na computação. Localizamos, dentre elas, a de Fonseca Filho (2007), Wazlawick (2016), Banks Costa (2008), Isaacson (2014), Plant (1995) e Abbate (2003). A obra de Wazlawick é interessante por ser densa, com detalhes, e ricamente ilustrada, apontando, já na contracapa, observações sobre o fato de as mulheres serem minoria na área da computação. Nesse contexto, os estudos de Eli Banks Costa são de fundamental importância, uma vez que trazem dados sobre o invento de Jacquard e os aspectos das origens da computação no século XIX, destacando a importância dos cartões perfurados na história da programação.

A evolução dos conceitos em informática está estritamente ligada à história da matemática e muitas mulheres aparecem neste campo de estudos, estando, no entanto, também ausentes nos livros, sendo, muitas vezes, substituídas pelos seus maridos ou pelos seus chefes de pesquisa. Conforme Pinheiro (2008 p. 22), a Ciência da Computação é um campo interdisciplinar desenvolvido num núcleo teórico das ciências exatas ou ciências duras e, muitas vezes, em muitas Universidades, a Computação aparece vinculada ou dentro de Departamentos de Matemática, tornando essa relação bastante estreita.

Ao compararmos a ciência da computação com as demais ciências, aquela é bem recente, embora por volta do ano 1120 já houvesse traduções sobre os algarismos e até 1550 eles já estivessem disseminados na Europa, principalmente na Itália. Na história da matemática, os desenvolvimentos da álgebra e da aritmética exerceram papel fundamental no desenvolvimento da computação, pois a ideia era tentar reduzir todo raciocínio a um processo mecânico, um esforço despendido, por exemplo, por Raimundo Lúlio (1235-1316).

É, na passagem do século XVIII para o XIX, que se observa um grande avanço na fundamentação da ciência da computação. O período que vai da lógica formal à lógica simbólica abre novas oportunidades para elaboração de cálculos cada vez mais abstratos, rumo à própria automatização do pensamento, principalmente, depois dos estudos de George Boole (1815-1864), que enfatizou a possibilidade de aplicar o cálculo formal em diferentes situações relacionadas a operações matemáticas com regras formais.

Conforme Matos (2019), no campo industrial, esse período é marcado por grandes transformações que afetaram a produção e o comércio, principalmente, na manufatura têxtil. Com o desenvolvimento das máquinas, a força muscular para operá-las tornou-se prescindível, era preciso apenas alguém que tivesse pouca força ou maior flexibilidade

para o trabalho, utilizando-se também do trabalho feminino e infantil de forma a submeter ao comando imediato do capital todos os membros da família dos trabalhadores, sem distinção de sexo nem idade (MARX, 2013, p. 468).

Na Europa do século XVIII, França, Alemanha e Inglaterra contavam com mão de obra infantil e feminina em suas fábricas, e Joseph Marie Jacquard (1752-1834) foi uma dessas crianças. Conforme o modelo educacional da época, quando as crianças aprendiam os ofícios por imitação do trabalho dos adultos, ele aprendeu o ofício com seu pai e se tornou encarregado da substituição de novelos de diferentes cores em teares, a fim de que a máquina reproduzisse determinados padrões em tecidos (para produzir apenas 1 centímetro de tecido, levava-se aproximadamente 30 minutos). Ainda naquela época, eram necessárias três pessoas para operar um tear: o leitor de desenhos, o puxador de laços e o tecelão (COSTA, 2008, p. 16).

Tear de Jacquard. Foto do arquivo pessoal da pesquisadora. The Science Museum of London

A fim de se livrar da tarefa altamente cansativa e repetitiva, Jacquard criou um sistema de cartões perfurados, tal que a máquina lia os cartões e executava as operações na sequência desejada (WALZLAWICK 2016). Segundo Costa (2008), o invento de Jacquard está intimamente ligado ao surgimento da computação, uma vez que o tear se tornou "programável" e foi, com o apoio de amigos, que Jacquard desenvolveu sua máquina de tear automatizada com cartões perfurados. Em 1811, existiam cerca de onze mil desses teares na França, número que viria a crescer exponencialmente em toda a Europa. Por sua invenção, cuja finalidade era liberar os operários de um trabalho penoso e repetitivo, Jacquard recebeu, em 1811, uma medalha de ouro e a Cruz da Legião de Honra.

Foi numa Londres repleta de trabalhadores assalariados (homens, mulheres e crianças), mão de obra explorada à exaustão e na efervescência do desenvolvimento das indústrias, dos pensamentos e dos poetas audaciosos e em confronto com a miséria do mundo, que nasceu Augusta Ada Byron - Lady Lovelace (1815-1852).

Ada era filha do poeta inglês Lord Byron e de Ann Isabella Milbanke, uma matemática que tinha o título de "Princesa dos Paralelogramos". Sua mãe a incentivou a estudar matemática e, para tanto, contratou tutores. Amiga de Ann, Mary Somerville, tutora e tradutora que trabalhava com matemática e astronomia em Cambridge, foi a responsável pela tradução do trabalho de Laplace (Mécanique Céleste) para o inglês, por volta de 1833.

George Gordon Byron (1788-1824)

George Gordon Byron (1788-1824), famoso como Lord Byron, foi um dos principais poetas britânicos do romantismo e nasceu em Londres no dia 22 de janeiro de 1788. Era filho de John Byron e Catherine Gordon de Gight. George possuía o título de sexto barão dos Byron, que lhe foi outorgado após a morte de seu avô em 1798. Estudou em Cambridge, onde fez o mestrado e, aos dezenove anos, casou-se em 1815 com Anne Milbanke, com quem ficou apenas um ano, relacionando-se depois com Claire, com quem teve uma filha, Allegra, que morreu de febre.

Byron influenciou várias gerações de poetas e escritores. No Brasil, sua influência foi principalmente sobre o poeta Álvares de Azevedo, da segunda fase do romantismo, conhecida como "geração Byroniana".

Lord Byron era muito famoso nos meios literários de Londres e sempre convidado para jantares e passeios, frequentava muitas festas e era também famoso por seus casos amorosos. Dentre estes, foi amante de Lady Caroline Lamb, casada com um poderoso aristocrata político, e, em umas das festas por ela organizada, Byron notou uma moça "vestida de modo mais simples". Seu nome era Annabella Milbanke, de 19 anos, a qual concluiu que ela daria uma "esposa adequada", ou seja, Annabella parecia o tipo de mulher que podia domar esses sentimentos e protegê-lo de seus excessos e, obviamente, ela também poderia ajudar a pagar suas muitas dívidas. De forma simples e sem muito entusiasmo, ele a pediu em casamento por carta e ela, atenta aos comentários sobre o conquistador Byron, resolveu recusar prontamente o seu pedido.

Com a recusa da moça, ele se afastou e passou a ter companhias consideradas na época como não apropriadas, dentre as quais sua meia-irmã, Augusta Leigh. Mesmo tendo fracassado no início, as insistentes investidas de Byron levaram Annabella a contrair matrimônio com ele em janeiro de 1815.

Annabella gostava de cálculos e tinha aulas de matemática, o que fez com que Byron a apelidasse de "Princesa dos Paralelogramos". Com o passar do tempo, ele usava sempre o seu gosto pela matemática para fazer piadas e debochar de Annabella: "Somos duas retas paralelas prolongadas ao infinito lado a lado que nunca se encontrarão", "Sua ciência favorita era a matemática […]. Ela era um cálculo andante" (ISAACSON, 2014).

Byron teve uma vida atribulada e, ao lutar na guerra pela independência grega do Império Otomano, acabou adoecendo, vindo a falecer em 19 de abril de 1824, tendo seu corpo transladado para a Inglaterra, mas o seu coração enterrado em terras gregas.

Augusta Ada Byron (1815-1852)

O nascimento da filha do casal ocorreu em 10 de dezembro de 1815. A recém nascida foi batizada de Augusta Ada Byron, sendo o primeiro nome uma homenagem à amada meia-irmã de Byron. Anabbelle, contudo, chamaria a filha pelo nome do meio. Depois do nascimento, Ana foi embora com Ada, que jamais voltou a ver o pai. Lord Byron deixou a Inglaterra em abril, tendo dado à mãe a custódia da filha, que nunca mais voltaria a ver, embora sempre buscasse notícias dela. É à filha que Byron dedica a abertura do canto 3 de *Childe Harold's Pilgrimage18*: *Teu rosto lembra tua mãe, bela criança! Ada! Tu, o fruto único de meus ramos? Vi em teus olhos riso e esperança, E nos separamos.* Ada não chegou a ver nem sequer um retrato de seu pai durante anos, mas Byron sempre carregava um retrato da filha junto de si e sempre enviava cartas, querendo saber dos gostos e aptidões da filha.

A mãe de Ada sempre se empenhou em dar uma educação primorosa para filha, ainda que nesse período as oportunidades educacionais não fossem as mesmas para meninos e meninas. Normalmente, o destino das meninas já estava traçado, ou seja, o mundo que correspondia ao casamento e filhos. No entanto, nem todas as mães queriam esse destino para as filhas, dentre elas Annabella, que contratou preceptores que ensinassem a Ada tudo que ela precisasse, principalmente matemática, com a intenção de que se afastasse das inclinações de seu pai, como a poesia.

Ada herdou do pai o temperamento poético e insubordinado, fato que assustava sua mãe, embora seu amor pelas máquinas fosse maior do que pela poesia. Lord Byron,

ao contrário da filha, era um ludita[2]. No primeiro discurso que fez na Câmara dos Lordes, em fevereiro de 1812, aos 24 anos, Byron defendeu os seguidores de Ned Ludd, ferrenho combatente dos teares mecânicos. Fiel a seu estilo e com uso de desprezo sarcástico, Byron ironizou os donos de moinhos de Nottingham, que defendiam um projeto de lei que tornaria a destruição de teares automatizados um crime punível com a pena de morte. "Essas máquinas para eles foram uma vantagem, na medida em que tornaram obsoleta a necessidade de empregar muitos operários, que em consequência foram deixados passando fome" (ISAACSON, 2014, p. 23), declarou Byron. "Os operários rejeitados, na cegueira de sua ignorância, em vez de se rejubilar com essas melhorias em artes tão benéficas à humanidade, julgaram-se sacrificados em nome de melhorias mecânicas" (ISAACSON, 2014, p. 23).

Ada viria a se casar com William King-Noel, barão que acabou se tornando o Conde de Lovelace. A partir deste momento, Ada perde o sobrenome de seu pai e recebe o nome de Augusta Ada King e o tratamento de Condessa de Lovelace.

No ano de 1833, Ada foi apresentada a Charles Babbage (1791-1871), cientista, matemático, filósofo, engenheiro mecânico e inventor inglês. Ada conheceu o dispositivo inventado por Baggage, a "Máquina Diferencial". Impressionada com o invento, Ada desenvolveu o que se pode chamar de a primeira linguagem de programação, distante por impressionantes cem anos antes do primeiro computador construído. A filha do grande poeta tornara-se, então, personagem fundamental para o desenvolvimento da computação, um fato tão impressionante quanto desconhecido nos tempos modernos e que não pode ser nem de longe desprezado. Como se fosse pouco, Ada Lovelace, admiradora do invento de Jacquard, desenvolveu e comparou os padrões algébricos a serem calculados pela Máquina de Babbage com os desenhos executados pelo tear de Jacquard (COSTA, 2008, p. 64).

Conforme Matos (2019), Babbage foi convidado a discursar no Congresso de Cientistas Italianos em Turim sobre sua Máquina Analítica, e quem fazia as anotações era um jovem engenheiro militar, capitão Luigi Menabrea, que mais tarde seria primeiro-ministro da Itália. Com a ajuda de Babbage, Menabrea publicou uma descrição detalhada da máquina em francês, em outubro de 1842.

2 Membro do movimento inglês do final do século XIX, que se opunha à mecanização e à industrialização. Os ludistas protestavam contra a substituição da mão-de-obra humana por máquinas. Ludita, In: Dicionário Priberam da Língua Portuguesa [em linha], 2008-2023, https://dicionario.priberam.org/ludita

Depois disso, um dos amigos de Ada sugeriu que ela traduzisse o texto de Menabrea para o *Scientific Memoirs*, um periódico dedicado a artigos científicos. Ao término, Babbage ficou em certa medida surpreso e perguntou porque não tinha ela mesma redigido um artigo original, ao que Ada respondeu simplesmente que isso não lhe havia ocorrido. Não era natural na época que mulheres publicassem artigos científicos (ISAACSON 2014 p. 39). Babbage lhe sugeriu então que acrescentasse algumas anotações ao trabalho de Menabrea, e Ada começou a trabalhar em uma seção que chamou de "Notas da tradutora", a qual acabou tendo mais do que o dobro do tamanho do artigo original, assinadas com um "A. A. L." (Augusta Ada Lovelace) (WAZLAWICK, 2016, p. 63).

Durante o verão de 1843, enquanto trabalhava em suas anotações, Ada e Babbage trocaram inúmeras cartas, de modo que, no outono, eles já haviam se encontrado várias vezes, depois de ela ter voltado à sua casa na praça St. James, em Londres. As notas de Ada contribuíram muito e descreveram a essência dos computadores modernos. Ela finalizou suas notas com a frase: "Podemos dizer com maior aptidão que a Máquina Analítica não tece padrões algébricos da mesma forma que o tear de Jacquard tece flores e folhas" (COSTA, 2008, p. 65).

Babbage manteve contato constante com Ada Lovelace e pedia para ela não modificar suas notas, mas foi na "Nota G", sua terceira contribuição, que Ada descreveu os detalhes do funcionamento do que hoje chamamos de programa de computador ou algoritmo. Na nota ela demostra que a Máquina Analítica poderia gerar os chamados "números de Bernoulli" e, assim, as operações poderiam ser feitas em sequências. Ao longo de seus apontamentos, Ada ajudou a inventar os conceitos de "sub-rotinas" e do "*loop* recursivo".

Ao escrever sobre a "sub-rotina", Ada não imaginaria que, depois dela, muitas mulheres viriam a utilizá-la também, dentre elas Grace Hopper, em Harvard; Kay McNulty e Jean Jennings, na Universidade da Pensilvânia. Embora fosse muito criativa, Ada não acreditava que as máquinas poderiam vir a "pensar" ou ter intenções próprias; a máquina não possui capacidade de aprendizado de forma independente. A Máquina Analítica não tem nenhuma pretensão de originar algo, escreveu em suas "Notas". Ela pode fazer tudo aquilo que soubermos ordenar-lhe que faça. Ela pode seguir análises; porém, não tem poder de antecipar quaisquer relações analíticas ou verdades (ISAACSON, 2014, p. 43). Um século depois, a afirmação de Ada seria refutada por Alan Turing em seu escrito publicado em 1950, *Computing Machinery and Intelligence*, sobre a inteligência artificial. Em sua objeção mais famosa contra o pensamento de Lady Lovelace, Turing argumenta que os computadores podem nos surpreender e que Ada não se havia dado conta disso, pois

fora impedida pelo contexto social em que viveu e escreveu.

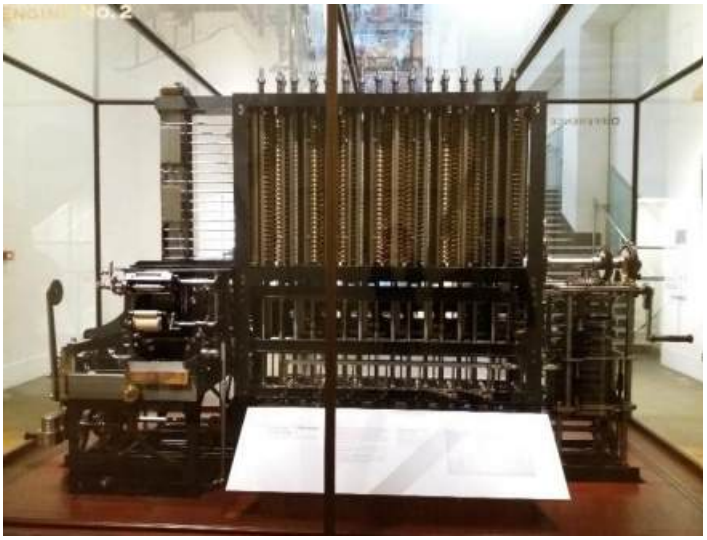
Existem alguns debates sobre o quanto do pensamento exposto nas notas era de Ada e o quanto era de Babbage e, sobre isso, Isaacson (2014 p. 43) escreve: "Entre os autores que escreveram sobre Ada Lovelace há quem a canonize e quem a desmascare. Há misóginos de plantão para atacar e criar falsos argumentos sobre as mulheres na história da ciência".

Sobre essa questão, os livros mais abrangentes são os de Toole, Wooley e Baum. Para um "desmascaramento" de Ada Lovelace, ver Bruce Collier, "*The Little Engines That Could've*", tese de doutorado, Harvard. Collier escreveu: "Ela foi uma maníaco-depressiva com assombrosos delírios sobre seus talentos. […] Ada era doida varrida, e além de problemas pouco mais contribuiu para as 'Notas'". Tais argumentos agressivos não são incomuns, ainda que, em suas memórias, Babbage dê a Ada boa parte do crédito de seu trabalho. Ele escreve:

> Discutimos várias ilustrações que podiam ser acrescentadas: sugeri muitas, mas a escolha ficou inteiramente por conta dela. O mesmo vale para o trabalho algébrico referente a vários problemas, exceto, na verdade, aquele que diz respeito aos números de Bernoulli, que eu havia me oferecido para fazer, a fim de que Lady Lovelace não precisasse ter esse incômodo. Ela me enviou isso de volta para que eu fizesse emendas, já que havia detectado um grave erro que eu havia cometido no processo (GREEN, 1864, p. 136).

Babbage e Lovelace iriam ter uma parceria promissora, ela tentaria angariar fundos para a construção de sua máquina para que nela continuassem trabalhando e a verdade é que a contribuição de Ada foi tão profunda quanto inspiradora. Infelizmente, a parceria não prosperou, pois Babbage não conseguiu recursos para a construção de sua máquina e Ada não publicou o seu artigo.

Em suas conversas com Babbage, Ada percebeu que poderia existir uma combinação entre funções lógicas e aritméticas, diferentemente das máquinas anteriores com funcionamento analógico (execução de cálculos usando medidas).



A máquina diferencial de Babbage n. 02. Desenhada 1847-49 e construída em 1985-2002. Science Museum of London. Foto do arquivo pessoal da pesquisadora.

A Máquina Diferencial, por outro lado, era "digital" (execução de cálculos usando fórmulas numéricas). Por seu trabalho, Ada Lovelace é considerada a patrona da arte e ciência da programação. Mesmo não estando a máquina de Babbage construída, as sub-rotinas de *loops* e saltos são largamente utilizadas na programação de computadores de hoje. Em suas anotações sobre o projeto de Charles Babbage, Ada incluiu suas próprias observações para o cálculo da sequência de Bernoulli. Essas séries de instruções constituem, de fato, o primeiro programa escrito e documentado na história.

Pintura de Ada Lovelace. Cortesia do Instituto Charles Babbage, Universidade de Minnesota.

Podemos entender que Ada extrapolou ao pensar na utilidade da máquina. Para ela, além de cálculos numéricos, a máquina poderia reproduzir e trabalhar com outras coisas, como cartas, notas musicais, entre outras coisas (COSTA, 2008, p. 65). Ada apresenta um algoritmo completo para computar os números de Bernoulli e esse algoritmo especificamente é considerado o primeiro programa de computador já escrito no mundo (WAZLAWICK, 2016, p. 65).

Durante grande parte de sua vida, a filha do poeta esteve doente e, em seu último ano de vida, lutou uma batalha cada vez mais dolorosa contra um câncer de útero, acompanhado de constante hemorragia. Sua saúde frágil fazia com que tivesse diversas crises, desmaios, ataques de asma e paralisias (ESSINGER, 2017). Conforme Schuartz (2006, p. 16), Ada, em muitas ocasiões, chegou a acreditar que a causa de sua histeria seria o uso de seu intelecto e chegou a escrever: "numerosas causas contribuíram para produzir os desequilíbrios passados e, no futuro, vou evitá-las. Um dos ingredientes (mas apenas um entre muitos) foi o excesso de matemática". Ada Lovelace foi acusada de dormir com seus tutores e muitos livros que tentaram destruir sua imagem de pesquisadora competente e produtiva salientam a sua vida sexual e nada sobre seu intelecto matemático; outros se referem ao fato de seu câncer ser um castigo por sua vida promíscua (PLANT, 1999, p. 35).

Com apenas 36 anos, no ano de 1852, Ada Lovelace morreu de câncer e foi sepultada, de acordo com um de seus últimos desejos, em um túmulo no campo ao lado do pai poeta que ela nunca conheceu e que havia morrido com a mesma idade.

Algumas considerações

Durante as leituras de História da Ciência, podemos notar como são trazidas as atuações femininas no processo histórico do desenvolvimento da tecnologia. Fica uma pergunta importante e provocativa: por qual razão ainda há uma chamada em relação ao espaço da computação e incentivos em forma de projetos de pesquisas, mas não há um desvelamento da participação das mulheres no processo da ciência da computação? Ada Lovelace costuma ser lembrada como a "encantadora" de números ou a "poetisa" dos números, em detrimento do seu saber matemático. Wazlawick (2016) se refere a ela como "matemática amadora", mas não se houve falar e nem se lê sobre homens da matemática, como Descartes, Leibniz, Pascal, etc., como sendo "amadores", "encantadores" ou "poetas" de números.

Como se, para eles, a compreensão dos números fosse racional e, para ela, fosse "mágica". Parece pouco tal argumento para se ter esta reflexão nas linhas deste texto, fruto de uma tese de doutorado, mas são esses "poucos" que ajudam a formar uma imagem que desqualifica a mulher cientista. Os números não "dançaram" para Ada Lovelace, como se pode ler na literatura corrente, mas ela os decodificou e calculou rigorosamente.

No Brasil, na maioria dos livros de história da computação e história da informática, não encontramos mais que três parágrafos curtos sobre Ada Lovelace, porém, em

outros países, como a Inglaterra, as referências são maiores, inclusive histórias em quadrinhos são dedicadas à primeira programadora da história da computação (PÁDUA, 2016). Algumas linhas, inclusive, se referem equivocadamente ao seu "gênio" e "comportamento" diferenciado, de forma caricata, em seu trabalho como pesquisadora, o que é bastante comum. Para homenageá-la, no ano de 1979, o Departamento de Defesa Americano deu o nome de Ada à primeira linguagem de programação. Tantos avanços foram possíveis posteriormente, como o ENIAC, programado por seis mulheres, sendo que as operações mais complexas do computador pautavam-se em um processo semelhante ao que foi desenvolvido por Ada Lovelace. Ela também escreveu sobre um computador que pudesse compor e tocar música. Trata-se do CSIRAC. Aos poucos o mundo vai se desprendendo do estigma que carrega a mulher e elas passam a serem vistas como protagonistas de suas próprias histórias.

Ada Lovelace foi uma mulher brilhante, competente e pesquisadora muito produtiva. A inteligência artificial está aí, quase onipresente e onisciente, e deve sua existência muito à Ada e outras mulheres. Se essas modernas ferramentas representam uma ameaça à sobrevivência humana, é um tema que ainda está em discussão. *Chatbots* e outros dispositivos não têm a experiência de uma vivência do passado e nem subconsciente; e talvez ainda não tenham autoconsciência. Contudo, do nosso lado, também não sabemos ao certo o que o futuro nos reserva. Que a sabedoria de Ada nos oriente na busca de um futuro sempre melhor, pois enquanto ela pensava em uma tecnologia que calculasse e tocasse músicas, a mesma tecnologia algum tempo depois guiou mísseis em guerras. Para além do conhecimento, temos que pensar sobre os usos sociais da ciência.

Referências

ABBATE, J. **Women and gender in the history of computing**. IEEE Computing Society, 2003.

BADINTER, E. **Les passions intellectuelles**. Paris, Fayard, 1999.

BARBOSA, R. **Mulheres e cibercultura**: notas sobre os dilemas das mulheres com as TIC na formação superior a distância. [s.l.: s.n.]. Disponível em: http://www.ufpb.br/evento/?searchPage=5. Acesso em: 09 de julho. 2017.

BARNETT, R. C. **A short history of women in science**: from stone walls to invisible walls. Chapter prepared for the American Enterprise Intitute. New York, EUA, 2014.

BARNETT, R. C. e SABATTINI, L. **A short history of women in science**: from stone walls to invisible walls. The American Enterprise Institute, New York, EUA, 2009.

BEYER, Grace Hopper. Disponível em: https://www.famousscientists.org/gracemurrayhopper/.2015. Acesso em 05/02/2019.

COHOON, J. M. & ASPRAY, W. **Women and Information Technology**: research on underrepresentation. The MIT Press, 2006.

COLLIER. "The Little Engines That Could've", tese de doutorado, Harvard, 1970. Disponível em: www.robroy.dyndns.info/collier/.

COSTA, E. B. L. **O invento de Jacquard e os computadores**: alguns aspectos das origens da programação no século XIX. Dissertação de mestrado – PUC – Orientadora: Dr. Maria Helena Roxo Beltran, São Paulo, 2008.

ESSINGER, J. **Ada´s Algorithm**. How Lord Byron´s daughter launcheg the digital age through the poetry of numbers. Printed by Gibson Square, 2017.

FILHO, C. F. **História da computação**: O caminho do pensamento e da tecnologia. Porto Alegre: EDIPUCRS, 2007.

FUEGI, J. FRANCIS, J. Babbage, passages from the life of a philosopher**.** "Lovelace & Babbage and the Creation of the 1843 'Notes'". **Annals of the History of Computing**, out. 2003.

GOYAL, A. Women in computing: historical roles, the perpetual glass ceiling, and current opportunities. IEEE **Annals of the history of computing**, vol 18. N. 3, 1996

GRACE HOPPER. **Biography**. https://www.biography.com/people/grace-hopper21406809. Acessado em 1 de abril de 2019. Publisher A&E Television Networks, January 23, 2019.

GREEN, L. **Charles Babbage**, passages from the life of a philosopher. Londres: 1864. p. 136.

ISAACSON, W**. Os inovadores**. Uma biografia da revolução digital. Trad. Berilo Vargas. São Paulo: Companhia das Letras. 2014.

KORDAKI, M. BERDOUISIS, I. Course Selection in Computer Science: Gender Differences. In Conference on Educational Sciences, 05-8 February 2013, Sapienza University of Rome, Italy, **Procedia - Social and Behavioral Sciences**, Volume, 2013.

KORDAKI, M. & BERDOUSIS, I. Achievements in computer Science courses: gender issues. **Proceedings of INTED2014**. Conference 10th-12th Valencia, Spain March 2014.

BIOGRAPHY.COM EDITORS. Grace Hopper Biography. **The Biography.com website**. Disponível em: https://www.biography.com/scientists/grace-hopper. Acesso em: 20 out. 2023.

SAITO, F. & TRINDADE, L. **Mulheres na ciência, matemática e na computação**. História da ciência. 1ª ed. São Paulo: Editora Livraria da Física, 2017.

MENABREA, L. F. Sketch of the Analytical Engine invented by Charles Babbage. **Bibliothèque Universelle de Genève**, n. 82, 1842. Disponível em http://psychclassics.yorku.ca/Lovelace/lovelace.htm#A. Acesso em: 13 set. 2023.

MINSKY, N. W. **El papel de la mujer em la ciência**. Vol. III, n. 1. Universidade de León Monterrey, México, 2005.

MOREIRA, H.; GRAVONSKY, I.; CALVALHO, M. & KOVALESKI, N. **Mulheres Pioneiras nas Ciências:** Histórias de conquistas numa cultura de exclusão. VIII Congresso Iberoamericano de Ciência, tecnologia e gênero. Curitiba: UTFPR, 2010.

MOREIRA, J. A.; QUEIROZ, C. & CARVALHO, M. E. P. Gênero e Inclusão de Jovens Mulheres nas Ciências Exatas, nas Engenharias e na Computação. In: Maria do Rosário de Fátima Andrade Leitão (Org.). **Gênero, Educação e Comunicação**. 1ed.Recife: EDUFRPE, 2016, v. 1, p. 43-64.

MUZI, J. L. C. & LUZ, N. S. Mulheres no campo da ciência e da tecnologia: avanços e desafios. In: **IV Simpósio Nacional de Tecnologia e Sociedade**, Curitiba, 2011.

PÁDUA, S. **The thrilling adventures of Lovelace and Babbage**. United States of American. Pantheon Books, 2016.

PERROT, M. Mulheres. In: PERROT, M. **Os excluídos da história:** operários, mulheres e prisioneiros. Seleção de textos e introdução Maria Stella Martins Bresciani; tradução de Denise Bottmann. Rio de Janeiro: Paz e Terra, 1988.

PERROT, M. **Minha história das mulheres**. Trad. Angela Correa. 2 ed, São Paulo: Contexto, 2017.

PINHEIRO, C. B. F. **A construção do conhecimento científico**: a Web semântica como objeto de estudo. Dissertação (Mestrado em Ciência da Informação) Faculdade de Filosofia e Ciências, Marília: Unesp, 2008.

PLANT, S. **The future looms**: weaving women and cybernetics. Body Science, vol 1, London, 1995.

PLANT, S. **Mulher digital**: o feminino e as novas tecnologias. Rio de Janeiro: Rosa dos Tempos, 1999.

PLATERO, R. **Globalización y tecnologías de información y comunicaciones**: las mujeres en el cyberactivismo, 2003. Disponível em: https://www.mujeresenred.net/IMG/pdf/ciberactivismo-r_platero.pdf. Acesso em: 22 set. 2023.

RATHGEBER. **Female and male cgiar scientists in comparative perspective**. Washington, DC, Consultative Group on International Agricultural Research, CGIAR|Center for Gender in Organizations, 2002

ROSSI, A. S. Women in Science: why so few? **Science**, v. 148, n. 3674, p. 1196-1202, 1965.

ROSSITER, M. W. **Women scientists in america:** before affirmative action. Baltimore: Johns Hopkins University Press, 1995.

SAMMET, J. E. **Programming languages:** history and fundamentals. Englewood Cliffs, New Jersey: Prentice Hall, 1969.

SANTOS, V. M. **Mulheres e homens na política da ciência e tecnologia**. Fortaleza: EdUECE; EDMETA, 2012.

SCHAFER, V. Femmes, genre et informatique: une question historique. **Bulletin de la société informatique de France** – numéro HS2, février, 2017

SCHAFER, V. Um constat global: mise em perspective historique et sociologique: Femmes, genre er informatique: une question historique. **Buletin de la Société informatique de France**. N. HS2, 2017.

SCHIEBINGER, L. Mais mulheres na ciência: questões de conhecimento. **História, Ciências, Saúde-Manguinhos**, v. 15, suplemento, p. 269-81, jun., 2008.

SCHIEBINGER, L. **O feminismo mudou a ciência?** Trad. de Raul Fiker. Bauru, SP: EDUSC, 2001.

SCHWARTZ, J. *et al.* Mulheres na informática: quais foram as pioneiras? **Cadernos Pagu**. 2006. Disponível em: http://scielo.br/scielo.php?script=sci. Acesso em: 19 de ago. 2017.

SHETTERLY, M. L. **Hidden Figures**: the American Dream and the untold story of black women mathematicians who helped win the space race. Collins Publishers, New York, NY, 2016.

TURING, A. M. Computing machinery and intelligence. England: **Mind**, volume 59, no 236, 1950.

WAZLAWICK, R. S. **História da computação**. 1 ed. Rio de Janeiro: Elsevier, 2016.

A TELEOSEMÂNTICA DE MILLIKAN: UMA ANÁLISE SISTEMÁTICA

Sérgio Farias de Souza Filho[1]

Resumo:

O que é para um estado mental representar o mundo? O que é a intencionalidade, a capacidade de estados mentais de representar a realidade? Ruth Millikan desenvolveu uma teoria teleológica da representação – a teleosemântica – a fim de solucionar este problema. O objetivo deste artigo é fazer uma análise e interpretação da teleosemântica de Millikan. Inicialmente, apresentaremos sua teoria das funções próprias e analisaremos detalhadamente sua teleosemântica. Posteriormente, veremos como Millikan responde a um problema que parece ameaçar a viabilidade de sua teleosemântica, o problema do conteúdo distante. Por fim, faremos uma breve análise da viabilidade desta teleosemântica, concluindo que, ainda que seja em última instância inviável, constitui inegavelmente um marco na filosofia da mente contemporânea.

Palavras-chave:

Filosofia da mente; intencionalidade; Ruth Millikan; teleosemântica.

A propaganda corrente nos conta que os dois principais problemas da filosofia da mente contemporânea são o *problema da consciência* e o *problema da intencionalidade* (FIELD, 1994, p. 34; CRANE, 2006, p. 5). A solução para o primeiro consiste em explicar como a mente pode ser consciente, enquanto a solução para o segundo consiste em explicar como a mente pode ter intencionalidade, isto é, capacidade representacional. Neste artigo focaremos no problema da intencionalidade, analisando uma das teorias da representação mais complexas e influentes: a *teleosemântica de Ruth Garrett Millikan*. Teorias teleológicas da representação – ou teleosemântica – determinam o conteúdo represen-

1 Professor de Filosofia da Universidade Federal Rural de Pernambuco, Recife-PE. Doutor em Filosofia pelo King's College London. Foi Pesquisador de Pós-Doutorado na Universidade Federal do Rio de Janeiro e na Universidade de São Paulo. E-mail: sergiofariasfilho@gmail.com

tacional do estado mental a partir de sua *função biológica*. Ou seja, estados representacionais são *estados biológicos* e representam aquilo que representam em virtude de suas funções biológicas. O mercado oferece diversas teleosemânticas, mas a de Millikan é certamente a principal versão e, reconhecidamente, a mais complexa e desenvolvida.[2]

A teleosemântica de Millikan é extensa e complexa, tendo sido refinada ao longo de décadas desde a publicação de sua obra-prima *Language, Thought, and other Biological Categories*, em 1984. O objetivo deste artigo é fazer uma interpretação e análise sistemática desta teleosemântica. No que se segue, focaremos nos aspectos mais relevantes para a compreensão de sua resposta ao problema da intencionalidade. Iniciaremos com a *teoria das funções próprias*, o ponto de partida de Millikan. Na segunda seção, apresentaremos sua *teleosemântica*. Na terceira, analisaremos como Millikan responde a um caso de indeterminação funcional que ameaça a viabilidade de sua teleosemântica, a saber, o *problema do conteúdo distante*. Por fim, na última seção faremos uma breve análise dos principais problemas que afligem a teleosemântica millikaniana enquanto uma teoria da representação, concluindo que embora dificilmente seja bem-sucedida, ela constitui um marco no debate contemporâneo quanto à intencionalidade.

A Teoria das Funções Próprias

A função biológica do coração é bombear sangue. A função biológica da glândula pineal é produzir melatonina (o hormônio do sono). Mas o que é uma *função biológica* em primeiro lugar? Como é possível para qualquer traço ou mecanismo biológico ter uma função? O que distingue a função de um traço biológico de outros efeitos deste traço? Por exemplo, por que a função do coração é bombear sangue e não emitir aquele som característico, dado que estes dois efeitos sempre ocorrem juntos? Millikan desenvolve uma *teoria das funções próprias* (*proper functions*) justamente para responder a este problema. Trata-se de uma versão das *teorias etiológicas* da função biológica, de acordo com as quais a função de um dado traço biológico é o *efeito* para o qual traços deste tipo foram *historicamente selecionados*. Ou seja, é o *efeito selecionado* que determina a função biológica.

Assim, a função do coração é bombear sangue – não emitir som – porque corações ancestrais foram selecionados para bombear sangue – não para emitir som. Afinal, é justamente o bombeamento de sangue que contribuiu para a aptidão biológica ao longo da

2 MCDONOUGH & PAPINEAU (2006). Para um estado da arte da teleosemântica, incluindo a apresentação e análise de suas principais versões, cf. NEANDER & SCHULTE (2022).

história evolutiva dos corações (i.e., o que aumentou a sobrevivência e sucesso reprodutivo de corações ancestrais), não a emissão de som. A teoria das funções próprias explica como é possível para traços terem função biológica, mas também abarca outros tipos de funções, tais como funções de artefatos, costumes e comportamentos. Millikan recorre à história do mecanismo para determinar sua função (1989a, p. 288):

Função Própria. Para um item A ter como função própria F, é necessário que **(1)** A originou-se como a reprodução de algum(ns) item(ns) prévio(s) que, em parte devido à posse das propriedades reproduzidas, efetivamente executou F no passado e A existe porque houve esta(s) execução(ões); ou **(2)** A originou-se como o produto de algum dispositivo prévio que, dada certas circunstâncias, executou F como sua função própria e normalmente causa a execução de F por meio da produção de um item como A.

As funções do primeiro tipo são *funções próprias diretas*, enquanto as do segundo tipo são *funções próprias derivadas*, já que estas últimas são derivadas das funções dos dispositivos que as produziram. Vejamos primeiro em que consistem as funções próprias diretas, já que este é o tipo mais fundamental de função.

A função própria de um item é definida pela história da seleção de itens deste tipo, em contraposição a uma definição da função em termos das propriedades presentes do item. Agora considere um item A que é uma reprodução de um item B. São diversos os exemplos na natureza e no cotidiano: fotocópias, pegadas, genes, artefatos, comportamentos resultantes de imitação etc. As propriedades por referência às quais *A* é uma reprodução de *B* são *propriedades reprodutivamente estabelecidas* de *A*, sendo *B* o *modelo* de *A*. Pegadas têm propriedades espaciais como reprodutivamente estabelecidas, enquanto genes têm propriedades genéticas como reprodutivamente estabelecidas[3]. Um item com função própria tem esta função enquanto um membro de um tipo de família reprodutivamente estabelecida (FRE). Itens similares formam uma FRE. Vejamos agora o que são explicações Normais.

Uma *explicação Normal* consiste em uma explicação de como uma FRE tem *historicamente executado* uma determinada função própria (MILLIKAN, 1984, p. 33-4.). Se esta FRE tem *F* como função própria, uma explicação Normal para a execução de *F* consiste em uma explicação preponderante daqueles casos nos quais *F* foi historicamente executada. As condições que devem ser citadas na explicação Normal da execução de F

3 Para a concepção de reprodução, cf. Millikan (1989, p. 19-23).

são as condições Normais para a execução da função, a saber, aquelas condições explicativas preponderantes sob as quais *F* foi historicamente executada. Considere novamente o coração, cuja função própria é bombear sangue. A explicação Normal para como o coração historicamente bombeou sangue deve contar como o coração é produzido, como ele funciona internamente e mencionar condições como a da regularidade de impulsos elétricos enviados ao coração, o recebimento de oxigênio etc. Entretanto, não há uma única explicação Normal para como um item exerce sua função porque há explicações Normais mais e menos aproximadas, sendo esta a explicação *mais aproximada* de como o coração bombeia sangue[4]. A explicação mais aproximada não pode se referir à fonte dos impulsos elétricos ou do oxigênio enviados ao coração, o que é permitido às explicações menos aproximadas. *Condições Normais* são precisamente aquelas condições que devem ser mencionadas na explicação Normal mais aproximada de como um item exerce sua função. No caso da evolução, as condições Normais são aquelas em que o item que executa a função está biologicamente adaptado.

Note que as condições Normais não são as condições *estatisticamente* mais comuns ou habituais sob as quais os membros de uma *FRE* têm existido (justamente para evitar confusão, Millikan as denomina "Normais" com "N" maiúsculo), pois as duas noções não são equivalentes: as condições historicamente mais comuns em que os membros de uma *FRE* têm existido podem não coincidir com as condições sob as quais eles têm historicamente executado suas funções próprias. Por exemplo, raros são os espermatozoides que historicamente conseguiram realizar suas funções próprias (exceção feita a funções próprias mais imediatas, tal como a de nadar), uma vez que raros são os espermatozoides que executam suas funções de fecundar um óvulo.

A intuição por trás da noção de função própria é que uma função F é uma função própria do item x, se x tem o caráter C em virtude de poder executar F por ter C. Ou seja, porque houve ancestrais de x que puderam executar F em virtude de terem tido o caráter C é que a função própria de x é F. Mas como assegurar esta relação de causalidade? O que nos assegura que x foi produzido porque F foi executado por ter tido C e não que x foi produzido porque F foi executado por ter tido o caráter D, por exemplo? Como assegurar a relação causal entre ter tido o caráter C e a possibilidade de no passado executar F? A resposta de Millikan é que há uma relação causal na direção de B para A quando houve uma correlação positiva entre Bs e As e o fato de esta correlação ter existido figura na

4 Millikan (1984, p. 33) menciona também a possibilidade de explicações Normais alternativas.

explicação da proliferação de Bs, do que se segue que Bs existem em parte porque Bs causaram As.

Dizer que há uma correlação positiva entre B e A é dizer que há uma proporção maior de As que não-As que são Bs e, consequentemente, vice-versa. A correlação ocorre relativamente a alguma amostra de coisas e essa amostra evidentemente deve conter coisas que não são Bs (assim como coisas que não são As). No caso de um organismo, o que nos garante que x foi produzido porque F foi executado via C e não via D é que há uma correlação positiva entre a produção de x e a execução de F via C na história evolutiva de x, mas não há tal correlação entre a produção de x e a execução de F via D. Explicações fazendo referência a correlações de certo tipo são dadas para explicar porque certos traços dos organismos foram selecionados e outros não. Millikan fala em correlação positiva para ilustrar porque certos traços foram escolhidos na história evolutiva em detrimento de outros. Isto posto, vejamos finalmente em que consiste a definição completa de função própria direta (MILLIKAN, 1984, p. 28):

Função Própria Direta. Para o membro m de uma Família Reprodutivamente Estabelecida R de caráter reprodutivamente estabelecido C, m tem F como função própria direta se e somente se:

(1) Alguns ancestrais de m executaram F.

(2) Em parte por ter existido, entre os ancestrais de m, uma conexão causal direta entre ter o caráter C e a execução de F, C correlacionou-se positivamente com F sobre um conjunto de itens S que tem entre seus membros os ancestrais de m e outras coisas que não são C.
(3) Uma explicação legítima que pode ser dada para m existir faz referência ao fato de C ter sido positivamente correlacionado com F sobre S, seja por causar diretamente a reprodução de m ou por explicar porque R se proliferou e, assim, porque m existe.

De acordo com esta definição, o tipo de função biológica que ordinariamente atribuímos a mecanismos biológicos como corações e fígados têm funções próprias diretas, assim como artefatos e pegadas. Mas note que não é uma condição necessária para que um mecanismo tenha uma função própria direta que ele seja capaz de executar sua função, afinal ele pode ser constitutivamente mal formado. Isso ocorre porque esta definição implica que o que determina a função própria do mecanismo não são suas propriedades presentes ou disposições de executá-la (ou de ser capaz de executá-la), mas a *história do mecanismo*, ou seja, suas propriedades históricas. Nas palavras de Millikan (1989a, p. 289), "the definition of proper function looks to history rather than merely to present properties or dispositions to determine function".

Funções próprias diretas podem ser *relacionais*. Um mecanismo possui uma função própria relacional se sua função é fazer ou produzir uma coisa que tenha uma relação específica com outra coisa. O exemplo clássico de função própria relacional é o mecanismo de variação da cor da pele do camaleão de acordo com a cor da superfície sobre a qual está sentado. Obviamente, este mecanismo foi selecionado por ter o efeito de tornar o camaleão invisível para os predadores, contribuindo, portanto, para a sobrevivência e reprodução da espécie. Assim, a função relacional do mecanismo é a de produzir uma determinada cor para o camaleão que tenha a relação "mesma cor que" com a superfície sobre a qual o camaleão está sentado.

Quando um mecanismo A tem uma função relacional ele deve produzir algo que tenha certa relação com B, do que dizemos que B está *assim situado* com relação a A. Se há algo que corresponda a B, então A adquire uma *função própria adaptada*. B é então o adaptador atual para A. Uma função adaptada não passa de uma função relacional adaptada a um contexto. Por exemplo, caso o camaleão esteja sentado sobre uma superfície verde e marrom, é uma função própria adaptada deste mecanismo produzir a cor verde e marrom para a pele do camaleão. O que quer que *A* produza quando executa sua função adaptada é um mecanismo adaptado. Neste caso, o mecanismo adaptado é a cor da pele verde e marrom e o adaptador para A é a cor da superfície.

Uma função adaptada, ao contrário de uma função relacional, não pode ser uma função própria direta de um mecanismo. A configuração "o camaleão tem a relação de 'mesma cor que' para com a superfície sobre a qual ele está sentado" é membra de uma *FRE*, mas a cor de pele verde e marrom não é membra de nenhuma *FRE,* já que é possível que esta cor de pele seja *inédita*, de modo que em nenhum momento prévio da história do mecanismo de variação tal cor tenha sido produzida. Mas se esta cor não é membra de nenhuma *FRE*, então ela não pode ter função própria direta. Antes, o que ela possui é uma *função própria derivada*. A função própria de um mecanismo adaptado é derivada da função própria do mecanismo que o produziu (exceção feita à própria produção desse mecanismo adaptado). Assim, o mecanismo adaptado do camaleão (i.e., a cor de pele verde e marrom) tem a função própria derivada de evitar que o camaleão seja detectado por predadores, porque o mecanismo do camaleão que produziu esta cor de pele especifica (i.e., o mecanismo de mudança de cor de pele) tem a função própria direta de evitar que o camaleão seja detectado por predadores.

Uma cor de pele específica tem a função própria derivada *F* de evitar que o camaleão seja detectado por predadores caso esta cor tenha se originado como o produto do mecanismo de variação de cor de pele que executou *F* como sua função própria e, normal-

mente, causa a execução de *F* por meio da produção de uma cor de pele específica. Mas como pode haver uma explicação Normal para como o mecanismo adaptado executou sua função derivada *F*, dado que, por um lado, uma explicação Normal é uma explicação de como um mecanismo historicamente executou uma função e, por outro, um mecanismo adaptado pode nunca ter sido produzido antes?

Como uma função adaptada é uma função relacional adaptada a um dado contexto, se a função relacional é executada Normalmente, então a função adaptada também é executada Normalmente (MILLIKAN, 1984, p. 42-4; 1989a, p. 289), ao passo que quando o mecanismo produtor não executa sua função relacional Normalmente, o mecanismo adaptado será mal formado (ou seja, mal adaptado) e com isto não será capaz de executar sua função adaptada. Uma explicação Normal para a execução de uma função adaptada é uma explicação geral de como um mecanismo produz ou faz coisas que têm certas relações com seus adaptadores. No caso do camaleão, uma explicação Normal de como o mecanismo de variação de cor produz uma cor específica é uma explicação de como o mecanismo produz esta cor de acordo com a cor da superfície atual sobre a qual o camaleão está sentado.

Esta é a teoria das funções próprias de Millikan. Sua apresentação detalhada aqui se faz necessária para compreender a teleosemântica porque tanto o mecanismo produtor da representação, como a própria representação, tem função própria. Mas, enquanto o primeiro tem função direta, a segunda tem derivada. Vejamos como isto é possível.

A Teleosemântica

Estados como crenças, desejos e certos estados sensoriais representam a realidade. Estados mentais representacionais representam a realidade de uma dada maneira e a maneira pela qual estes estados representam a realidade é o seu *conteúdo representacional*. Por exemplo, minha crença que Recife está ao norte de Maceió tem o mesmo conteúdo que a sentença escrita no papel Recife está ao norte de Maceió, a saber, *Recife está ao norte de Maceió*. Ambas representam a mesma condição – Recife como estando localizada ao norte de Maceió. Mas o que confere o status de *representação* a um dado estado? Em virtude de quê um estado é representacional? Por fim, assumindo-se que se trata de um estado representacional, em virtude de quê tem este e não aquele conteúdo? Este é o problema da intencionalidade que Millikan procura solucionar através de sua teleosemântica.

Para Millikan, o que determina a *categoria biológica* de um item é sua função própria. Um item pertence à categoria biológica do coração se tem a função própria desta categoria, a saber, bombear sangue. Representações também formam uma categoria biológica e, por conseguinte, para um item constituir ou não uma representação depende de ter ou não a função própria desta categoria. Mas enquanto no caso do coração a única seleção envolvida é a evolutiva, no caso das representações há seleções ontogenéticas também envolvidas, notadamente, seleções por aprendizado. Assim, há tanto funções próprias filogenéticas como ontogenéticas. Ademais, não apenas as representações inatas têm funções próprias, mas também representações adquiridas ao longo do desenvolvimento do organismo e representações produzidas pela primeira vez na história da espécie.

A teleosemântica millikaniana é pluralista no que concerne a sua etiologia, mas isto não quer dizer que todas as funções envolvidas sejam igualmente fundamentais na determinação do conteúdo. De fato, a função evolutiva é a mais fundamental. Sem a seleção evolutiva que moldou os sistemas representacionais dos organismos biológicos, as outras formas de seleção que constituiriam outros tipos de função biológica não teriam sequer espaço de atuação. Dessa maneira, aqui nos concentraremos no problema de como funções próprias determinam o conteúdo daquelas representações, cujo único processo de seleção relevante para a determinação de seu conteúdo é a seleção evolutiva[5].

Considere um estado representacional de um organismo que foi produzido por algum mecanismo deste organismo, o *produtor da representação*. Millikan sustenta que a fim de determinar o conteúdo da representação, deve-se focar no mecanismo que *usa ou consome* a representação. Embora ambos os mecanismos sejam relevantes para determinar o conteúdo, é o *mecanismo consumidor* que possui o papel preponderante. É este também que faz com que um dado item seja uma representação, ou seja, que confere o status de representação: "[i]t is the devices that use representations which determine these to be representations and, at the same time [...] determine their content" (MILLIKAN, 1989b, p. 284).

Millikan sustenta que um sistema representacional é dividido em duas partes ou aspectos: o *produtor* e o *consumidor* da representação. O primeiro produz a representação para que o segundo a use ou consuma. Para compreender essa distinção, vejamos um caso em que os mecanismos consumidor e reprodutor pertencem a organismos distintos.

5 Para uma apresentação da atuação de outros processos de seleção na determinação do conteúdo de outras representações, cf. Millikan (1984; 1990).

Considere o caso da dança da espécie de abelha *Apis mellifera* (VON FRISCH, 1967). Há um mecanismo da abelha que tem a função de produzir, após a descoberta de uma fonte de néctar, uma dança específica que direciona as outras abelhas ao local do néctar. Estas abelhas espectadoras voam então nesta direção para coletar néctar e transportá-lo para a colmeia. A dança da abelha representa, portanto, o local do néctar. É evidente que o produtor da representação é o mecanismo produtor da dança presente na abelha dançante, ao passo que o consumidor é o mecanismo das abelhas espectadoras que usam a dança para direcionar o voo ao local do néctar. Mas em outros casos, talvez, estes mecanismos pertençam a um mesmo organismo, como no caso do anuro que detecta uma mosca e a captura. Aqui, o produtor é o sistema visual do anuro que produz uma representação da mosca, enquanto que os consumidores são o sistema motor e o sistema digestivo que respectivamente captura e digere a mosca.

Para Millikan, há dois tipos de representações, *indicativas* e *imperativas*. As primeiras são aquelas representações que devem ser determinadas pelos fatos e descrevem ao mecanismo consumidor que é o caso. Já as imperativas são as representações que devem determinar os fatos, ditando ao consumidor o que ele deve fazer. Estes dois tipos não são mutuamente excludentes, sendo possível a uma representação simultaneamente pertencer a ambos. Por exemplo, a dança da abelha é uma representação simultaneamente indicativa e imperativa, pois *descreve* o que é o caso para as abelhas espectadoras – o local do néctar – e também as *dita* o que fazer – voar em direção ao néctar. Aqui nos restringiremos às representações indicativas[6].

Aqui são necessárias algumas observações quanto às funções próprias dos mecanismos produtores e consumidores da representação constituída pela dança da abelha. O produtor tem como função própria imediata produzir uma dança que tem uma certa relação de correspondência com o local do néctar, de modo que uma variação no local do néctar corresponde a uma variação na forma da dança. Esta função é, portanto, *relacional*. Uma função menos imediata deste mecanismo é produzir, como resultado da dança, o voo das abelhas espectadoras em direção ao néctar. Já a função própria do consumidor da dança é a função relacional de produzir uma direção de voo correspondente à dança e, portanto, correspondente ao local do néctar. Note que a dança é um adaptador imediato para a direção do voo da abelha espectadora, mas a linha do voo é também adaptada ao adaptador da dança da abelha – i.e., o local do néctar –, já que é uma função do produtor produzir uma dança que tenha certa correspondência com o local do néctar.

6 Para a teoria das representações imperativas, cf. Millikan (1984, 1986, 2004, 2009).

Millikan propõe quatro requerimentos que um estado deve satisfazer para ser representacional[7]. No que se segue, "Normalmente" é uma abreviação de "quando executa suas funções próprias de acordo com uma explicação Normal".

(1) Requerimento da FRE. Uma representação é membro de uma *FRE* com *funções próprias diretas*.

Para compreender em que sentido uma representação como a da abelha pode pertencer a uma FRE dotada de função própria, é necessário distinguir funções próprias *variantes* e *invariantes*. As funções próprias derivadas de um mecanismo adaptado podem ser funções invariantes ou variantes. No caso das invariantes, a função derivada do mecanismo adaptado não é derivada de um adaptador, ao passo que no caso das variantes ela é uma função derivada adaptada, sendo estritamente derivada do mecanismo produtor e do adaptador. Por exemplo, suponha que uma dança específica de uma abelha aponta para o sudeste como o local do néctar. Esta dança tem como *função invariante* a de mostrar, em geral, às abelhas espectadoras o local em que há néctar e tem como *função derivada adaptada* – e *variante* – levar as abelhas espectadoras ao sudeste, pois é ali que nesta situação específica há néctar. Em suma, o voo da abelha espectadora em direção ao sudeste tem como função própria derivada *invariante* levar ao néctar em geral e tem como função própria *variante* levar ao néctar naquele local específico.

O mesmo mecanismo adaptado com funções próprias *derivadas*, na medida em que exibe um caráter *concreto*, pode também ser membro de uma FRE com funções próprias *diretas*, na medida em que exibe um caráter mais *abstrato* (MILLIKAN, 1984, p. 42). Uma dança específica da abelha tem funções próprias que são derivadas do mecanismo que a produziu e de seu adaptador (o néctar em um local específico), mas considerada meramente como uma dança que está de acordo com as regras sintáticas gerais para as danças das abelhas, esta dança é membro de uma FRE de ordem superior com funções próprias diretas. Sua função própria direta mais imediata é a função relacional de mover as abelhas

7 Em *Language, Thought and Other Biological Categories*, estas quatro condições são aquelas que um estado deve satisfazer para ser o que Millikan (1984, p. 97-9) denomina "ícone intencional". Nesta obra ela utiliza "representação" em um sentido estrito, de modo que representações primitivas, como a do local do néctar por parte da dança da abelha, seriam ícones intencionais. Entretanto, em *Biosemantics* (1989b) e *Varieties of Meaning* (2004) ela utiliza "representação" de maneira mais abrangente, de modo a incluir também representações primitivas. É neste último sentido que utilizamos "representação".

em certa direção, de tal modo que a forma concreta desta dança tem uma correspondência para com esta direção. Diante disto, temos que o que é invariável na execução da dança é sua *forma sintática geral*, o caráter Normal de funcionamento da FRE das danças das abelhas, ao passo que o variável é o conteúdo de ir *nesta ou naquela direção*, a depender do local atual do néctar, ao qual esta forma sintática invariante é imediatamente adaptada.

(2) Requerimento da Cooperação. Uma representação Normalmente está entre dois mecanismos cooperantes – um *produtor* e um *consumidor* – que foram selecionados pela evolução para se ajustar um ao outro, sendo a presença e a cooperação de um mecanismo uma *condição Normal* para a execução das *funções próprias* do outro.

Os mecanismos produtores e consumidores da dança são membros de FREs que foram selecionados pela evolução para cooperar um com o outro na execução de funções invariantes comuns (*e.g.*, obter mel). A presença e cooperação de um mecanismo é uma condição Normal para a execução das funções próprias do outro. Logo, a presença e cooperação do produtor da dança é uma condição Normal para a execução das funções próprias do consumidor da dança, porque caso o produtor esteja ausente ou em mau funcionamento, o consumidor não orientará adequadamente as abelhas espectadoras a respeito da direção em que devem voar para obter mel.

(3) Requerimento da Adaptação. Normalmente uma representação tem a função de *adaptar* o mecanismo consumidor às condições sob as quais as funções próprias do consumidor possam ser executadas.

O atual local do néctar é o adaptador original ao qual a dança adapta o mecanismo consumidor das abelhas espectadoras. Ao ser sinalizado, através da dança, sobre o local atual do néctar, o mecanismo consumidor adapta correspondentemente a direção do voo da abelha espectadora e a leva ao néctar.

(4) Requerimento do Mapeamento. A explicação Normal de como uma representação adapta o mecanismo consumidor, de modo a que ele possa executar suas funções próprias, faz referência ao fato de a representação fazer um *mapeamento* sobre algo de acordo com uma função de mapeamento específica.

Representações são mecanismos que devem fazer um dado mapeamento sobre objeto(s) no mundo a fim de cumprirem suas funções próprias, ou seja, Normalmente elas mapeiam de uma certa forma sobre este(s) objeto(s) quando executam suas funções próprias. Por exemplo, a dança da abelha mapeia, de acordo com certas regras, sobre uma configuração real de objetos – o néctar, o sol e a colmeia.

Diante destes quatro requerimentos, podemos finalmente estabelecer qual a função própria do mecanismo produtor, consumidor e do próprio estado representacional. O requerimento da cooperação estabelece que a presença e cooperação entre si dos produtores e consumidores da representação é uma *condição Normal* para que eles possam executar suas *funções próprias*. Mas como a produção de uma representação pode ter um efeito evolutivamente benéfico para o organismo? Considere um calo resultante do uso de uma roupa apertada que tem o efeito benéfico de proteger a pele de danos futuros. Diríamos que este calo representa o local em que a roupa estava? De um ponto de vista teleosemântico, certamente não. Mas o que faz com que a dança da abelha, mas não o calo, constitua genuinamente uma representação? Em primeiro lugar, o mecanismo produtor do calo não foi selecionado para produzir representações do local da roupa, ao passo que o mecanismo produtor da dança foi selecionado justamente para produzir representações do local do néctar. A história de seleção do calo mostra que ele foi selecionado para proteger a pele, sendo, portanto, esta sua função. Note que a produção de uma suposta representação do local da roupa seria um efeito colateral, não um efeito benéfico, da produção do calo. Em segundo lugar, a produção da dança é um meio de sinalizar às abelhas companheiras o local do néctar para que elas possam levá-lo à colmeia e, assim, contribuir para a reprodução da espécie, mas qual mecanismo consumiria a sinalização do local da roupa no caso da produção do calo? Nenhum, simplesmente *não há consumidor*.

A lição que Millikan (2004, p. 72-3) tira disto é que uma representação é produzida com o propósito de ser uma representação para algum consumidor, afinal não há sentido em um organismo produzir uma representação se nada irá reconhecê-la enquanto tal e consumi-la. Produtores foram selecionados para cooperar com os consumidores, que, por sua vez, foram selecionados para cooperar com os produtores. O que um mecanismo faz ajuda o outro e vice-versa. A representação é o estado que Normalmente está entre ambos.

A *função do produtor* é tão somente produzir o que seus consumidores necessitam. Mas como a representação produzida será interpretada ou consumida? Em que consiste interpretar ou consumir corretamente uma representação? A resposta de

Millikan (2004, p. 76) é que a representação será usada para *guiar* seus consumidores na execução de suas funções próprias e tal execução será bem-sucedida apenas se estiver de acordo com o que está sendo representado, ou seja, apenas se a execução estiver *de acordo* com a representação e esta *corresponder* ao estado de coisas representado. Normalmente, a execução das funções próprias dos consumidores será bem sucedida apenas porque o efeito da representação é adaptá-la ao estado de coisas representado. Note que é justamente isto o que estabelece o requerimento da adaptação.

Mas disto decorre que a execução Normal das funções dos consumidores da representação exige um *isomorfismo* entre a representação e o representado: variações no estado de coisas representado devem corresponder às variações na representação. Dado que as funções dos consumidores foram selecionadas para variar com a representação, há então uma função de mapeamento de acordo com a qual a representação deve *corresponder* ao mundo para que os mecanismos que a consumam possam ter êxito na execução de suas funções próprias. A função de mapeamento de uma representação nada mais é que a maneira pela qual ela mapeia sobre o que está sendo representado.

A partir disso, fica fácil constatar qual a função própria do produtor. Se sua função é apenas produzir o que o consumidor necessita para executar Normalmente suas funções próprias e tudo o que ele necessita para isto é que a representação produzida corresponda ao estado de coisas representado de acordo com uma função de mapeamento, então a *função do mecanismo produtor* é produzir uma representação que *corresponda* ao estado de coisas representado de acordo com tal função de mapeamento.

O consumidor é simplesmente um mecanismo que explora o mapeamento entre a representação e o representado para executar Normalmente suas funções próprias, do que decorre que é uma condição Normal para a execução destas funções a presença e cooperação do *produtor*. Obviamente o consumidor pode ainda executá-las caso o produtor esteja ausente ou caso a representação não esteja de acordo com a função de mapeamento, mas neste caso a execução das funções próprias do consumidor seria acidental. Já para o produtor executar sua função, ele necessita da presença e cooperação do consumidor, uma vez que o que determina a função de mapeamento envolvida na representação produzida são as necessidades do consumidor (MILLIKAN, 1989b, p. 286), de maneira que caso o consumidor esteja ausente, ou não coopere, nenhuma função de mapeamento será determinada e o produtor não poderá produzir uma representação de acordo com esta função.

Tendo determinado em que consiste a função própria do mecanismo produtor, vejamos em que consiste a *função própria da representação*. A função do produtor é a função

direta de produzir uma representação que *corresponda* ao que está sendo representado de acordo com certa função de mapeamento. A produção da representação é o *meio* ao qual o produtor recorre para adaptar o consumidor às condições sob as quais suas funções próprias possam ser executadas. Sendo esta a função direta do produtor, segue-se que a *função própria derivada da representação* será adaptar o mecanismo consumidor às condições sob as quais as funções próprias do consumidor possam ser executadas. A maneira pela qual a representação faz isto é correspondendo a certo estado de coisas de acordo com uma função de mapeamento. Note que é justamente isto o que estabelece o requerimento da adaptação.

No caso da abelha, uma representação específica do local do néctar tem a função derivada de adaptar os consumidores às condições sob as quais suas funções possam ser executadas em virtude do produtor da representação ter a função direta de adaptar o consumidor às condições sob as quais suas funções próprias possam ser executadas. Tal representação não passa de um *mecanismo adaptado* ao local em que o néctar se encontra no ambiente.

Até agora nos restringimos a analisar os critérios de Millikan para um item ser uma representação. Para um estado membro de uma *FRE* com funções próprias ser uma *representação*, deve haver Normalmente uma cooperação entre o produtor e o consumidor (*requerimento da cooperação*), o estado em questão Normalmente deve adaptar o consumidor às condições sob as quais as funções próprias deste mecanismo possam ser executadas (*requerimento da adaptação*) e a explicação Normal de como esta adaptação ocorre faz referência ao fato de este estado fazer um mapeamento de acordo com uma função de mapeamento (requerimento do mapeamento). Contudo, estes requerimentos, por si só, não determinam o que está sendo representado, i.e., o conteúdo. Representações são estados que Normalmente devem fazer um mapeamento sobre o mundo a fim de cumprirem suas funções próprias. Seja P uma representação. Dado que há inúmeros mapeamentos possíveis de P sobre o mundo, como distinguir um mapeamento dos demais para determinar o conteúdo de P? É preciso um critério para escolher, por princípio, uma função de mapeamento em detrimento das demais. Tal critério passa, evidentemente, pela história de seleção (MILLIKAN, 1984, p. 100):

Conteúdo Representacional. O item *P* é uma representação do que quer que P mapeie que deve ser mencionado na explicação Normal mais aproximada da execução das funções próprias de seus consumidores tal como adaptados a *P*. O conteúdo de P é o estado de coisas ao qual P deve corresponder a fim de que seus consumidores possam *executar Normalmente* suas funções próprias.

Ou seja, *P* representa o estado de coisas ao qual *P* adapta seu consumidor. É uma condição Normal para a execução das funções próprias dos consumidores da representação que a representação e o representado estejam de acordo entre si.

Uma condição Normal para a execução de uma função própria é uma condição que deve ser mencionada na explicação Normal mais aproximada da execução desta função. A explicação Normal mais aproximada da execução das funções próprias do consumidor de um item não faz referência a qualquer evento que ocorra antes de sua produção. Na cadeia de eventos que ocorre entre o início da produção do item até o término de seu consumo, tal explicação parte do ponto em que as atividades do consumidor têm início e explica como este historicamente executou suas funções. Assim, a explicação Normal mais aproximada de como os consumidores de *P* executam suas funções não pode fazer referência a como *P* foi produzido. Ao dar a explicação Normal mais aproximada de como os consumidores de *P* executam suas funções próprias, é preciso mencionar apenas o fato de alguma variável no ambiente circundante ser mapeada por *P*, não o fato de como *P* foi produzido. Por exemplo, a explicação Normal mais aproximada de como os consumidores da dança da abelha executam suas funções próprias faz referência ao fato de o néctar estar a certa distância do sol e da colmeia, mas não faz qualquer referência a como tal dança foi produzida.

O conteúdo da representação é determinado pelo estado de coisas ao qual a representação deve corresponder para que seus consumidores possam executar Normalmente suas funções próprias. Disto se segue que o conteúdo não se assenta sobre a univocidade da função dos consumidores da representação, mas na igualdade das condições Normais para a execução destas funções. Esta peculiaridade da teleosemântica de Millikan terá um papel fundamental na sua resposta ao problema do conteúdo distante, como veremos.

Uma questão, entretanto, surge ao refletirmos sobre a abordagem millikaniana para o conteúdo representacional. A condição Normal de haver certa correspondência entre a representação e o representado é uma condição Normal entre tantas outras. Por exemplo, é uma condição Normal para que as abelhas espectadoras possam obter néctar que haja uma correspondência entre a coreografia da dança e o local do néctar, mas é igualmente uma condição Normal para a execução desta função a presença de oxigênio, de modo a que estas abelhas possam respirar e, assim, voarem para o néctar. Mas ora, porque então o conteúdo da dança depende da condição Normal de haver uma correspondência entre a dança e o local do néctar e não da condição Normal da presença de oxigênio?

A resposta de Millikan é simples e direta: o mecanismo produtor da dança foi selecionado para produzir uma dança que mapeie sobre o local do néctar, não sobre a presença de oxigênio. O tempo e o local da dança variam não com a presença ou ausência de oxigênio, mas com o tempo e local do néctar (mais precisamente, com o tempo e o local do néctar tal como relacionado com o sol e a colmeia). Mas para compreender em que consistem os aspectos variantes e invariantes de uma representação, é necessário ver em que consiste precisamente o tipo geral de função de mapeamento de uma representação.

Quando uma representação de um estado de coisas é verdadeira, ela é relacionada a este estado de coisas da seguinte maneira (MILLIKAN, 1984, p. 107): (I) o estado de coisas é uma condição Normal para a execução das funções próprias diretas da representação; (II) há operações sobre a representação que têm uma correspondência um para um com operações sobre o estado de coisas; (III) qualquer transformação da representação resultante de uma destas operações tem como uma condição Normal para a execução das funções próprias da representação uma transformação correspondente no estado de coisas.

A tese por trás desta concepção do mapeamento da representação sobre o estado de coisas remete ao *Tractatus logico-philosophicus* de Ludwig Wittgenstein (1922 [2017]) por sustentar que o que corresponde, em primeira instância, às transformações na representação são transformações no estado de coisas, não transformações nos elementos do estado de coisas[8]. O que quer que seja considerado como sujeito a um conjunto de transformações é articulado. A representação é articulada não em elementos, mas em aspectos variantes e invariantes. O que não muda ao longo de todas as transformações possíveis sobre uma representação é o seu aspecto invariante, enquanto que o seu aspecto variante é aquilo que é modificável ao longo do conjunto de transformações.

Transformações na dança da abelha (*e.g.*, girar o ângulo do eixo da dança em 20º no sentido horário) correspondem às transformações um para um na relação entre o sol, a colmeia e o néctar que está sendo mapeada. A dança representa o local do néctar através da representação da relação entre o sol, o néctar e o ambiente, de modo que transformações na dança correspondem às transformações biunívocas no local do néctar relativo ao sol e à colmeia. É difícil especificar exatamente o que é *invariante* na coreografia da dança, mas o que é invariante no estado de coisas representado são os *relata* da relação mapeada: sol–néctar–colmeia. Assim, não é possível uma transformação na dança que corresponda a uma substituição do sol pela lua no que está sendo representado, de modo que a dança

8 Para uma defesa desta tese, cf. Millikan (1984, p. 102-7).

resultante mapeia lua–néctar–colmeia. Entretanto, podem ocorrer transformações quanto à distância entre o sol, o néctar e a colmeia, o que demonstra que este é um aspecto *variante* do estado de coisas representado. Uma vez que a função de mapeamento, de acordo com a qual a dança mapeia sobre o mundo, não faz qualquer referência à presença de oxigênio, segue-se que a presença de oxigênio não está sendo representada pela dança.

Por fim, as condições Normais para a execução das funções dos consumidores da representação são aquelas em que tais consumidores estão biologicamente adaptados, dado que a seleção envolvida é a seleção evolutiva. Para determinar o conteúdo de uma representação devemos, primeiro, olhar para aquelas condições na história evolutiva nas quais os consumidores da representação contribuíram para a adaptação da espécie. Posteriormente, devemos descobrir qual mapeamento entre a representação e o mundo foi requerido para que esta contribuição pudesse ocorrer, ou seja, qual mapeamento nestas ocasiões permitiu esta contribuição para a adaptação da espécie. O conteúdo da representação é precisamente aquele estado de coisas que foi *mapeado* nestas ocasiões. Por exemplo, considere a batida na água da cauda de um castor que faz com que os outros castores fujam do local. A batida na água representa *perigo* porque na história evolutiva dos castores que consumiram esta representação apenas quando a batida mapeou sobre um *predador* presente e os castores posteriormente fugiram é que houve contribuição adaptativa, já que tal fuga evitou a *captura*.

Indeterminação Funcional: O Problema do Conteúdo Distante

A teleosemântica de Millikan tem sido enaltecida como uma das principais teorias da representação mental desde a sua primeira formulação. Contudo, também tem sido alvo de intenso escrutínio e diversos problemas a ameaçam. Aqui focaremos no *problema do conteúdo distante*, um caso clássico de indeterminação funcional que, à primeira vista, parece inviabilizar a determinação do conteúdo por parte da teleosemântica millikaniana. Ora, se esta determina o conteúdo a partir da noção de função própria e atribuições de funções próprias são indeterminadas, segue-se que o conteúdo é ele mesmo *indeterminado*, o que seria uma consequência inaceitável.[9] No que se segue apresentaremos o que são casos de indeterminação funcional *em geral* para em seguida focarmos no problema do conteúdo distante.

9 Para uma avaliação de outros casos de indeterminação funcional que assolam a teleosemântica millikaniana, cf. Ryder & Kingsbury & Williford (2013); Neander & Schulte (2022).

Casos problemáticos de indeterminação funcional são aqueles nos quais parece haver razões igualmente plausíveis para atribuir funções incompatíveis a um dado mecanismo. Isto resulta em uma indeterminação na atribuição de sua função, já que é indeterminado quando o mecanismo está em bom ou mau funcionamento. Abordaremos os casos de atribuições funcionais através do clássico exemplo que deu início a este debate: as bactérias anaeróbicas (DRETSKE, 1994, p. 164-8).

Algumas bactérias marinhas possuem ímãs internos que se alinham ao campo magnético da Terra e com isso alinham a própria bactéria. No hemisfério norte, estes ímãs se inclinam em direção ao campo geomagnético do norte, fazendo com que a bactéria se mova para baixo, em direção ao fundo do mar. Uma vez que ambientes ricos em oxigênio são letais para a bactéria, o ímã serve para livrá-la do oxigênio, na medida em que a afasta da superfície marítima – um ambiente rico em oxigênio. Ao levar a bactéria em direção ao norte geomagnético e com isto para o fundo do oceano, o ímã contribui para sua sobrevivência, já que este é um ambiente pobre em oxigênio. No caso de bactérias do hemisfério sul, os ímãs são invertidos e direcionam a bactéria para o sul geomagnético, o que a leva para o fundo do mar, tendo, portanto, o mesmo benefício.

Caso uma bactéria do hemisfério sul seja transportada para o hemisfério norte, seu ímã a levará à autodestruição já que fará com que ela se mova em direção à superfície, já que estará indo em direção ao sul geomagnético. Esta autodestruição também ocorre caso ponhamos próximo à bactéria uma barra magnética orientada em direção oposta ao campo geomagnético, também a levando para a superfície.

À primeira vista, este parece ser um caso de representação falsa – uma vez que no habitat natural da bactéria seu ímã interno a direciona para um ambiente com pouco oxigênio, parece razoável dizer que a função do ímã é direcionar a bactéria para o ambiente com pouco oxigênio. Assim, quando na presença de uma barra magnética, o mecanismo falsamente representa o ambiente pobre em oxigênio, já que estará direcionando a bactéria para a superfície.

Casos de indeterminação funcional surgem quando se questiona o que garante que esta é a descrição correta da função do ímã. Por que descrever sua função como a de direcionar a bactéria para um ambiente *livre de oxigênio*? Por que não descrevê-la como a de direcionar a bactéria para o *norte geomagnético*? Ou como a de direcionar a bactéria ao *campo magnético prevalecente*?

A depender de como se descreve a função do ímã, em certos casos haverá mau funcionamento e, assim, falsa representação, e em outros não. Se sua função é direcionar a bactéria para um ambiente livre de oxigênio, o sistema estará em mau funcionamento quando levar a bactéria a um ambiente rico em oxigênio, havendo, portanto, falsa representação. Mas se sua função é direcionar a bactéria para o campo magnético prevalecente, o sistema estará funcionando perfeitamente bem quando, sob influência da barra magnética, levar a bactéria para um ambiente rico em oxigênio, sendo este agora um caso de representação verdadeira.

Deve haver algo de errado neste último caso, mas assumindo-se como a função do sistema apontar a direção do campo magnético prevalecente, não se pode responsabilizá-lo por este erro. Dretske (1994) sugere que, neste caso, o erro talvez esteja na *correlação* do ambiente – entre a direção do campo magnético e a direção de condições anaeróbicas –, que faz com que o ímã interno sirva para direcionar a bactéria ao local em que há pouco oxigênio.

Isto mostra que enquanto não esteja determinada a função do ímã, não há como determinar o seu *status de bom funcionamento*, independentemente do que seja responsável pelo erro. Isto é, não há como especificar em quais casos há mau funcionamento (logo, representação falsa) e em quais há bom funcionamento (logo, representação verdadeira). Talvez o mais intuitivo seja sustentar que a função do ímã é direcionar a bactéria para ambientes anaeróbicos – não para o campo magnético –, porque o que garante sua sobrevivência é se dirigir ao ambiente anaeróbico. Contudo, recorrer tão somente à necessidade biológica não determina por si só a função do ímã. É justamente aqui que entra em cena o *problema do conteúdo distante*.

Dado que um sistema *O* necessita de *F* e que o mecanismo *M* permite a *O* detectar *F*s, não se segue que *M* representa *F*, porque se *F* e *G* são correlacionados no ambiente natural de *O*, então há duas maneiras para *M* detectar *F*s: *M* representa a presença de *F*s, direcionando assim *O* para o que ele necessita ou *M* representa a presença de *G*s e posto que no ambiente natural de *O* sempre que algo instancia *F*, também instancia *G*, segue-se que ao direcionar *O* para *G*s, *M* estará também direcionando *O* para *F*s (DRETSKE, 1994, p. 167).

Esta objeção é particularmente poderosa porque há vários exemplos de ambientes nos quais há este tipo de correlação. No caso da bactéria, sua necessidade é estar em um local livre de oxigênio e o mecanismo que a permite detectar este local é seu ímã interno, mas tal detecção pode ser feita através das *mais diversas funções*: apontar para a direção do

campo magnético prevalecente, para ambientes livres de oxigênio etc. No ambiente natural da bactéria, o local que instancia estas propriedades é rigorosamente o mesmo: o fundo do mar. Da mesma maneira, não se segue do fato que um animal necessita de vitamina C e que é portador de um mecanismo que o permite detectar vitamina C que este mecanismo tem a *função* de detectar vitamina C. Afinal, esta necessidade é também satisfeita caso o animal aponte para alimentos ricos em vitamina C, como laranja e limão.

O que ocorre nestes casos é que se um mecanismo sensorial é capaz de detectar apenas uma propriedade do ambiente, segue-se que se este mecanismo direciona o organismo para uma *propriedade distante* (*e.g.*, ausência de oxigênio), ele também irá direcioná-lo para alguma *propriedade mais próxima* (*e.g.*, campo magnético local). Se houve seleção natural para o mecanismo direcionar o organismo para onde há instância da propriedade distante *F*, também houve seleção natural para o mecanismo direcioná-lo para onde há instância da propriedade próxima *G*, posto que, como visto, o direcionamento para *G*s acarreta em direcionamento para *F*s.

O *problema do conteúdo distante* é que parece ser indeterminado se a função própria do mecanismo produtor é detectar a *condição distante* ou a *condição próxima*. Logo, a função própria da representação também é *indeterminada*. Mas o que dizer do mecanismo consumidor? Seria sua função levar a bactéria em direção à condição anaeróbica ou ao campo magnético prevalecente? A teleosemântica de Millikan parece implicar que o conteúdo da representação é também indeterminado. A fim de solucionar este problema, cabe a Millikan explicar o porquê de a função da representação ser a detecção da propriedade distante F, não da propriedade próxima G. Ou seja, explicar como a representação pode representar a condição distante sem estar representando a condição próxima.

Millikan sustenta que uma das vantagens de sua teleosemântica é que a solução do problema do conteúdo distante passa por um enfoque nas condições em que o *mecanismo consumidor* da representação executa suas funções próprias. O conteúdo da representação produzida pelo ímã é determinado por uma *condição Normal* que deve ser satisfeita para que o mecanismo consumidor da representação possa executar sua função própria, a saber, a condição Normal de haver uma correspondência entre a representação e o estado de coisas representado. Assim, no caso da bactéria anaeróbica o conteúdo é *ambiente livre de oxigênio* porque os consumidores da representação necessitam ser direcionados ao ambiente livre de oxigênio – não ao campo magnético – para que possam executar Normalmente suas funções próprias, independentemente da questão ulterior se o ambiente livre de oxigênio é também o ambiente em que está o campo magnético.[10]

10 "What the magnetosome represents is only what its *consumers* require that it correspond to in order

Os ímãs internos foram selecionados para produzir representações de condições anaeróbicas, justamente para guiar os consumidores da representação na execução Normal de suas funções. Para os consumidores, pouco importa que sejam direcionados para o campo magnético ou para o fundo do mar, o que importa é que sejam direcionados para condições anaeróbicas, afinal o direcionamento para um ambiente em que está o campo magnético, mas que não seja livre de oxigênio, seria mortal para a bactéria. Disto se segue que nenhuma propriedade mais próxima está sendo representada pelo ímã, apenas a *propriedade distante* de ser um ambiente *livre de oxigênio* está sendo representada.

Segundo Millikan, a função do ímã é detectar condições anaeróbicas. Caso coloquemos próximo da bactéria uma barra magnética orientada em direção oposta ao campo geomagnético, o ímã a orientará em direção à superfície marítima, um ambiente rico em oxigênio. Neste caso, o imã estará em mau funcionamento, na medida em que não direcionou a bactéria para condições anaeróbicas. Mas seria mesmo *plausível* dizer que o ímã está em mau funcionamento quando, em virtude da presença da barra magnética, direciona a bactéria para a superfície?

Ora, caso a barra não estivesse presente, o ímã desempenharia seu papel de direcionar a bactéria para o fundo do mar. Como bem observou Dretske (1994), o ímã está funcionando perfeitamente bem quando, sob influência da barra magnética, direciona a bactéria para a superfície, afinal o responsável por ele ter erroneamente apontado para a superfície foi a *barra magnética*, não um *defeito* em seu funcionamento. Millikan (1991, p. 161) argumenta que mesmo levando isto em consideração o ímã está em mau funcionamento, já que não executa sua função de detectar condições anaeróbicas. O que ocorre aqui é que há *duas noções* distintas de *mau funcionamento* por trás desta aparente incoerência.

Por um lado, há o *sentido estrito* de "mau funcionamento": um mecanismo está em mau funcionamento caso não execute sua função devido a algum defeito em seu *funcionamento interno*. Por exemplo, uma cafeteira estará em mau funcionamento se mesmo nela havendo pó de café, ela não produza café em virtude da danificação de seu fusível térmico. Por outro lado, há o *sentido amplo* de "mau funcionamento": um mecanismo está em mau funcionamento caso *simplesmente não execute suas funções*, ainda que a causa disto não

to perform *their* tasks. Ignore, then, how the representation (a pull-in-a-direction-at-a-time) is normally produced. Concentrate, instead, on how the systems that react to the representation work, on what these systems need in order to do their job. What they need is only that the pull be in the direction of oxygen-free water at the time" (MILLIKAN 1989b, p. 290).

seja um defeito em seu funcionamento interno. Por exemplo, uma cafeteira elétrica pode estar em perfeito funcionamento interno, mas caso nela não haja pó de café, ela não estará exercendo sua função de produzir café.

O ímã da bactéria está em mau funcionamento no *sentido estrito* se por algum defeito biológico ele não estiver funcionando internamente bem, não executando então sua função de direcionar a bactéria às condições anaeróbicas, ainda que todas as condições externas necessárias para a execução desta função sejam satisfeitas. O ímã estará em mau funcionamento no *sentido amplo* caso simplesmente não cumpra sua função, mesmo que esteja internamente em perfeito estado. O que explica isto é que as condições externas indispensáveis para a execução de sua função não estão satisfeitas, isto é, a *correlação do ambiente* está falhando. Em suma, só há bom funcionamento do ímã no sentido amplo caso *simultaneamente* haja bom funcionamento interno e as condições externas indispensáveis estejam satisfeitas, ao passo que há bom funcionamento do ímã no sentido estrito caso o ímã esteja em bom funcionamento interno, independentemente da satisfação destas condições externas.

A solução de Millikan para o problema do conteúdo distante consiste em determinar o conteúdo a partir do que os *consumidores* necessitam da representação a fim de executar Normalmente suas funções próprias, a saber, que a representação efetivamente corresponda ao estado de coisas representado de acordo com uma função de mapeamento. O mapeamento necessário para os consumidores executarem suas funções é aquele mapeamento que foi necessário para os ancestrais destes consumidores executarem suas funções, contribuindo assim para a *adaptação* da espécie. O conteúdo da representação do ímã é *ambiente livre de oxigênio* porque à luz dos consumidores ancestrais desta representação, constata-se que eles necessitavam que a representação mapeasse sobre *condições anaeróbicas* a fim de executar a função de produzir o movimento da bactéria nesta direção.

O apelo à condição Normal para a execução da função dos consumidores da representação permite a Millikan determinar o conteúdo independentemente do quão distante esteja a propriedade representada na cadeia causal de estímulos que causou a produção da representação. A propriedade de ser um ambiente anaeróbico é mais distante que a de ser o campo magnético prevalecente, já que é detectando instâncias desta última que o ímã detecta instâncias da primeira. Contudo, a propriedade representada é a de condições anaeróbicas porque a condição Normal que deve ser satisfeita para a execução das funções dos consumidores é que a representação mapeie sobre o ambiente livre de oxigênio. Millikan defende que o conteúdo é determinado a partir do *benefício adaptativo* resultante do *consumo* da representação.

Conclusão

No centro da teleosemântica de Millikan está a tese que o conteúdo é determinado pela condição Normal para a execução da função própria do consumidor da representação. Assim, o que Millikan propõe é uma teleosemântica *baseada no consumidor*. Note como sua solução para o problema do conteúdo distante depende fundamentalmente disto. De fato, à luz da supracitada tese, Millikan tem êxito em solucionar este problema. Mas seria a teleosemântica baseada no consumidor a maneira adequada de desenvolver o ponto de partida de todas as teleosemânticas, a saber, a tese de que o conteúdo é determinado pela função biológica do estado representacional?

Muitos teleosemanticistas discordam de Millikan, argumentando que a teleosemântica baseada no consumidor deve ser substituída por uma *teleosemântica baseada no produtor*: o conteúdo deve ser determinado a partir da função biológica do mecanismo produtor do estado representacional (DRETSKE, 1994; NEANDER, 2017; SCHULTE, 2018; SOUZA FILHO, 2018). Ainda que a teleosemântica millikaniana tenha êxito em solucionar o problema do conteúdo distante, ela talvez colapse diante de outro caso de indeterminação funcional, a saber, o *problema dos papéis causais complexos* (DRETSKE, 1994; NEANDER, 1995; PAPINEAU, 2003; SOUZA FILHO, 2013). De todo modo, o debate quanto à viabilidade destes dois tipos de teleosemântica tem sido uma das principais fontes de divisão interna entre os teleosemanticistas.

Outro problema que aflige a teleosemântica de Millikan é que ela seria muito *liberal* quanto a que estados sensoriais são genuinamente representacionais. Ou seja, ela seria uma teoria da representação que trata diversos estados sensoriais que claramente não são representacionais como estados genuinamente representacionais (FODOR, 1986; BURGE, 2010; SCHULTE, 2015), tais como os estados da bactéria anaeróbica, planária, ameba, etc. O desafio para Millikan seria então estabelecer quais são as *condições mínimas* que um dado estado sensorial deve satisfazer para constituir um estado representacional genuíno e demonstrar que sua teleosemântica é plenamente compatível com as condições mínimas para intencionalidade. Contudo, cremos que ela dificilmente terá sucesso nesta empreitada (anteriormente já argumentamos que a teleosemântica millikaniana é incompatível com as condições mínimas para a intencionalidade, cf. SOUZA FILHO, 2022).

Por fim, há objeções gerais à teleosemântica que evidentemente também se aplicam à teleosemântica de Millikan. Dentre estas, talvez a mais problemática seja aquela que a teleosemântica é incapaz de explicar a capacidade representacional de esta-

dos representacionais mais *sofisticados* (NEANDER & SCHULTE, 2022). Note que até aqui aplicamos a teleosemântica de Millikan apenas aos estados representacionais sensoriais *primitivos*, tais como os da bactéria anaeróbica, abelha e castor. Mas como determinar, em termos da noção de função biológica, os conteúdos das representações de número imaginário, democracia, átomo e gênero de faroeste?

Evidentemente, este é um problema não apenas para a teleosemântica, mas para diversas teorias da representação (especialmente aquelas de caráter naturalista). Contudo, tem se argumentado que este problema é particularmente desafiador para a teleosemântica na medida em que esta procura determinar o conteúdo representacional a partir da noção de *efeitos selecionados* (PEACOCKE, 1992). Millikan (1984, 2000, 2004, 2017) tem procurado dar conta deste problema, mas não está claro que ela obtém sucesso nesta missão.

Nosso objetivo neste artigo foi interpretar e analisar a complexa teleosemântica desenvolvida por Millikan ao longo das últimas décadas. Nesta última seção fizemos uma breve análise dos principais problemas que afligem a viabilidade desta teoria, sendo nossa conclusão que a teleosemântica de Millikan é dificilmente viável. Não obstante, é inegável a sua enorme e decisiva contribuição para a nossa compreensão da natureza da intencionalidade, constituindo um marco na filosofia da mente contemporânea[11].

Referências

BURGE, T. **Origins of objectivity**. Oxford: Oxford University Press, 2010.

CRANE, T. **The mechanical mind**. Oxford: Routledge, 2016.

DRETSKE, F. Misrepresentation. In: STICH, S. P. & WARFIELD, T. A. (Orgs.). **Mental representation:** A reader. Oxford: Basil Blackwell, 1994. p. 157-73.

FIELD, H. Mental representation. In: STICH, S. P. & WARFIELD, T. A. (Orgs.). **Mental representation:** A reader, Oxford: Basil Blackwell, 1994. p. 34-77.

FODOR, J. Why Paramecia don't have mental representations. **Midwest Studies in Philosophy**, v. 10, n. 1, 3–23, 1986.

11 Este artigo é um desenvolvimento e refinamento de uma pesquisa inicialmente realizada ao longo de meu mestrado no Programa de Pós-Graduação Lógica e Metafísica da Universidade Federal do Rio de Janeiro, quando tive apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Agradeço a Guido Imaguire por críticas e sugestões ao longo das primeiras versões dos manuscritos que posteriormente resultaram neste artigo.

MACDONALD, G. & PAPINEAU, D. **Teleosemantics**. Oxford: Oxford University Press, 1987.

MILLIKAN, R. **Language, thought, and other biological categories**. Cambridge, Massachusetts: MIT Press, 1984.

MILLIKAN, R. Thoughts without laws: Cognitive science with content. **The Philosophical Review**, v. 95, n. 1, p. 47-80, 1986.

MILLIKAN, R. In defense of proper functions. **Philosophy of Science**, v. 56, n. 2, p. 288-302, 1989a.

MILLIKAN, R. Biosemantics. **Journal of Philosophy,** v. 86, n. 6, p. 281-97, 1989b.

MILLIKAN, R. Truth rules, hoverflies and the Kripke-Wittgenstein paradox. In: MILLER, A. & WRIGHT, C. (Orgs.). **Rule-following and meaning**. Montreal e Kingston: McGill-Queen's University Press, 2002. p. 209-33.

MILLIKAN, R. Speaking up for Darwin. In: LOEWER, B. & REY, G. (Orgs.). **Meaning in mind:** Fodor and his critics. Cambridge, Massachusetts: Blackwell, 1991. p. 151-64.

MILLIKAN, R. **On clear and confused ideas:** An essay about substance concepts. Cambridge: Cambridge University Press, 2000.

MILLIKAN, R. **Varieties of meaning**. Cambridge, Massachusetts: MIT Press, 2004.

MILLIKAN, R. Biosemantics. In: BECKERMANN A.; MCLAUGLHIN B.P. & WALTER S. (Orgs.). **The Oxford handbook of philosophy of mind**. Oxford: Oxford University Press, 2009. p. 394-406.

MILLIKAN, R. **Beyond concepts:** Unicepts, language, and natural information. Oxford: Oxford University Press, 2017.

NEANDER, K. **A mark of the mental:** In defense of informational teleosemantics. Cambridge, Massachusetts: MIT Press, 2017. .

NEANDER, K. & SCHULTE, K. Teleological theories of mental content. In: ZALTA, E. (Org). **Stanford encyclopedia of philosophy**, 2022. Disponível em: < https://plato.stanford.edu/archives/sum2022/entries/content-teleological/>. Acesso em 09 jul. 2023.

PEACOCKE, C. **A study of concepts**. Cambridge, Massachusetts: MIT Press, 1992.

RYDER, D.; KINGSBURY, J. & WILLIFORD, K. **Millikan and her critics**. Chichester: Wiley Blackwell, 2013.

SCHULTE, P. Perceptual representations: A teleosemantic answer to the breadth-of-application problem. **Biology & Philosophy**, n. 30, v. 1, p. 119–136, 2015.

SCHULTE, P. Perceiving the world outside: How to solve the distality problem for informational teleosemantics. **Philosophical Quarterly**. v. 68, n. 271, p. 349–69, 2018.

SOUZA FILHO, S. F. **Seguir regras e naturalismo semântico.** Dissertação (Mestrado em Filosofia) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2013. Disponível em: https://ppglm.files.wordpress.com/2008/12/dissertacao-ppglm-sergio-souza.pdf. Acesso em 20 jul. 2023.

SOUZA FILHO, S. F. **Naturalising intentionality:** A teleological approach. Tese (Doutorado em Filosofia) – King's College London, Londres, 2018. Disponível em: < https://kclpure.kcl.ac.uk/ws/portalfiles/portal/103712183/2018_Farias_de_Souza_Filho_Sergio_1465852_ethesis.pdf. Acesso em 20 jul. 2023.

SOUZA FILHO, S. F. A dual proposal of minimal conditions for intentionality. **Synthese**, 200: 115, p. 1-22, 2022.

VON FRISCH, K. **The dance language and orientation of bees**. Cambridge, Massachusetts: Harvard University Press, 1967.

WITTGENSTEIN, L. **Tractatus logico-philosophicus**. São Paulo: Editora da Universidade de São Paulo, 2017 [1922].

A CRÍTICA DE JOHN R. SEARLE À NOÇÃO DE INCONSCIENTE E PERCEPÇÃO INCONSCIENTE

João Paulo Maciel de Araujo[1]

Resumo:

O presente texto tem por objetivo apresentar as críticas de Searle à noção de inconsciente e percepção inconsciente. Para esse propósito, é preciso introduzir alguns conceitos referentes à sua teoria da percepção, uma vez que suas críticas têm como pressuposto uma defesa da consciência e da intencionalidade perceptual. De acordo com Searle, dada a dificuldade que é prover uma descrição adequada da consciência enquanto objeto de investigação empírica, filósofos e cientistas se sentem mais inclinados a investigar processos cognitivos subjacentes ao nível da consciência, isto é, processos que não dependem da consciência para serem observados. Esses processos foram chamados de processos inconscientes e o objetivo de Searle é mostrar o quão problemático pode ser a noção de inconsciente. O que comumen-

Abstract:

This work aims to present Searle's criticisms of the notion of the unconscious and unconscious perception. For this purpose, it is necessary to introduce some concepts referring to his theory of perception because his criticisms are based on a defense of consciousness and perceptual intentionality. According to Searle, given the difficulty of providing an adequate description of consciousness as an object of empirical investigation, philosophers and scientists feel more inclined to investigate underlying cognitive processes at the level of consciousness, that is, processes that do not depend on consciousness to be observed. These processes have been called unconscious processes and Searle's aim is to show how problematic the notion of the unconscious can be. What are commonly known as unconscious processes

1 Doutor em Filosofia pelo programa integrado de Pós-Graduação em Filosofia UFPB-UFPE-UFRN. Mestre em Filosofia pela Universidade Federal de Pernambuco com período sanduíche na Universidad de Buenos Aires pelo Programa Capes PPCP-Mercosul. Professor horista no colegiado de filosofia da Universidade Estadual de Roraima. Membro do grupo de pesquisa Escola Amazônica de Filosofia – EAF.

te é conhecido como processos inconscientes nas ciências cognitivas são, na verdade, processos neurobiológicos interpretados como sendo intencionais e, portanto, tomados, muitas vezes, como estados mentais. Na visão de Searle, um estado mental inconsciente é aquele que, em princípio, pode tornar-se consciente (princípio de conexão). Na percepção inconsciente, o princípio de conexão também deve estar presente se queremos oferecer uma explicação coerente dos processos perceptuais inconscientes. Portanto, toda a discussão de Searle em torno do inconsciente e, consequentemente, da percepção inconsciente é exclusivamente dependente de seu modelo explicativo de uma teoria intencional da consciência e da percepção.

Palavras-chave:

percepção, intencionalidade, consciência, inconsciente, percepção inconsciente.

in the cognitive sciences are actually neurobiological processes interpreted as being intentional and therefore often taken to be mental states. In Searle's view, an unconscious mental state is one that in principle can become conscious (connection principle). In unconscious perception, the connecting principle must also be present if we are to offer a coherent explanation of unconscious perceptual processes. Therefore, all of Searle's discussion of the unconscious and, consequently, of unconscious perception, is exclusively dependent on his explanatory model of an intentional theory of consciousness and perception.

Keywords:

perception, intentionality, consciousness, unconscious, unconscious percep-

Introdução

As considerações de Searle sobre a percepção inconsciente tornaram-se notórias apenas com a publicação de *Seeing Things As They Are* (2015). Todavia, essas considerações compõem o menor capítulo de seu livro, sendo, portanto, algo muito discreto em sua produção filosófica e, consequentemente, em sua teoria da percepção. Até então, todas as descrições dadas por Searle (1983; 2004; 2012) acerca da percepção eram apenas no horizonte da consciência e da intencionalidade. Em sua teoria da percepção, a intencionalidade desempenha um papel muito importante, sendo ela, uma noção extremamente cara ao seu pensamento filosófico como um todo. No que diz respeito ao tema da percepção

inconsciente, há em Searle (1991; 1992; 2004) uma discussão muito mais abrangente, que é a discussão sobre o inconsciente.

Nesse sentido, podemos considerar a discussão sobre o tema da percepção inconsciente algo menor ou um subproduto de uma discussão maior sobre o inconsciente. Quando comparamos, em seus escritos, o que já foi publicado sobre o inconsciente e o que já foi publicado sobre percepção inconsciente, é minimamente intrigante o porquê de o tema da percepção inconsciente ter sido quase completamente ignorado pelo estudioso ao longo das décadas.

No que concerne à literatura em filosofia da mente/percepção, ao menos desde a década de 60, os filósofos analíticos já falavam em percepção inconsciente, um exemplo notório é o clássico *A Materialist Theory of the Mind* (1968), de David Armstrong. Então, por que Searle se esquivou tanto deste tema? A resposta é muito simples: não há espaço na teoria de Searle para uma explicação dos fenômenos perceptuais inconscientes, uma vez que sua teoria funciona no horizonte da consciência e da intencionalidade. O que Searle faz é apresentar uma série de críticas à noção de inconsciente, que, em sua visão, é confusa e incoerente.

Portanto, a partir dessas questões, pretendo apresentar como Searle compreende o fenômeno da percepção inconsciente e como esse tema está conectado com um tema maior, que é o tema do inconsciente. Para este propósito, inicialmente apresentarei o que Searle entende por percepção e por inconsciente, para só depois avançar no que Searle está considerando como processos perceptuais inconscientes.

1. Percepção

A teoria da percepção de Searle possui dois modelos explicativos distintos, mas que não são excludentes. Assim, é possível notar uma continuidade e amadurecimento a partir do seu primeiro modelo até o atual. O primeiro modelo explicativo da percepção é oriundo dos anos oitenta. Costumo chamá-lo de *proto* teoria da percepção, localizada no capítulo dois da obra *Intencionalidade* (1983). O outro, um modelo acabado e pormenorizado, onde muitos dos *insights* centrais já estavam dados desde os anos oitenta, está presente em *Seeing Things As They Are* (2015). Trata-se de uma obra exclusivamente dedicada à filosofia da percepção. Comparado com o primeiro modelo, este segundo possui revisões e abrangências conceituais que não existiam no capítulo dois de *Intencionalidade* (1983).

Nas palavras de Searle (2015, p. 3), em *Intencionalidade*, "haviam algumas imprecisões e incompletude em minha explicação".

Agora em sua teoria atual, Searle precisa contrastá-la com outras visões e explicar por que a sua concepção seria melhor. Para isso, ele lança mão de uma noção que chamou de *Bad Argument*. O *Bad Argument* seria uma visão equivocada da percepção que, segundo Searle (2015), foi defendida pela maior parte dos filósofos ao longo dos séculos. O *Bad Argument*, portanto, encerra a ideia de que "nós nunca percebemos diretamente objetos e estados de coisas no mundo, mas percebemos apenas diretamente nossas experiências subjetivas" (Searle, 2015, p. 11). Poderíamos afirmar que o *Bad Argument*, tal como é descrito por Searle, já havia sido antecipado na própria modernidade filosófica por Thomas Reid quando este afirmou que "Todos os filósofos, de Platão ao Sr. Hume, concordam que não percebemos os objetos externos imediatamente e que o objeto imediato da percepção deve ser alguma imagem presente na mente" (Reid, 1785, p. 86). Segundo Searle (2015), qualquer argumento que pretenda tratar a experiência perceptual como objeto da experiência real ou possível seria uma forma de *Bad Argument*. Para entendermos melhor esse ponto é preciso estar atento à distinção que Searle faz entre conteúdo e objeto.

De acordo com Searle (2015, p. 37), "duas experiências perceptuais podem ter conteúdos de mesmo tipo, mas uma tem objeto enquanto que a outra não"). O conteúdo intencional consiste na experiência subjetiva que ocorre dentro de nossas cabeças; em contrapartida, o objeto intencional revela-se como qualquer objeto do mundo externo capaz de causar em nós experiências subjetivas perceptuais. Essa noção foi influenciada pela teoria causal da percepção de Paul Grice[2]. Entretanto, em Searle, a causalidade assume formas intencionais, pois considera o sujeito da percepção consciente e ativo em todo o processo.

Em sua explicação da percepção, Searle (1983; 2004; 2012; 2015) defende um realismo direto acerca do tipo de acesso que temos aos objetos do mundo externo. O realismo direto afirma que as coisas que percebemos no mundo, percebemo-las da maneira como elas realmente são. Como o próprio termo ("realismo direto") aponta, temos um acesso direto, portanto, não mediado ao mundo objetivo.

O realismo direto é também conhecido por realismo ingênuo (*naïve*). Entretanto, Searle (2015) faz questão de frisar que o realismo dele não pode ser ingênuo[3], reservando

2 GRICE, H. Paul. The Causal Theory of Perception. Proceedings of the Aristotelian Society. Supp. vol. xxxv, 1961. p. 121-53.

3 Não obstante, em *Intencionalidade* (1983), Searle não parece ter esse cuidado de separar o seu realismo

este termo para os disjuntivistas, que também endossam um realismo direto (ingênuo). Não é meu propósito explicar os pormenores do que seja o disjuntivismo em percepção, até porque esta proposta possui alguns matizes[4] que não teríamos como dar conta aqui. Todavia, retornarei ao disjuntivismo para mostrar uma questão que faz toda diferença entre o realismo direto de Searle e o realismo direto (ingênuo) que eles defendem.

Além do disjuntivismo, uma postura em filosofia da percepção bastante atraente e dominante no âmbito das discussões filosóficas é o realismo indireto, também conhecido como representacionalismo. Esta postura estabelece que todas as nossas percepções de objetos físicos no mundo natural são sempre percebidas indiretamente. Dessa forma, o que percebemos de maneira imediata e direta são apenas objetos internos não físicos como impressões, ideias, *sense-data*, e assim por diante[5]. Agora que sabemos que Searle defende um realismo direto em sua explicação perceptual, vamos a um outro aspecto importante a ser levado em conta, a saber, para Searle (1983; 2004; 2012; 2015), percepção pressupõe consciência e intencionalidade.

Definir consciência pode ser algo extremamente complicado dada a miríade de concepções que este conceito pode abarcar. Contudo, isso não nos impede de caracterizá-la de maneira sucinta para o nosso objetivo em questão, que é oferecer uma breve imagem acerca da teoria da percepção de Searle. Além de um fenômeno biológico causado por processos neuronais em nosso cérebro, Searle (2015, p. 46) afirma que "consciência (*consciousness*) consiste em todos os nossos estados (processos, eventos, etc.) de sentimento, senciência ou de consciência (*awareness*)". Trata-se de um fenômeno que abrange muitas esferas de nossa vida psíquica.

Em nossa vida prática, a consciência está presente em nós desde o primeiro momento que despertamos de nosso sono e só encerra quando voltamos a dormir. Todavia, como bem observa Searle (2015, p. 47), "sonhos são uma forma de consciência, embora

direto de um realismo ingênuo ao declarar que "Não é meu objetivo neste capítulo entrar nas disputas tradicionais concernentes à filosofia da percepção; no entanto, a tese que estou defendendo sobre a intencionalidade da experiência visual talvez seja mais clara se nos desviarmos um momento para contrastar essa visão realista ingênua (*naive realist view*) com seus grandes rivais históricos, a teoria representativa e o fenomenalismo" (Searle, 1983, p. 58).

4 Uma visão sucinta dos tipos de disjuntivismo está no capítulo 7 (*varieties of disjunctivism*) de SOTERIOU, Matthew. **Disjunctivism**. New York: Routledge, 2016. Para uma discussão mais abrangente, ver BYRNE, Alex & LOGUE, Heather (Eds.). **Disjunctivism: Contemporary Readings**. Massachusetts. The MIT Press, 2009.

5 Para uma descrição acurada do representacionalismo, ver o cap. 4 de MAUND, Barry. **Perception**. (Central Problems of Philosophy) Chesham: Acumen, 2003.

bem diferente da consciência desperta". Além do mais, para Searle (1992; 2004; 2015), é possível traçar algumas características da consciência como, por exemplo, qualitatividade, subjetividade, unidade e irredutibilidade.

Com a intencionalidade não é muito diferente, ela também é considerada por Searle (1984; 2015) como parte de nossa biologia, possuindo, portanto, seu lugar de direito na natureza. Em termos definicionais, Searle (1983, p. 1) caracteriza a intencionalidade como "aquela propriedade de muitos estados e eventos mentais pela qual estes são dirigidos para, ou acerca de objetos e estados de coisas no mundo". No grau de dificuldade, de acordo com Searle (2004), o problema da intencionalidade só estaria abaixo do problema da consciência, sendo este um subproduto ou um espelho do problema da consciência.

Desde Brentano (1876), a intencionalidade vem sendo majoritariamente caracterizada como a marca de nossos estados mentais, que, por seu turno, se distinguem ontologicamente dos estados de coisas no mundo. A intencionalidade pode ser sobre estados de coisas, mas não é ela mesma os objetos e estados de coisas em si. Por essa razão, apenas estados mentais exibem intencionalidade. Assim, estados mentais como crenças, desejos, esperanças, medos, sonhos etc. são sempre acerca de algo, isto é, possuem um conteúdo intencional. Isso faz com que a intencionalidade esteja intimamente relacionada com a forma como nós representamos o mundo.

Em sua explicação intencionalista da percepção, Searle (1983; 2004; 2012; 2015) desenvolve toda uma linha de argumentação segundo a qual consciência e intencionalidade são características imprescindíveis dos fenômenos perceptuais. Para Searle, uma explicação da percepção que não leve em conta a consciência e a intencionalidade não deve ser levada a sério. Sendo a consciência o pano de fundo dessas questões a nível de fenômenos e a neurobiologia o nível mais baixo de condições de possibilidade para os nossos estados mentais, Searle (2015) considera que há uma relação intrínseca entre intencionalidade e percepção. Dessa forma, fenômenos perceptuais não podem ser concebidos senão como dotados de intencionalidade. Isso faz com que a explicação dos fenômenos perceptuais em Searle seja intrinsecamente intencional, além, é claro, de que o tipo de acesso que temos aos objetos do mundo natural seja direto (realismo).

Percepção pertence à classe dos estados mentais e, como vimos, estados mentais exibem intencionalidade, sendo muitas vezes acerca de objetos e estados de coisas no mundo. Há três casos paradigmáticos para explicar nossas experiências visuais: percepção verídica, ilusão e alucinação. A teoria da percepção de Searle reponde a cada um desses três

casos. Quando nossa experiência visual é causada por objetos externos, chamamos essa experiência visual de percepção verídica (*the good case*). Na descrição de Searle (2015), há um objeto intencional no mundo natural que causou minha experiência visual subjetiva, a essa experiência Searle atribui um conteúdo intencional. Quando tenho uma ilusão perceptiva, ou seja, se estou a ver algo que acredito ser um X, mas que na verdade é um Y, o que ocorreu foi uma falha no modo como minha percepção capturou o objeto em questão. Às vezes, temos consciência da ilusão, como no exemplo do bastão que assume uma aparência curva quando imerso na água. Na ilusão, ainda existe um objeto externo que causou minha percepção, embora o conteúdo da percepção seja ilusório. Em contrapartida, na alucinação (*the bad case*), o que existe é apenas a experiência visual (conteúdo intencional), mas sem um objeto do mundo externo responsável por causar a experiência. Na alucinação, o sujeito tem uma experiência visual de algo que não corresponde nem de forma verídica, nem de forma ilusória ao mundo.

Apesar de não haver um objeto (causal) da experiência perceptual, a experiência visual de caráter alucinatório compartilha de um elemento comum com os outros casos. Trata-se do elemento fenomenal ou conteúdo intencional. Searle (2015) não faz uma distinção ontológica entre o conteúdo intencional de uma percepção verídica e o conteúdo intencional de uma alucinação. Do ponto de vista fenomenológico, Searle (2015, p. 170) decide "considerar que casos de percepção verídica e casos de alucinação correspondentes são exatamente os mesmos". Em outras palavras, há em Searle uma defesa do elemento comum entre o caso verídico e o caso alucinatório.

Mais acima mencionei que os disjuntivistas defendiam um realismo ingênuo acerca da percepção. O que essencialmente distingue o realismo direto de Searle do realismo direto (ingênuo) dos disjuntivistas é a defesa da tese do elemento comum. Em sua defesa do realismo direto (ingênuo), os disjuntivistas não endossam a tese do elemento comum entre o conteúdo intencional de uma percepção verídica e o conteúdo intencional de uma alucinação. Este ponto, que distingue Searle dos disjuntivistas, revela um elo frágil em sua defesa do realismo direto, que, numa certa medida, pode ser interpretado como uma teoria representacional da percepção.

Assim, podemos considerar três características principais na teoria da percepção de Searle: a primeira, é sua defesa do realismo direto; a segunda, a tese de que percepção envolve intencionalidade; e terceira, a tese de que existe um elemento comum entre os casos verídicos e alucinatórios. Não pretendo desenvolver aqui as implicações de cada uma

dessas características, seus problemas e inconsistências[6]. Ao contrário, meu propósito foi apenas prover uma imagem razoável e sucinta de sua teoria da percepção para avançarmos em nosso objetivo.

2. O Inconsciente

As considerações de Searle sobre a problemática do inconsciente surgem em contraposição à falta de interesse dos filósofos e cientistas cognitivos em prover um programa de estudos e investigação adequados para a consciência. Em *Consciousness, Unconsciousness and Intentionality* (1991), um texto que antecipa e serve de base para algumas das reflexões sobre o tema do inconsciente na *Redescoberta da Mente* (1992), Searle (1991, p. 45) afirma que "uma das coisas mais surpreendentes no último meio século na filosofia analítica da mente é a escassez de trabalhos sérios sobre a natureza da consciência". Dada a dificuldade que é prover uma descrição adequada da consciência enquanto objeto de investigação empírica, filósofos e cientistas se sentem mais inclinados a investigar processos cognitivos subjacentes ao nível da consciência, isto é, processos que não dependem da consciência para serem observados. Esses processos foram chamados de processos inconscientes e o objetivo de Searle (1991; 1992; 2004) é mostrar o quão problemático pode ser a noção de inconsciente.

Em *Mind* (2004), Searle delineia quatro tipos de inconsciente explorando sua natureza e modos de existência para então determinar quais, em sua visão, são problemáticos e quais não são. Ele começa ingenuamente perguntando se o inconsciente mental realmente existe, isto é, se pode existir um estado que é literalmente mental e ao mesmo tempo inconsciente. Para Searle (2004), esses estados não seriam subjetivos e tampouco qualitativos pelo fato de não serem parte de um campo unificado de consciência. Com o intuito de fortalecer a linha de argumentação que pretende desenvolver, Searle (2004, p. 238) recorre à clássica noção cartesiana de consciência:

> Para Descartes, é óbvia a resposta à pergunta: *Existem estados mentais inconscientes?* A ideia de um estado mental inconsciente é uma autocontradição. A mente é definida por Descartes como *res cogitans* (coisa pensante) e "pensar"

6 Exploro essas questões de modo aprofundado em ARAUJO, J. P. M. Representacionalismo e realismo direto na teoria da percepção de John. R. Searle. In: SOUZA, Marcus José Alves de & LIMA FILHO, Maxwell Morais de (Orgs.). **Escritos de Filosofia IV**: Linguagem e Cognição. Porto Alegre. Editora Fi, 2020.

> para Descartes é apenas outro nome para a consciência. Assim, a ideia de um estado mental inconsciente seria a ideia de uma consciência inconsciente, uma simples autocontradição.

A noção cartesiana de consciência, bem como sua conexão com o pensar, é algo que perdura até os nossos dias atuais. Todavia, apesar de seu apelo intuitivo, ela passou a ser sistematicamente questionada a partir do advento da psicanálise. Na abordagem que Searle (2004, p. 238) faz do inconsciente, o modelo explicativo de Freud seria apenas uma tentativa de descrever um tipo de estado mental que em tese seria desprovido de consciência, afirmando que "o problema com esta imagem é que é muito difícil fazer algum sentido nela". Vamos supor que você tenha em mente o seguinte pensamento (consciente): "Wittgenstein só apreciava filmes de faroeste". Vamos agora repetir o procedimento só que subtraindo a consciência, isto é, o mesmo pensamento só que inconscientemente. Como seria tal coisa possível?

Para entendermos esse ponto, voltemos aos quatro tipos de "inconsciente" delineados por Searle (2004), que resumidamente são definidos como: (1) pré-consciente; (2) inconsciente reprimido; (3) inconsciente profundo e (4) não-consciente. O primeiro deles é um caso fenomênico e, portanto, não problemático. O próprio termo vem da psicanálise, seria aquilo que Freud[7] (1912) chamou de pré-consciente. Mas em termos searleanos, trata-se daquela dimensão disposicional que nossas crenças possuem enquanto possibilidade de se tornarem conscientes. Por exemplo, a minha crença de que "Wittgenstein só apreciava filmes de faroeste" é uma crença que não precisa estar presente em minha consciência o tempo todo. Da mesma forma, minha crença de que "Boa Vista é a capital de Roraima" ou que "Damurida é uma comida apimentada", até poucos segundos atrás, não fazia parte do meu campo consciente. Aqui é revelado o caráter intencional de nossas crenças porque são sempre acerca de algo, muito embora, elas não precisam necessariamente serem conscientes. Dado o conjunto infinito de nossas crenças, elas estão muitas vezes em *stand by*, de modo que podem ser acessadas indefinidamente por nós. Portanto, aqui temos um caso que Searle (2004) considera não problemático.

7 Em seu artigo "*Algumas observações sobre o conceito de inconsciente na psicanálise*", Freud apresenta três concepções para o inconsciente. Seguindo a tríplice noção Inconsciente/Pré-consciente/Consciente Freud delineia: (1) concepção descritiva do inconsciente, isto é, o inconsciente em estado latente embora capaz de consciência; (2) concepção dinâmica do inconsciente, ou seja, recalque ou repressão em relação à certos conteúdos indesejáveis no âmbito da consciência e (3) concepção sistemática do inconsciente, a saber, o inconsciente enquanto algo que possui suas próprias leis de funcionamento e que diferem radicalmente da atividade da consciência. Disso podemos notar o quão problemático pode ser o conceito de inconsciente em Freud se o tomarmos unilateralmente.

O segundo caso (o inconsciente reprimido), ao contrário do primeiro, Searle considera problemático. O *insight* basilar sobre esse tipo de inconsciente é a ideia segundo a qual um indivíduo tem um ou mais estados mentais que interferem causalmente em seu comportamento. O indivíduo neste caso desconhece completamente a existência deste estado mental (inconsciente), frequentemente negando-o. Trata-se da genuína repressão descrita por Freud. De acordo com Searle (2004, p. 240), "esses são casos em que o estado mental inconsciente funciona causalmente, mesmo quando inconsciente". Para ilustrar esse ponto, um bom exemplo prático é quando alguém está sob efeito da hipnose. Na hipnose, o indivíduo age por um motivo que não tem consciência e que, em alguns casos (em condições normais), esta ação seria contrária ao que ele pensa e acredita.

O terceiro caso (inconsciente profundo), não opera mais no terreno da psicanálise, mas, sim, no das ciências cognitivas. Searle considera esse caso tão problemático quanto o segundo caso acima descrito. O indivíduo não pode trazer o estado mental à consciência porque esse tipo de processo (inconsciente) "não é o tipo de coisa que pode formar o conteúdo de um estado intencional consciente" (Searle, 2004, p. 241). Em seu texto, Searle (2004) usa como exemplo o modelo explicativo das ciências cognitivas. Via de regra, é afirmado que uma criança quando aprende uma língua ela o faz aplicando de maneira "inconsciente" regras computacionais de uma gramática universal. O mesmo ocorre para muitas de nossas percepções visuais, que são operações computacionais "inconscientes" de *inputs* em nossa retina, um processo subjacente à nossa consciência[8].

> Em ambos os casos, tanto na aquisição da linguagem quanto na formação de percepções, as regras computacionais não são o tipo de coisa que poderiam ser pensadas conscientemente. Em última análise, elas se reduzem inteiramente a sequências massivas de zeros e uns, e tudo o que a criança pode fazer quando pensa, ela não pode pensar em zeros e uns, e de fato os zeros e uns são apenas uma maneira de falar. Os zeros e uns existem na mente do observador e formam um modo de descrição do que se passa inconscientemente na mente da criança (Searle, 2004, p. 241).

Por fim, o quarto caso de inconsciente (o não consciente) descrito por Searle é algo que pertence à nossa dimensão neurobiológica. Este é mais fácil e intuitivo de compreen-

8 Há, na filosofia da mente e nas ciências cognitivas, toda uma literatura e discussão em torno desses estados ou processos subjacentes à nossa consciência. Eles são conhecidos pelo termo "estados subdoxásicos" e encerram processos que ocorrem em nossa cognição, dos quais nós não temos a mínima consciência. Um texto paradigmático sobre o assunto é o artigo de Stephen Stich (1978), "*Beliefs and Subdoxastic States*".

der e é considerado por Searle, assim como o primeiro caso, não problemático. Grosso modo, poderíamos afirmar que o que se passa em minha retina e no meu lobo occipital não é acessado diretamente e conscientemente por mim. O que tenho acesso são apenas os conteúdos fenomênicos e intencionais de minha percepção visual. Trata-se de processos que para ocorrerem não precisam de um indivíduo consciente. Searle (2004) usa como exemplo a nossa medula; mesmo se estivéssemos inconscientes, ela continuaria controlando nossa respiração. Em suma, nós não temos acesso a esses processos da mesma forma que acessamos os conteúdos da nossa consciência, pois não se tratam de fenômenos mentais.

Podemos resumir esses quatro casos de inconsciente descritos por Searle na categoria dos que são considerados fenômenos mentais e dos que não o são. O primeiro e segundo casos residem no âmbito fenomênico de nossos estados mentais, enquanto que o terceiro e quarto escapam à categoria de fenômenos mentais. A estratégia de Searle (2004, p. 242) consiste em mostrar que "a forma de compreender os casos reprimidos segue o modelo do primeiro, o pré-consciente; e a maneira de entender o terceiro, os casos inconscientes profundos, segue o modelo do quarto, os casos não conscientes".

No que concerne aos casos problemáticos de inconsciente, Searle desenvolve alguns questionamentos. Para os casos de tipo (2), ele faz a seguinte pergunta: "Como pode um estado mental reprimido existir e funcionar como um estado mental quando está completamente inconsciente?" (Searle, 2004, p. 243). Como vimos, para os casos de tipo (1), não há problema, uma vez que é perfeitamente comum uma pessoa estar inconsciente e ainda assim possuir uma gama de crenças em *stand by*. Em outras palavras, eu não preciso estar consciente de todas as minhas crenças para dizer que as tenho. Mas com os casos de tipo (2) isso não é tão simples, pois o estado inconsciente (reprimido) teria poderes causais no comportamento humano. Segundo Searle (2004, p. 243), "parece-me que quando atribuímos esses estados mentais inconscientes a um agente, estamos atribuindo características neurobiológicas capazes de causar consciência".

Para resolver o problema dos casos de tipo (2), Searle (2004, p. 245) propõe que "o mesmo tipo de processo neurobiológico que pode causar um estado consciente também pode causar um comportamento apropriado para ter esse estado consciente". Dito de outro modo, os casos de tipo (1) e (2) estão intimamente relacionados com a nossa neurobiologia, ou seja, com o funcionamento do nosso cérebro. O cérebro é condição *sine qua non* para que possa haver tais estados. Esta é uma tese defendida por Searle que ficou conhecida como naturalismo biológico e ela foi primariamente proposta em 1992 com a publicação da obra *A Redescoberta da Mente*. Lá vemos Searle (1992, p. 1) afirmar que:

> Os fenômenos mentais são causados por processos neurofisiológicos no cérebro e são eles próprios características do cérebro. (...) Eventos e processos mentais fazem parte de nossa história biológica natural tanto quanto a digestão, a mitose, a meiose ou a secreção enzimática.

Nesse sentido, é esperado que Searle proponha uma explicação naturalista para o inconsciente, erradicando qualquer possibilidade de tratar o inconsciente como uma entidade metafísica que não possua nenhuma relação com o mundo natural. Ao menos para os dois primeiros casos de inconsciente, nenhum mistério paira sobre sua explicação.

No que concerne aos casos de tipo (3), Searle (2004) é resoluto ao afirmar que não existem tais casos. Sua tese sobre esses casos é simples e direta. Para Searle (2004), aquilo que ele está chamando de inconsciente profundo nas ciências cognitivas não passa de processos neurobiológicos interpretados como sendo intencionais e, portanto, tomados muitas vezes como estados mentais. Dentro da tese do naturalismo biológico de Searle, existem processos neurobiológicos capazes de causar na consciência uma gama de estados mentais. Por outro lado, esses processos descritos nos casos de tipo (3) se comportam como se fossem intencionais. Porém, "na medida em que o estado mental não é nem mesmo o tipo de coisa que poderia se tornar o conteúdo de um estado consciente, não é um estado mental genuíno" (Searle, 2004, p. 246).

Diferentemente dos casos de tipo (1) e (2), os casos de tipo (3) jamais poderão chegar à consciência como um conteúdo intencional da mente. Na visão de Searle, um estado mental inconsciente é aquele que em princípio pode tornar-se consciente. Ele chamou isso de "Princípio de Conexão" (*Connection Principle*), ou seja, "a noção de inconsciente está logicamente conectada à noção de consciência" (Searle, 2004, p. 246). Vimos no início do texto que consciência e intencionalidade são noções bem caras ao pensamento filosófico de Searle. Ademais, fenômenos intencionais possuem formas aspectuais, na medida em que permitem ver ou experienciar um fenômeno de uma determinada perspectiva. Se a intencionalidade é aquela característica de nossos estados mentais de serem sempre sobre alguma coisa, a forma aspectual é a idiossincrasia do modo em que a própria coisa se apresenta na consciência. O que Searle conclui é que nos casos de tipo (3) não há forma aspectual e a razão já sabemos, porque os casos de tipo (3) não são autênticos estados mentais.

Com isso, Searle assimila os casos de tipo (3) aos de tipo (4), uma vez que eles são desprovidos de qualquer traço que possa ser identificado como um autêntico estado mental. Trata-se de um grande equívoco afirmar que estados cerebrais seriam estados in-

conscientes ocorrendo abaixo do radar da consciência. Nossa neurobiologia é totalmente desprovida de forma aspectual. Searle usa como exemplo um homem que tem o desejo de beber água, mas que não tem o desejo de beber H_2O. A razão é muito simples, ele não sabe o que é H_2O:

> Mas o comportamento externo será exatamente o mesmo nos dois casos: o caso de desejar água e o caso de desejar H_2O. Em cada caso, ele procurará beber o mesmo tipo de bebida. Mas os dois desejos são diferentes. Como essa diferença pode ser captada no nível da neurofisiologia? A neurofisiologia, descrita em termos de força sináptica e potenciais de ação, nada conhece sobre a forma aspectual (Searle, 2004, p. 247).

Dois outros aspectos que Searle leva em consideração em sua discussão sobre o tema do inconsciente são: (A) razões para ação e (B) seguir regras. Não pretendo me deter nesses tópicos, mas para fins de uma visão resumida farei breves considerações. No que concerne às razões para agir, à explicação de nossas ações a partir de razões, Searle adota uma postura davidsoniana de senso comum. Davidson (1963) afirma que uma razão justifica ou racionaliza uma ação se, e somente se, a razão nos leva perceber que o agente viu em sua ação alguma característica, consequência ou aspecto da ação que o agente desejava, prezava, considerava benéfica, obrigatória etc.

Searle (2004, p. 250) considera a noção de razão um elemento chave para a explicação da ação humana, na qual "o conteúdo da explicação deve corresponder ao conteúdo da mente do agente cujo o comportamento está sendo explicado". Mas como funciona isso no âmbito do inconsciente? Assim, como razões explicam nossas ações, a postulação do inconsciente também tem por objetivo explicar nossas ações, ou ao menos uma parcela de nossas ações, que normalmente uma explicação racional não daria conta.

Como observa Searle (2004, p. 250), "a razão pela qual dizemos que as pessoas têm motivações inconscientes é que não encontramos outra maneira de explicar algumas formas de seu comportamento". Todavia, Searle chama atenção para o fato de que as razões que justificam nossas ações não são simples razões, elas o são apenas na superfície. Na verdade, elas, muitas vezes, são um conjunto de razões complexas que constituem aquilo que Searle (2000) desde *Rationality in Action* chamou de razão total (*total reason*).

> Razões são sempre proposicionais na forma e algo é uma razão apenas se for parte de uma razão total. O ponto chave para a discussão do inconsciente é este.

> Existem algumas formas de comportamento humano que só fazem sentido se postularmos uma razão para a ação da qual o próprio agente não tem consciência (Searle, 2004, p. 251-2).

Mais uma vez, postular uma razão na qual o agente não tem consciência deve ser algo que a princípio poderia tornar-se consciente para o agente. Desse modo, o inconsciente só tem sentido de ser postulado se ele está conectado em alguma medida com a consciência.

Além das razões para ação, um outro aspecto considerado por Searle é o de seguir uma regra. Ao menos, desde Wittgenstein (1953), os filósofos discutem o que significa seguir uma regra. Em termos wittgensteinianos, seguir uma regra é uma instituição pública atravessada pela dimensão intersubjetiva de nossa linguagem. É a partir daí que Wittgenstein (1953) afirma que não pode existir uma linguagem privada, porque falar uma língua é seguir regras e ninguém pode seguir uma regra privadamente pelo simples fato de que acreditar seguir uma regra não é seguir uma regra. Os critérios de correção de regras são sempre realizados na esfera pública de nossas relações sociais, e não a partir de estados subjetivos isolados desconectados de todo o resto. Searle (2004) considera o seguir regras uma subcategoria especial de razões para ação e lança mão desse *background* para discutir se podemos seguir regras inconscientemente.

Dessa forma, Searle (2004, p. 253-5) elenca seis características do comportamento quando seguimos regras. A primeira delas (1) afirma que o conteúdo da regra deve funcionar causalmente na produção do comportamento; regras condicionam nosso comportamento. A segunda (2) depende da primeira e afirma que as regras têm propriedades lógicas que são comuns aos estados intencionais e aos atos de fala diretivos de quando obedecemos a uma ordem. A terceira (3) é corolário de (1) e (2), pois toda regra deve ter um conteúdo intencional que revela uma forma aspectual concernente à própria regra. A quarta (4) preza pelo voluntarismo de seguir uma regra, ou seja, para que uma regra seja capaz de guiar o comportamento, ela tem de ser algo que o agente possa seguir voluntariamente e isso implica também em deixar de seguir a própria regra. A quinta (5) diz que regras estão sempre sujeitas a interpretações, isto é, podemos seguir a mesma regra, mas a partir de outras motivações ou justificações. Por fim, (6) afirma que seguir regras só pode ocorrer em tempo real (aqui e agora), no qual são determinadas as condições de satisfação, ou seja, se estamos ou não performando a regra corretamente.

De acordo com Searle (2004), essas seis características são elencadas apenas ao nível da consciência e quando postulamos o seguimento de regras no nível inconsciente

fica difícil manter algumas das características descritas acima. O alvo de Searle (2004, p. 256) quando postula o seguir regras a nível inconsciente são as ciências cognitivas, para ele "muitas postulações de seguimento inconsciente de regras, como nas explicações da ciência cognitiva da percepção visual e aquisição da linguagem, não atendem a essas condições".

Searle (2004, p. 256) conclui suas observações afirmando que "A noção de inconsciente é uma das concepções mais confusas e mal pensadas da vida intelectual moderna". Para ele, precisamos de uma noção coerente do inconsciente, esta noção deve se ajustar ao nosso conhecimento da realidade incluindo o que sabemos sobre o cérebro. Uma chave para essa compreensão mais coerente é o princípio de conexão, que, como vimos, conecta a noção de inconsciente à noção de consciência. Apesar de demonstrar não estar inteiramente satisfeito com sua conclusão e de não conseguir pensar numa explicação alternativa melhor, Searle (2004, p. 257) insiste que:

> Dizer de um agente que ele tem tal e tal estado intencional inconsciente, e que esse estado está funcionando ativamente na causa de seu comportamento, é dizer que ele tem um estado cerebral que é capaz de causar esse estado de forma consciente, mesmo que em um caso particular possa ser incapaz de causá-lo de forma consciente por causa de dano cerebral, repressão, etc.

É notória a problemática em torno da ontologia do inconsciente em Searle. Como já foi afirmado ao longo do texto, todo o seu esforço consiste em tentar obter uma explicação do inconsciente que seja compatível ou consistente com uma visão naturalista do mundo físico e com o papel dos estados mentais no mundo.

3. Percepção Inconsciente

Desde *Intencionalidade* (1983), Searle se dedicou apenas a prover uma explicação da percepção consciente, ignorando completamente as discussões em torno dos fenômenos perceptuais inconscientes. O fato é que em sua teoria da intencionalidade perceptual não há espaço para uma explicação da percepção inconsciente, uma vez que todo o seu foco é na explicação da percepção consciente. Mas o que significa então o tópico da percepção inconsciente em Searle? Da mesma forma que a discussão do inconsciente, a questão da percepção inconsciente recebe uma conotação crítica da parte de Searle. A

diferença agora é que Searle irá focar suas críticas no modo como as ciências cognitivas descrevem o fenômeno em questão.

Em sua obra *Seeing things as they are*, Searle (2015) desenvolve toda uma teoria intencional da percepção focando apenas na percepção consciente e dedicando apenas um pequeno capítulo para tratar da percepção inconsciente. Neste capítulo, ele começa explanando resumidamente o que ela já havia feito em *A Redescoberta da Mente* (1992) e em *Mind* (2004) sobre o inconsciente para, a partir de então, discutir os casos que as ciências cognitivas estão chamando de percepção inconsciente. Vale ressaltar que aquilo que Searle está chamando de ciências cognitivas é apenas um recorte de alguns autores que trabalharam essas questões. Na atualidade, há um vívido debate sobre a percepção inconsciente no âmbito filosófico e na psicologia empírica que Searle ignorou ou desconhece completamente[9].

Ao tratar do tema da percepção inconsciente, Searle (2015) reproduz uma ideia que já nos é familiar desde a publicação de *A redescoberta da mente* (1992). Trata-se da ideia de que sempre houve "uma suspeita sobre a consciência como um nível genuíno de compreensão do comportamento humano e da cognição humana" (Searle, 2015, p. 208). Essa suspeita é traduzida pela visão de que a consciência não desempenha um papel importante em nossa cognição e comportamento, que questões que envolvem percepção ou até mesmo ação voluntária são no fundo processos inconscientes. De acordo com Searle (2015), essa visão foi amplamente implementada por aquilo que ele chamou de "ideologia da metáfora do computador", que tinha como pano de fundo resultados experimentais. Com o intuito de lançar luz sobre o tema, Searle (2015) analisa três casos experimentais paradigmáticos sobre processos perceptuais inconscientes, a saber: (1) visão cega, (2) potencial de ação e (3) reflexos. Veremos cada um desses casos.

No que consiste a visão cega? O modo como conscientemente percebemos as coisas visualmente depende de uma área de nosso cérebro chamada córtex visual primário. Qualquer dano severo nessa área ocasiona aquilo que os neurocientistas chamam de cegueira cortical. O fenômeno da visão cega ocorre quando o sujeito com algum tipo de dano nessa área, que compromete parte significativa de seu campo visual, responde de algum modo a um estímulo mostrado em seu ponto cego que numa situação corriqueira o sujeito não conseguiria enxergar conscientemente. Searle (2015) introduz essa discussão citando o trabalho de Weiskrantz[10], um dos pioneiros nesse campo de pesquisa.

---

9 Para uma defesa atual dos processos perceptivos inconscientes, ver: BLOCK, Ned. **The Anna Karenina Principle and Skepticism about Unconscious Perception**, 2015. Para uma linha argumentativa mais crítica desses processos, ver: PHILIPS, Ian. **Unconscious Perception Reconsidered**, 2018.

10 WEISKRANTZ, Lawrence. Blindsight: **A Case Study Spanning 35 Years and New Developments**. Ox-

A visão dos pacientes acometidos por esse tipo de dano possui uma significativa parte da extensão de seus campos visuais reduzidos. Esses pacientes são literalmente cegos de uma parte de seus campos visuais, eles não veem manchas, escuridão ou qualquer coisa do tipo. Literalmente, eles nada veem, pois não existe essa parte de seus campos visuais. Um detalhe importante é que a retina, bem como o nervo óptico, não é comprometido, apenas a área cortical da visão.

No experimento de Weiskrantz (2009), ele pede para o paciente fixar seus olhos no centro das linhas que se cruzam em uma tela de computador. Feito isso, o próximo passo é fazer piscar rapidamente um X ou um O na região do campo visual acometido pela cegueira. Dada a velocidade do estímulo, o paciente não consegue mover os olhos, mas ele pode relatar o que está acontecendo. Em seus relatos, o paciente afirma que lhe pareceu ver um X ou um O numa região que até então não percebia. Com o tempo, o paciente vai ficando mais habilidoso, acertando em mais de 90% das vezes que é submetido ao teste.

Na concepção de Searle (2015, p. 209), "[h]á claramente algo de uma forma intencional de informação sendo recebida na parte do campo visual do paciente onde ele é cego". Em sua interpretação do experimento, Searle (2015, p. 209) considera que "esta experiência mostra claramente que existem formas de percepção intencional que não são conscientes". Aqui temos aquilo que Searle chamou de princípio de conexão, da mesma forma que na discussão sobre o inconsciente, nesses processos perceptuais, a percepção inconsciente se revela como aquilo que em princípio poderia tornar-se parte da consciência.

Searle (2015) afirma que, para Weiskrantz (2009), o aspecto mais crucial de seu experimento é mostrar que não existe apenas uma via neuronal, mas várias vias neuronais no sistema visual e que nem todas são conscientes. O problema dessa discussão é que ela permanece na superfície. Searle a toma como algo autoevidente e, para isso, cita pesquisas posteriores ao experimento de Weiskrantz que apoiam esses dados[11]. Portanto, apesar de interessantes, Searle (2015, p. 214) considera os casos de visão cega algo muito específico e exemplos marginais de percepção, "ninguém pode dirigir um carro, ou mesmo escrever um livro ou assistir a um filme, usando apenas os recursos da visão cega".

O segundo caso experimental é o do potencial de ação (*readiness*). Esse caso é bastante famoso no cenário filosófico devido às implicações dos experimentos de Benjamin Libet[12], que terminaram contribuindo para o fortalecimento da ideia de que o livre-arbí-

---

ford: Oxford University Press, 2009.

11 MILNER, David; GOODALE, Mel. **The Visual Brain in Action**. Oxford: Oxford University Press, 2006.

12 Seus famosos artigos da década de 80 são: *Time of conscious intention to act in relation to onset of cerebral*

trio é uma ilusão e que, muitas de nossas escolhas que julgamos tomar conscientemente, já são inconscientemente iniciadas a nível neuronal. Ou seja, já são previamente determinadas pelo nosso cérebro. Mencionando não apenas Libet, mas também pesquisadores alemães[13] pioneiros na década de 70, Searle (2015, p. 210) afirmou que "os resultados de suas pesquisas pareciam mostrar que o início da ação era inconsciente, ou seja, a ação foi iniciada antes que o agente estivesse consciente do que estava fazendo".

O experimento consiste em monitorar um sujeito que foi instruído a realizar algumas ações simples, como esticar sua mão e apertar um botão. Ele também foi previamente instruído para que toda vez que apertasse o botão olhasse para um relógio e conferisse o exato momento em que ele decidiu apertar o botão, ou seja, o momento em que sua intenção de agir teve início. O resultado empírico do experimento mostrou que, durante o processo, houve um lapso de tempo (350 milissegundos) entre uma atividade cerebral motora do sujeito e sua consciência ao iniciar a ação. A partir desses dados, a interpretação majoritária do experimento encerra a ideia de que o nosso cérebro decide que vai apertar o botão antes de estarmos conscientes de nossa decisão. Tudo se passa como se fôssemos um piloto que tem a ilusão de estar pilotando uma aeronave, quando, na verdade, a aeronave está sendo conduzida pelo piloto automático e nós somos apenas os passageiros.

Nas considerações de Searle (2015), os desdobramentos dessa discussão revelam uma filosofia ruim, ancorada num projeto experimental também ruim; uma espécie de postura filosófica que toma a consciência como algo menos importante e que, portanto, não deve ser levada a sério na investigação empírica.

> Todos os tipos de pessoas que melhor deveriam saber, disseram que os experimentos de Libet refutam o livre arbítrio e mostram que nosso comportamento é de fato determinado. Talvez o livre-arbítrio seja falso, mas os experimentos de Libet não mostram nada disso. Um estudo recente sugere a possibilidade de que os resultados experimentais resultem da exigência de que os sujeitos olhem para o relógio. Talvez seja o relógio que produz o potencial de ação. Se você conduzir o mesmo experimento em que o sujeito decide não se mover, o mesmo "potencial de ação" acontece (Searle, 2015, p. 211).

---

activity (readiness-potential), *The unconscious initiation of a freely voluntary act* (1983), e *Unconscious cerebral initiative and the role of conscious will in voluntary action* (1985).

13 Ver: DEECKER, Lüder; GRÖZINGER, Berta & KORNHUBER, H. H. Voluntary finger movement in man: Cerebral potentials and theory. **Biological Cybernetics**, vol. 23, 1976, p. 99-119.

Sobre este tópico, Searle conclui que o experimento em si não refuta o livre-arbítrio. Em termos descritivos, o que o experimento na verdade mostra é apenas que ocorreu um aumento da atividade na área motora suplementar antes que o sujeito tomasse consciência de iniciar sua ação. Searle termina insistindo que o experimento não apresenta nada de consistente que invalide nossa concepção de livre-arbítrio. Ele afirma que "os sujeitos já haviam decidido fazer alguma coisa - em meu jargão, eles haviam formado uma intenção prévia - mas não havia nenhuma sugestão de que a situação anterior à ação fosse de alguma forma suficiente para causar a ação" (Searle, 2015, p. 214).

Por fim, o terceiro caso (3), sobre os reflexos. Todos nós já tivemos algumas experiências na qual os nossos reflexos atuaram de modo imediato antes mesmo que tivéssemos consciência. O caso mais comum é quando encostamos em algo quente com o nosso braço e de imediato o retiramos sem sequer ter tempo de pensar e tomar uma decisão. Dito de outra forma, iniciamos o movimento corporal antes mesmo de o percebermos. De acordo com Searle (2015), essa lacuna entre nossa consciência e os reflexos ocorre porque nosso cérebro pode levar até meio segundo ao processar um sinal de entrada até à consciência. Por isso, nosso corpo inicia a ação antes da percepção. O problema não é o fato de que nossos reflexos ocorrem numa velocidade maior que nossa percepção, o problema é que isso termina, nas palavras Searle (2015, p. 213), sendo "a ponta de um iceberg de pesquisas que sugerem que a consciência realmente não importa muito".

A visão *standard,* desde os experimentos de Libet, é que a consciência teria o papel de monitorar nosso comportamento, mas não de iniciá-lo, ou seja, seu início se dá inconscientemente e, uma vez que tomamos consciência, só nos resta vetá-lo ou não. A ideia é muito simples, nós não escolhemos a vontade de desejar o que desejamos ou fazer o que fazemos, mas podemos nos recusar a realizá-la, isto é, "a consciência pode vetar uma ação que de outra forma teria ocorrido" (Searle, 2015, p. 213). Searle cita como exemplo recente dessa postura o trabalho de Marc Jeannerod[14], o qual afirma que o processamento mental consciente é melhor elucidado quando pressupomos processos inconscientes realizando verdadeiramente suas atividades.

> A consciência pode funcionar como uma espécie de policial para guiar nossas ações e pode até mesmo vetar certos tipos de ações, mas o verdadeiro motor que impulsiona a cognição e o comportamento humano é inconsciente. Acho

14 JEANNEROD, Marc. **Motor Cognition**: What Actions Tell the Self. Oxford: Oxford University Press, 2006.

> que essa visão está totalmente equivocada e não é apoiada pela evidência experimental (...). (Searle, 2015, p. 213).

A conclusão de Searle (2015) em torno desse debate é que, ao contrário desse tipo de literatura, a consciência é de extrema importância, basta pensarmos numa série de atividades cotidianas da qual precisamos da consciência. Poderíamos nos imaginar dirigindo um carro, pedalando uma bicicleta, lendo um texto ou escrevendo-o inconscientemente?[15] Para Searle (2015), a resposta é de natureza negativa. De toda forma, a consciência importa e ela continua sendo um desafio tanto no cenário filosófico quanto científico.

Considerações finais

Como foi afirmado ao longo do texto, o tema da percepção inconsciente é apenas um subconjunto da discussão sobre o inconsciente. Ainda em sua atual teoria da percepção, Searle enfatiza a ideia vista na seção sobre o inconsciente, a saber, para que alguma coisa seja considerada um fenômeno mental inconsciente, esta coisa deve ser algo que possa tornar-se consciente. Isso é uma condição indispensável para que possa haver aquilo que Searle chama de realidade psicológica, com conteúdo intencional e forma aspectual.

Como vimos, muito de sua crítica gira em torno do inconsciente profundo, que se comporta como se fosse mental, mas que na verdade são apenas processos neurobiológicos. Portanto, toda a discussão de Searle em torno do inconsciente e, consequentemente, da percepção inconsciente é exclusivamente dependente de seu modelo explicativo de uma teoria intencional da consciência e da percepção. Searle não nega que existem processos mentais inconscientes, muitos processos neurobiológicos estão acontecendo neste exato momento em nossa experiência perceptiva. Todavia, isso não significa que se trata de alguma realidade mental, mas apenas de algo que é condição de possibilidade para os diversos estados mentais.

15 Um desafio para essa questão são os casos de sonambulismo. Embora muito específicos, ainda assim se configuram como um contrafactual à questão formulada. Tenho um texto em andamento sobre o tema, mas ainda é preciso amadurecer alguns *insights*.

Referências

ARMSTRONG, David, M. **A materialist theory of the mind**. London. Routledge & Kegan Paul. 1968.

ARAUJO, J. P. M. Representacionalismo e realismo direto na teoria da percepção de John. R. Searle. In: SOUZA, Marcus José Alves de & LIMA FILHO, Maxwell Morais de (Orgs.). **Escritos de filosofia IV:** Linguagem e Cognição. Porto Alegre. Editora Fi, 2020.

BLOCK, Ned. The Anna Karenina principle and skepticism about unconscious perception. **Philosophy and Phenomenological Research**, Vol. 93, No. 2, 2016, p. 452-9.

BRENTANO, Franz. **Psychology from a empirical standpoint**. New York. Routledge. 1995 (original 1874).

DAVIDSON, Donald. Actions, reasons, and causes. **The Journal of Philosophy**, Vol. 60, No. 23, (1963), p. 685-700.

DEECKER, Lüder; GRÖZINGER, Berta; KORNHUBER, H. H. **Voluntary finger movement in man**: Cerebral potentials and theory. Biological Cybernetics, vol. 23, 1976, p. 99-119.

FREUD, Sigmund. Algumas observações sobre o conceito de inconsciente na psicanálise. (original 1912). In: FREUD, S. **Obras Completas**, vol. 10, Observações psicanalíticas sobre um caso de paranoia relatado em autobiografia ("O caso Schreber"), artigos sobre técnica e outros textos (1911-1913). (*Tradução* Paulo César de Souza). São Paulo. Companhia das Letras, 2011.

GRICE, H. Paul. The causal theory of perception. **Proceedings of the Aristotelian Society**, Supp. vol. xxxv, 1961, p. 121-53.

JEANNEROD, Marc. **Motor cognition**: What actions tell the self. Oxford: Oxford University Press, 2006.

LIBET, Benjamin. 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, no. 3, 1983. p. 623-42.

LIBET, Benjamin. Unconscious cerebral initiative and the role of conscious will in voluntary action. **Behavioral and Brain Sciences**, Vol. 8 , Issue 4, 1985 , p. 529-39.

MARTIN, Michael G. F. The limits of self-awareness. In: BYRNE, Alex & LOGUE, Heather (Eds.). **Disjunctivism**: Contemporary Readings. Massachusetts. The MIT Press, 2009, p. 271-317.

MAUND, Barry. **Perception**. Chesham. Acumen, 2003.

MILNER, David; GOODALE, Mel. **The visual brain in action**. Oxford: Oxford University Press, 2006.

PHILIPS, Ian. Unconscious perception reconsidered. **Analytic Philosophy** Vol. 59 No. 4, 2018 p. 471–514.

REID, Thomas. **Essays on the Intelectual Powers of Man**. Edited by A. D. Woozley. London. Macmillian. 1941 [1785]

SEARLE, John. **Intentionality**: An essay in the philosophy of mind. Cambridge University Press, 1983.

SEARLE, John. **Intentionality and it's place in nature**. *Synthese* 61 (1984) p. 3-16.

SEARLE, John. Consciousness, unconsciousness and intentionality. **Philosophical issues**, 1991, Vol. 1, Consciousness (1991), p. 45-66

SEARLE, John. **The rediscovery of the mind**. Massachusetts: The MIT Press, 1992.

SEARLE, John. **Rationality in action**. Massachusetts: The MIT Press, 2000.

SEARLE, John. **Mind**: A brief introduction. Oxford University Press, 2004.

SEARLE, John. Perceptual intentionality. **Organon F**, 19, 2012, p. 9-22.

SEARLE, John. **Seeing things as they are**: A theory of perception. Oxford University Press, 2015.

SOTERIOU, Matthew. **Disjunctivism**. New York: Routledge, 2016.

STICH, P. Stephen. Beliefs and subdoxastic states. **Philosophy of Science**, Vol. 45, No. 4 (Dec., 1978), p. 499-518.

WEISKRANTZ, Lawrence. **Blindsight**: A case study spanning 35 years and new developments. Oxford: Oxford University Press, 2009.

WITTGENSTEIN, L. **Philosophical investigations**. Translated by G. E. M. Anscombe. Revised by Peter Hacker and Joachim Schulte. Oxford: Wiley-Blackwell. (german-english bilingual edition) 2009 [1953].

EL PROBLEMA DE LA MENTE Y EL CUERPO SOLUCIONADO: NIETZSCHE Y DENNETT

Mariano Rodríguez González
(Universidad Complutense de Madrid)

Resumen:

No es difícil advertir las coincidencias entre Nietzsche y Dennett. Este trabajo subraya su trascendencia. Y es que se puede decir que los dos filósofos habrían pretendido "resolver" el problema *filosófico* de la mente y el cuerpo de una manera básicamente similar: desde el continuismo o gradualismo de la evolución biocultural de la mente humana, o sea, desde el no-esencialismo que desmiente a Platón y a Aristóteles y en general a la tradición dominante del pensamiento occidental. Pero, eso sí, desde dos paradigmas muy diferentes que con Shannon se iban a separar a mediados del siglo XX, para desconsuelo de un autor actual como T. W. Deacon (2011). El de la termodinámica o la energía, y el de la información. Por eso hay que tomarse muy en serio que dos pensadores enmarcados en dos tradiciones contrapuestas vengan a coincidir en lo fundamental de su solución. Porque entonces esta coincidencia es una razón poderosa para considerar si no se habría resuelto ya el viejo problema filosófico de la mente y el cuerpo.[1]

Abstract:

The coincidences between Nietzsche and Dennett have been noted. This paper underlines the significance of these coincidences. It can be said that the two philosophers would have tried to "solve" the *philosophical* problem of mind and body in a basically similar way: from the continuity or gradualism of the biocultural evolution of the human mind, that is, from the non-essentialism that disproves Plato and Aristotle and in general the dominant tradition of Western thought. But, yes, from two very different paradigms that with Shannon were to separate in the mid-twentieth century, to the dismay of a current author like Deacon (2011): that of thermodynamics or energy and that of information. That is why we should take very seriously the fact that two thinkers framed in two opposing traditions come to agree on the fundamentals of their solution. Because then this coincidence is a powerful reason to consider whether the old philosophical problem of mind and body has already been solved.

1 En el sentido que tendría, para nosotros, resolver un problema filosófico, o "misterio", a diferencia de resolver un problema científico. A la pregunta esencial de por qué el problema mente-cuerpo habría dejado ya de ser un "problema filosófico", respondería Dennett en CE: «A mystery is a phenomenon that people don't

Palabras clave:

continuismo, gradualismo, información, energía, problema filosófico.

Keywords:

Continuity, Gradualism, Information, Energy, philosophical Problem.

I. Introducción

En lo que sigue presentamos una argumentación que mostraría cómo, tanto Nietzsche, tomando pie en la biología y la fisiología de su época, como Dennett, basándose en el evolucionismo darwiniano y la Inteligencia Artificial (DENNETT, 1991, 2017), habrían podido resolver *definitivamente* el tradicional problema filosófico de la relación psicofísica. Un problema este que, para pensadores como Popper (1974), sería el fundamental del pensamiento moderno, e incluso de toda la filosofía occidental[2].

Como mínimo, es seguro que estos dos filósofos se esforzaron en resolver o disolver el célebre problema. Por parte de Nietzsche, con lo que Abel (2001) iba a denominar en su estudio de la filosofía nietzscheana de la mente "*modelo del continuo*", entendido éste en el sentido en el que se podría afirmar que la formación de cristales en el mundo inorgánico es ya "pensamiento". O sea, que una continuidad natural llevaría sin grandes saltos desde las estructuras de los cristales hasta las estructuras del pensamiento humano.

> Nuestros pensamientos más elevados y audaces son partes del modo de ser de la "realidad". Nuestro pensamiento es de la misma materia que todas las cosas (NIETZSCHE, 2008, p. 838).

Y en el caso de Dennett, con su *gradualismo*. La tesis de que la *comprensión* se despliega por grados, igual que en general la conciencia; la tesis de que no sería, ninguna de las dos, cuestión de todo o nada. Hay que «cerrar la herida dualista» que abriera Descartes en su día[3]. Y cerrarla, por supuesto, abrazando un naturalismo general, si bien en

know how to think about—yet» (1991, Pos 455).

2 «La filosofía occidental consiste fundamentalmente en representaciones del mundo que no son sino variaciones sobre el tema del dualismo del cuerpo y la mente, así como en problemas de método relacionados con ellas» (POPPER 1974, p. 147).

3 O la herida que empezó a abrirse con la separación parmenídea entre los sentidos y la razón como dos modos absolutamente diferentes de conocimiento, según Nietzsche el origen de la posterior invención platónica de alma.

diferentes versiones para nuestros dos filósofos, naturalismo realista "suave"[4], en el caso de Dennett, y perspectivista en el de Nietzsche[5].

Para suturar esa herida dualista, Dennett va a partir de Darwin y Turing, en la estrategia *decisiva* de seguirle el rastro a *la evolución de la mente*. Porque Darwin y Turing habrían realizado una «extraña inversión del razonamiento», como la denominará Dennett siguiendo a un crítico decimonónico de la idea de selección natural. *Inversión* en el sentido de que ese razonamiento hace concebibles o viables las *competencias* completamente desprovistas de *comprensión*. Cuando lo que el sentido común nos ha venido diciendo hasta ahora es que la comprensión sería *conditio sine qua non* de toda competencia: "evidentemente", para hacer una cosa primero hay que *saber* cómo hacerla, y por eso habría entrado en este punto el Divino Hacedor con su Divina Sabiduría. Pero una máquina universal de Turing, como el ordenador digital tipo von Neumann, es capaz de computar cualquier cálculo, sin "saber" en absoluto aritmética, ni por tanto comprender el sentido de lo que hace. *Del mismo modo*, los animales no humanos se involucran en conductas beneficiosas para su supervivencia y reproducción, que entonces tendrían todo el sentido o toda la "racionalidad". Pero está claro que sin tener la menor idea de por qué ni para qué lo hacen.

Ahora bien, al mostrarnos competencias sin comprensión, e incluso, *la posibilidad real de una cascada de competencias que irían escalando hasta la emergencia de la comprensión*, y por tanto, comprensión a medias, y a cuartos; y si reconocemos que con esta mostración habríamos suturado por fin la herida dualista, resolviendo el problema filosófico de lo mental, entonces inmediatamente nos ha surgido otro problema sin duda también difícil, el de para qué sirve y cómo surgió la comprensión, y con ella la conciencia en el sentido humano del término. Es decir, surge entonces el asunto de la enigmática *función de la conciencia*, con el que iba a ocuparse Nietzsche (2014, p. 868-9) magistralmente en el aforismo 354 del Libro V de *La gaya ciencia*[6].

---

4 El *mild realism* de Dennett no deja de asemejarse, en lo que respecta a la matización interpretacionista que supone su importante doctrina de las *stances* (posturas, posiciones), a lo que en mi opinión podría ser una versión lúcida del perspectivismo nietzscheano, precisamente aquella que por fin consiga hacer compatibles el interpretacionismo nietzscheano con el peculiar realismo final del pensador alemán.

5 Nietzsche no dejará de criticar a "los que hoy naturalizan en el pensamiento", refiriéndose con ello a los que extraen en su época un materialismo ingenuo a partir de las ciencias naturales.

6 Nietzsche dará, a título de conjetura, la célebre respuesta que más abajo veremos, respuesta recogida por Dennett, y muy parecida a la de Marx que también cita con aprobación el americano. Pero es de notar que en su libro de 2009 un filósofo de la mente de la talla de Thomas Metzinger llegará a listar un buen número de funciones concretas que la investigación científica reconocería en la actualidad a la conciencia.

Concentramos en este trabajo el problema de la mente y el cuerpo en su versión de problema de la conciencia, pero a su vez, este problema de la conciencia lo enfocaremos como el problema de la comprensión. Y esto por muchas razones: en primer lugar, se trataría de la cuestión filosófica esencial que en nuestros días levanta la IA, y no hay que pasar por alto que Dennett, como nos recordaba Boden (2018), es un filósofo esencialmente inspirado en la IA. ¿Es la inteligencia artificial inteligencia *real*? O sea, ¿comprenden lo que hacen las máquinas que consideramos inteligentes?

No se sabe cómo responder a esta pregunta de un modo terminante. Se trata, cómo no, de la cuestión que vuelve a plantear la crítica que Searle (1985) le dirigiera a Turing y al proyecto de la IA fuerte, el de que su famoso Test no pasaría de ser puro operacionalismo irrelevante. Caemos así en la cuenta de que en este asunto de la *comprensión* podremos resumir del mejor modo lo esencial del problema de la mente y el cerebro. Porque en él, que sin duda incluye la *experiencia* de la comprensión, se darían cita las tres cuestiones centrales de la conciencia, la intencionalidad, y el yo, aquellas que para un pensador como Feigl (1967) conforman el problema de la mente y el cuerpo. Y si, en segundo lugar, nos situamos en la perspectiva histórica, tenemos que decir que la perplejidad dualista, que planteó nuestro problema en Descartes, habría brotado *también* del mayúsculo enigma del científico moderno que es capaz de *comprender* el universo mecánico, y que entonces, "lógicamente", es inconcebible que forme parte del universo mecánico.

Como iba a reiterar en el siglo XX el mismo Popper (1994/1997), filósofo también dualista pero en el sentido débil del "emergentismo" de propiedades, un sistema nervioso no podría *entender* jamás la resolución de un problema matemático, como tampoco la podría entender un ordenador por mucho que ejecute el programa adecuado, y por eso el ordenador no es más que un sustituto del lápiz y el papel, solo que más sofisticado—y decir lo contrario sería de todo punto *incomprensible*[7]. Así que el cerebro igual que el ordenador no sería más que herramienta del Yo: *El yo y su cerebro* (1977/1980). Pero entonces se nos plantea la peliaguda cuestión de qué es propiamente ese Yo que usa el cerebro, es decir, reaparece el problema de la mente y el cuerpo.

Conviene mostrar qué entendemos por "comprensión", porque los filósofos hacen del término interpretaciones muy diferentes, como es usual. Para Wittgenstein la experiencia de comprender (el ¡*eureka*!) debía ser operacionalizada como la capacidad de se-

7 Como le resultaba también incomprensible a Platón la imposible capacidad corporal de entender ideas universales, *y por eso mismo tuvo que inventarse el 'alma' espiritual*, como piensa R. Rorty en su excepcional obra de 1979 *La filosofía y el espejo de la naturaleza.*

guir una regla[8], porque solo si lo hacemos así puede ser comprendida la comprensión (1988, §§ 143-154). Pero en el lado en apariencia contrario, Nietzsche había caracterizado en uno de sus apuntes a la *experiencia* constitutiva de la comprensión como experiencia *emocional*. Es decir, una experiencia *que en último término remite a las pulsiones*[9], por lo que no es posible comprender la comprensión, y entonces hay motivos para mostrarse escéptico respecto a su justeza epistémica, lo que significa que, en principio, no puedo estar seguro de si de verdad entiendo lo que creo entender[10]. Ahora bien, en el ámbito de la filosofía de la mente de inspiración cognitivista, la comprensión tendría que ver, por supuesto, con el significado lingüístico o la intencionalidad psíquica. En Dennett tiene que ver la comprensión con nuestra capacidad de representarnos razones para actuar, y de evaluarlas y criticarlas. Se trata así de la racionalidad que se le atribuye a la especie humana, y que se concentra en Dennett en la explicación por propósitos o fines, entendida principalmente en un sentido instrumental y en el marco darwiniano.

Ahora bien, dar con la solución de un problema que nos preocupa se nos muestra, también, como una *experiencia* humana, experiencia peculiar, puede que placentera o incluso dolorosa. Podemos pensar así que, en el problema fundamental de la filosofía de la mente, que es el de la mente y el cuerpo, la cuestión de la comprensión guardaría una relación esencial con la de la conciencia. Y vamos a poner aquí de manifiesto el sorprendente aire de familia entre las respectivas teorías de la conciencia de nuestros dos autores principales. Pues sucede que, para ambos, en tanto conciencia propiamente humana, o sea, conceptual y lingüística, *la conciencia emerge a partir de la imperiosa necesidad de comunicación, y por eso tendría un carácter esencialmente social.* Se trata entonces, con esto, de un impulso filosófico, compartido por los dos filósofos, radicalmente contrario a la posición solipsista cartesiana[11].

---

8 «¡No pienses ni una sola vez en la comprensión como 'proceso mental'! — Pues ésa es la manera de hablar que te confunde. Pregúntate en cambio: ¿en qué tipo de caso, bajo qué circunstancias, decimos 'Ahora sé seguir'?, quiero decir, cuando se me ha ocurrido la fórmula—En el sentido en el que hay procesos (incluso procesos mentales) característicos de la comprensión, la comprensión no es un proceso mental. (La disminución y el aumento de una sensación dolorosa, la audición de una melodía, de una oración: procesos mentales)» (WITTGENSTEIN, 1988, I, §154, 155).

9 «¿Qué quiere decir '*entender una idea*'? La idea induce una representación, la representación induce percepciones, y las percepciones excitan sentimientos, y cuando la piedra llega por fin al fondo se produce un sonido sordo: llamamos 'entender' a esta sacudida del fondo. No hay aquí ni causa ni efecto, sólo asociación: cierta palabra suele despertar cierta representación: *cómo* sea esto posible, nadie lo sabe. Nuestro 'entender' es algo inentendible, y esa última resonancia en las pulsiones no es sino otra cosa desconocida más, nueva y grande. – (…)» (NIETZSCHE, 2008, 645-46).

10 Con lo que podríamos incluso pensar que Nietzsche tiene que concluir pensando de la comprensión algo muy parecido a lo que habíamos visto pensaba Wittgenstein, porque en realidad habríamos llegado a la imposibilidad de establecer criterios de corrección en un lenguaje que solo yo pueda entender.

11 Por supuesto que hay diferencias e incluso puntos de oposición entre Nietzsche y Dennett, cuya explora-

II. Comprensión y competencias.

Esa aportación de Darwin al conocimiento humano que fue *el algoritmo de la adaptación*, la explicación mecánica, ciega, de la evolución de las especies por selección natural, no la habríamos comprendido, en lo filosófico de su alcance, hasta la llegada del ordenador digital, la célebre máquina universal de Turing. El golpe maestro de la interpretación dennettiana es el de desvincular ese algoritmo adaptacionista de su realización concreta en la biosfera, para llevarlo, en calidad de fórmula informacional clave, y neutral respecto de su medio de implementación, a todos los niveles de la realidad, por supuesto también el cultural (THOMPSON, 2009). Por eso los dos científicos británicos, Darwin y Turing, habrían podido llevar a cabo lo que llama Dennett «una extraña inversión del razonamiento», inversión que se puede resumir en la idea de las «razones o justificaciones de flotación libre» (*free-floating rationales*). El ordenador digital ejecuta cualquier tarea computacional sin necesidad ninguna de comprender el sentido de lo que hace, aunque la tarea desde luego lo tenga si la contemplamos bajo un cierto aspecto. Pero pensemos por otra parte, por ejemplo, en el pájaro que finge tener un ala rota *para engañar*, sugiriéndole la perspectiva de una caza fácil, al depredador que ha detectado sus huevos puestos en tierra, *y que entonces deja de interesarse en ellos* (DENNETT, 2017, p. 91). O consideremos, por si eso fuera poco, a la larva de la mosca *caddis*, que está provista de una trompa recogedora de plancton muy parecida en su diseño a una trampa humana efectivamente usada para capturar langostas (DENNETT, 2009, p. 1063 a y b). Hay entonces *razones*, que habría seguido la madre naturaleza o la selección natural, para todas estas conductas, y todos estos diseños por el estilo. Unas razones que los humanos podemos descubrir con el método habitual de los biólogos, el de la *ingeniería inversa*: considerar al organismo natural como un artefacto. Con este tipo de "razones", en realidad, *de lo que se trata de la idea de "función", a medio camino entre el puro mecanismo de la postura física y la teleología de la postura intencional*.

Esta inversión darwiniana pudo parecer absurda a muchos lectores en aquellos años del siglo XIX, sin duda, pero en la actualidad el desarrollo de la IA a partir de Turing

---

ción llevaría directamente a la problemática relación del filósofo alemán con Darwin y los darwinistas de su época. Pero nos vamos a centrar aquí solo en su pronunciado paralelismo y en sus sorprendentes convergencias, sin entrar a debatir otras cuestiones. Como la de "Nietzsche contra Dennett", por ejemplo: habría sido el alemán más riguroso en la lucha contra el antropomorfismo (ANDRESEN, 2015). O como la de "Dennett contra Nietzsche": reivindicación entusiasta del darwinismo, que admite y postula *Cranes*, grúas, pero en absoluto *Skyhooks*, ganchos celestes obsoletos y científicamente inaceptables como el de la voluntad de poder: «Nietzsche's idea of a will to power is one of the strangest incarnations of skyhook hunger, and, fortunately, few find it attractive today» (DENNETT, 1999, Pos 9137).

nos habría mostrado que es una posibilidad perfectamente real[12]. Así que los dos genios de la ciencia nos han enseñado de un modo incontestable, Darwin y Turing, que hay por todas partes *competencias sin comprensión*. Por 'comprensión', Dennett se refiere a algo así como el entender humano, o sea, la habilidad de valorar los principios que explican por qué una cosa es como es. Mientras que 'competencia' quiere simplemente decir la habilidad que tiene un organismo de lograr un fin, bien sea encontrar comida, impresionar a una pareja, resolver una ecuación» (RATHKOPF, 2017, p. 1357).

Sería la comprensión esa capacidad que se pone de manifiesto, en su forma más madura hasta la fecha, en la especie *homo sapiens*, la capacidad de representarme la situación y representarme a mí mismo en ella, a fin de poner en marcha la acción intencional más adecuada, a partir del descubrimiento *consciente* de determinadas *razones*. Unas razones caracterizadas como *ancladas*, frente a las que flotan de manera libre en los ámbitos inconscientes de la biología o de las máquinas inteligentes. Aplicar así *la postura o posición intencional* a un sistema, orgánico o mecánico, no es otra cosa que atribuirle un cierto grado de comprensión *o de racionalidad* (DENNETT, 1997). Y ello no solo como estrategia de explicación y predicción sino también por supuesto de comprensión[13].

Para Nietzsche, como ya hemos referido, sería la comprensión, básicamente, una experiencia emocional, y entonces perteneciente al ámbito de la conciencia, pero, al final, también al de la pulsión. Una experiencia que por tanto no nos traerá la seguridad de haber efectivamente comprendido, también porque toda experiencia consciente tendría para Nietzsche la naturaleza de la falsificación. Mientras que nos podríamos aventurar a decir que las competencias dennettianas tendrían que ver con lo que el filósofo alemán denomina, la «inteligencia más elevada y amplia» (NIETZSCHE, 2010, p. 448)[14], que es para él la de la corporalidad, es decir, los procesos pulsionales (corporales) inconscientes, esos que «no son yo, pero hacen yo».

12 A lo que es lógicamente posible lo restringe lo que es físicamente posible, pero a lo que es físicamente posible lo viene a recortar de forma drástica lo que sería viable desde el punto de desarrollo alcanzado en este momento por la evolución biológica. Esto nos lo recuerda Dennett, sin duda para poner coto a los excesos modales de tantos *thought experiments*.

13 De manera que tendríamos: explicaciones causales que corresponden a la posición física; razones libremente flotantes que se relacionan con la posición del diseño en la dimensión propiamente funcional de los sistemas orgánicos y en las máquinas (explicaciones funcionales); razones, en tanto lingüística o conscientemente representadas, capaces de guiar de arriba abajo nuestras acciones, y que son detectadas sólo cuando ocupamos la posición intencional (explicaciones intencionales).

14 Las agencias y los agentes plurales de Marvin Minsky (1986), uno de los padres de la IA que más han influido en el pensamiento de Dennett, se dejarían comparar bien con la idea nietzscheana de la corporalidad pulsional.

El camino que asciende acumulativamente desde las competencias más básicas hasta la emergencia de la comprensión al estilo humano ha sido el de un "progreso" paulatino, gradual, continuo, sin ninguna frontera nítida entre las diversas formas por las que ha venido transcurriendo. Pero serían los *memes*[15], para Dennett, como puente necesitado para salvar la sempiterna brecha[16], la genuina condición necesaria de la comprensión. El cerebro de un mono sería, a fin de cuentas, para nuestro pensador, una máquina bayesiana que detecta patrones relevantes en el entorno a fin de calcular la probabilidad de patrones no detectados actualmente. Pero para que ese cerebro se convierta en una *mente consciente* del tipo humano, lo que hará falta es que sea infestado por *memes*. Es decir, por modos de hacer, hábitos, prácticas, herramientas de pensamiento (*thinking tools*) gregorianas[17] (DENNETT, 2017, p. 98), diseñadas en otro lugar. Los memes son verdaderas unidades informacionales[18] de replicación cultural sobre cuya reproducción diferencial, que explicaría la evolución de las culturas, estaría sin embargo operando una vez más la selección natural. Esos memes pueden ser, como los virus, beneficiosos o perjudiciales para la supervivencia del organismo huésped, pero para ellos lo que cuenta es su propia reproducción, igual que le ocurría al tan egoísta gen egoísta (DAWKINS, 1976). Una invasión memética semejante habría ocurrido hasta ahora solo en el *homo sapiens* (DENNETT, 2017, p. 388-9).

En definitiva, fue la transmisión *cultural* de la información lo que llevó de las meras competencias a la comprensión, transformando el cerebro animal en una mente consciente propiamente dicha, del tipo humano. La teoría memética tiene tanta importancia para Dennett porque su concepto básico nos va a permitir garantizar al modo naturalista el principio gradualista que resuelve o disuelve el problema psicofísico. O sea, el meme le permitirá al filósofo eludir el eterno problema del huevo y la gallina, porque ese concepto aseguraría que hay «una profunda conexión entre la dinámica del cambio cultural y la dinámica del cambio biológico: los dos son fundamentalmente informacionales y los dos se hallan sujetos a la selección natural» (RATHKOPF, 2017, p. 1360).

---

15 Cf. Dawkins, Richard (1976): *El gen egoísta. Las bases biológicas de nuestra conducta*. Barcelona, Salvat Ediciones, 1993, pp. 247-263.

16 La brecha, o esa *Unüberbrückbarkeit* en la que Wittgenstein ponía lo esencial del problema del cerebro y la conciencia (la *imposible imposibilidad*). *Cfr. Ph.Unt.* I, 412.

17 Así denominados por Dennett en honor del psicólogo Richard Gregory, quien llamó nuestra atención sobre estas *herramientas del pensamiento*.

18 No en el sentido de la teoría matemática de la información, sino en el de información *semántica*. Pero entendida ésta, a su vez, no tanto como conocimiento de hechos abstractos, sino como conocimiento de cualquier cosa que ayude a un organismo a comportarse de forma adaptativa. Es decir, información semántica, pero en el sentido de información dirigida a un *know-how* (nos habla Dennett en este punto de un *design worth getting)*.

Sin duda, el meme más importante para la evolución de las mentes es la palabra, y la transmisión cultural que conformará la conciencia de los individuos no es sino una tradición lingüística y literaria. Nada nuevo hay en esto, ya Popper había subrayado que es nuestra interacción lingüística con el M3 de los productos culturales lo que da a luz a la "mentalidad". Pero el hecho de partir de las cuatro funciones del lenguaje hecho y derecho, incluyendo la superior función argumentativa, le habría llevado a Popper inevitablemente de vuelta al dualismo (RODRÍGUEZ GONZÁLEZ, 2008). Mientras que el concepto de meme, entendido como mecanismo cultural replicante que se propaga por simple imitación, como las modas, permite a Dennett perseverar en el materialismo científico salvaguardado por el principio del continuo. Previo al nacimiento del juego del sentido tendríamos el juego meramente imitativo con los fonemas que no significan nada pero que serán la base de todo significado lingüístico[19].

Podemos asumir, en este punto, que las ideas nietzscheanas de pulsión y de instinto se localizarían justamente, asimismo, en este quicio entre natura y cultura. Incluso llegará a decir el filósofo alemán que las pulsiones *se nos enseñan en el Estado* (NIETZSCHE, 2008, p. 798)[20]. Pero lo que me interesa subrayar, por supuesto, es la importancia de una lectura *funcionalista* de la pulsión y el instinto como los trata Nietzsche. Nos dirá, por ejemplo, de la pulsión, que en realidad es simplemente un nombre para cubrir nuestra ignorancia respecto del preciso mecanismo de naturaleza físico-química que estaría en obra en las conductas de los organismos (NIETZSCHE, 2008, p. 328)[21]. Resulta ilustrativa también, en esta misma línea «reduccionista»[22], la comparación nietzscheana de la pulsión

19 La teoría memética es la única capaz de explicar o servir de marco inicial de explicación del hecho de que después del suicidio de un adolescente se eleva la tasa de suicidios de adolescentes durante semanas o incluso meses, el llamado "efecto Werther". Puesto que está claro que suicidarse ni favorece el éxito reproductivo ni es algo que por su valor pueda suscitar en el sujeto ninguna suerte de adhesión racional. «Entonces podemos y tenemos que decir que la idea de suicidio se extiende porque ha sido diseñada para reproducirse en su nicho ecológico, es decir, en un conjunto de cerebros humanos» (RATHKOPF, 2017, p. 1361).

20 «Así pues: el Estado originariamente no oprime a lo que diríamos los individuos: ¡éstos no existen todavía! Les hace posible la existencia a los hombres, animales gregarios. ¡En él se nos *enseñan* las *pulsiones*, los *afectos, no son nada originario*!¡Carecen de 'estado de naturaleza'!» (NIETZSCHE, 2008, p. 798).

21 «En general, la palabra instinto [*Trieb*] no es más que un cómodo recurso y se emplea por doquier ahí donde los efectos regulares en los organismos aún no han sido remitidos a sus leyes químicas y mecánicas» (NIETZSCHE, 2008, p. 328).

22 Es clarificadora la distinción dennettiana entre un reduccionismo en sentido trivial, en el que todos estamos de acuerdo (nadie dudará de que el presidente del gobierno es un conglomerado de robots, las macromoléculas), y un reduccionismo avaricioso (*greedy*) que es completamente insensato (intentar explicar las medidas de gobierno del presidente a partir del comportamiento de las macromoléculas que lo integran). (Cfr. Dennett,1999).

con ese circuito que forman la mano del pianista que toca el piano, los nervios del pianista, y los procesos que se desarrollan en algún centro de su cerebro (NIETZSCHE, 2020, p. 217)[23]. Las partes del cuerpo ligadas telegráficamente ¡¡eso es la pulsión!!, junto con una meta inconsciente. Si podemos admitir que el meme sería como una especie de software, un programa, una máquina virtual o una aplicación que descargo en mi móvil, no vendría mal repasar, en otra ocasión, las consideraciones de Abel (2001) sobre un funcionalismo nietzscheano para reforzar la solidaridad profunda entre los dos autores.

Con ese *meme* crucial que es la palabra llegó el lenguaje como tal, lo que nos iba a colocar a los humanos en el territorio de la comprensión. O bien, si queremos seguir a Dennett en su uso continuo de "metáforas" informáticas, el lenguaje nos abriría la posibilidad de diseñar nuestros productos culturales *de arriba abajo* (*top down*), diseñarlos a la manera *inteligente* propia de la GOFAI, con un diseño des-darwinizado o apartado del "ensayo y error" universal. Con el lenguaje humano se les entrega por primera vez a unas criaturas del planeta Tierra la inaudita posibilidad de acceder a lo que Sellars había bautizado como *espacio lógico de las razones*[24]. Con el lenguaje humano surgen razones de una clase nueva, *razones en tanto conscientemente representadas*, que entonces ya habrían dejado de ser como todas las que rigieron el darwinista proceso anterior. Este proceso anterior es de tipo *bottom-up*, el del «pensar sin esfuerzo», pensar no consciente pero muy efectivo, como el que modelan las redes neuronales artificiales. Un proceso que llamaríamos "no inteligente", en suma, pero sólo en ese sentido de contraste que quiere darle Dennett. En el ámbito inconsciente de este «pensar» tienen lugar los eventos "causalmente" efectivos de esa "inteligencia más abarcadora" nietzscheana que es el cuerpo pulsional.

A esa "semicomprensión" animal, por llamarla de algún modo—tal vez mejor una no-comprensión, porque no sería nada más que un mero *saber-hacer*—, la comprensión del tipo humano añade entonces la capacidad de representar(se) explícitamente, mediante palabras, diagramas u otras «herramientas de autoestimulación» del pensar, otros *memes*, cualquier cosa o tópico, para considerarlo, analizarlo, evaluarlo, criticarlo (DENNETT, 2017, p. 300).

Lo que hay que tener en cuenta, sobre todo, es que habría sido la imperiosa necesidad biológica de mantener la relación con el otro, en nuestra forma de vida de especie necesitadamente social, la que exigió el enorme desarrollo que ha alcanzado en nosotros

23 «(…) La mano del pianista, la vía hasta allí y un sector del cerebro forman juntos un órgano (el cual debe *aislarse* para poder contraerse fuertemente). *Las partes del cuerpo ligadas telegráficamente—es decir, instinto* (*Trieb*). ¡Schopenhauer ha proporcionado además el *fin inconsciente*!» (NIETZSCHE, 2020, p. 217)

24 *Cf.* Wilfrid Sellars, 1963a.

los humanos la práctica comunicativa de dar razones de lo que hacemos y lo que pensamos, dárnoslas también a nosotros mismos, así como por supuesto de evaluarlas. Y es que hemos dejado de ser animales simplemente skinnerianos y popperianos, seríamos ya animales gregorianos, o sea, muy bien adiestrados en la práctica con *thinking tools* como los *memes*, herramientas que multiplican la potencia y el alcance del pensar. Sin ellas no podríamos siquiera formular nuestros pensamientos, no podríamos pensarlos a nuestro modo humano.

Haber alcanzado este "milagro" de la comprensión como experiencia de la conciencia no sería en realidad ningún milagro, sino simplemente el resultado de nuestra costumbre reflexiva de criticar nuestras propias razones en lo que hace a su coherencia y adecuación. Esta reflexión autocrítica ahora mismo se estaría intentando llevar sobre los sistemas de la IA con el fin de dotarlos de atisbos de conciencia. Pero al menos de momento no se consigue, y la verdad es que no habría ningún indicio serio de que se vaya a conseguir en un futuro próximo o ni siquiera a medio plazo.

En otro orden de cosas, la importancia universal de la filosofía procedería de que ella ha venido siendo el apartado cultural destinado al cultivo de la comprensión. Se trata con la filosofía de la meta-representación como profesión, esa práctica puesta a punto por Sócrates, Platón y Aristóteles. Sería el de la filosofía el ideal de comprenderlo todo, incluida la comprensión misma, el ideal de consumar la auténtica humanidad de los humanos. Pero en nuestros días hay indicios de que se aproxima el que algunos denominan «Gran Oscurecimiento», por cuanto los sistemas de la IA habrían ingresado en una fase post-inteligente o post-comprensiva. Y es que la gran ciencia, como la del tipo CERN, elabora modelos y teorías y proyectos de investigación que, propiamente hablando, ningún científico individual llega a comprender. Y no es solo que debamos hablar ya de inteligencia grupal o colaborativa, es que el *Deep-Learning* lleva a los investigadores a conclusiones que funcionan a pesar de que nadie las comprenda con claridad. Estos sistemas de la IA como Deep-Blue, Go+ y EMI muestran incluso una evidente creatividad, en el ajedrez, en la composición musical, pero una creatividad paradójica para el criterio tradicional, pues que sería de tipo mecánico, totalmente ajena a la comprensión (DENNETT, 2017, p. 322). En este sentido, la obsolescencia del humano en la que tantos muestran hoy tanto interés puede significar, de un modo perfectamente coherente, la obsolescencia de la comprensión, que sería al mismo tiempo la pérdida de valor cultural de la filosofía.

III. Teoría de la conciencia.

Así que la teoría evolucionista darwiniana y la teoría memética, comprendidas en su unidad en tanto versando sobre campos en los que se implementa el mismo algoritmo de la adaptación, arrojarían luz sobre los dos problemas fundamentales, estrechamente vinculados entre sí, de la filosofía de la mente, el del significado o la intencionalidad, y el de la conciencia. Pertrechados con las dos, podremos resolver, por fin, el problema central de la relación mente-cuerpo, para lo cual resultará crucial, como vimos, que dejemos de entender la conciencia al modo esencialista, es decir, en el sentido de una facultad una, del tipo *todo o nada*, que se tiene o no se tiene, para pasar a enfocarla al modo continuista. La conciencia sería entonces cuestión de grado, habría menos y más conciencia.

Al decir de sus críticos Dennett sería un eliminativista más, para el que la experiencia consciente no es nada real. Lo cual resulta contradictorio con lo que vimos acerca de la comprensión y su importancia. Pero él mismo se ha obstinado en desmentir este diagnóstico simplificador. Claro está que la conciencia "existe", que "es real", del mismo modo que lo son los colores o la voluntad libre (DENNETT, 2017, p. 222-4). Como ocurre con Nietzsche, la incomprensión radica en que la conciencia no es lo que la mayoría de la gente ha venido pensando que es, una especie de pantalla cinematográfica ante el Ojo de la Mente. Ya desde el comienzo de su carrera, el Dennett (1981) de la postura o la actitud intencional habría mantenido la misma posición intermedia entre el realismo y el interpretacionismo de lo mental, posición que nada tiene que ver con ningún pedestre irrealismo de lo mental.

Si se ha venido creyendo, falsamente, que Dennett niega la realidad de la conciencia es porque lo que sí niega, aparte del Teatro o Cine Cartesiano, es la consiguiente validez de la *auto-fenomenología* como método para su estudio *científico*. Exactamente lo mismo en Nietzsche. La perspectiva cartesiana de primera persona no vale porque nos encadena a una consideración que no puede abandonar la psicología del sentido común, con todos sus artefactos y sus mitos, incapacitándonos así para acceder a una auténtica *ciencia (natural) de la conciencia*. Por eso aboga Dennett por el enfoque hetero-fenomenológico. Porque lo que puedo aprender de las experiencias conscientes del otro a través de sus *palabras* descriptivas es lo mismo que puedo aprender de mis propias experiencias subjetivas, pero con la gran ventaja de que puedo corregirlo y hacerlo avanzar en cuanto a su rendimiento epistémico. Tenemos que recordar en este punto las devastadoras críticas nietzscheanas a la introspección como método de la inicial psicología decimonónica. Y es que, para él, la concienciación es en todo caso un proceso constructivo, es decir,

falsificador. Lo más profundo y real que puedo saber de mí, quiéralo o no, me vendrá siempre de las observaciones ocasionales que los otros que mejor me conocen me hacen a mí mismo sobre mí[25], basándose en lo que digo y hago (NIETZSCHE, 2014, p. 244). Es la introspección, sencillamente, un *unnatural act*, sentencia hoy Dennett (2017, p. 349), y por eso, cuando le pedimos al sujeto que haga "introspección", le ponemos en un brete.

Si le seguimos el rastro a una resolución del tradicional problema de la mente y el cuerpo, resolución que, para él, en el fondo equivale a la superación definitiva del planteamiento dualista inaugurado por Descartes, entonces no podremos pasar por alto el modo tajante en que Dennett, dejándose llevar como de costumbre por su inspiración computacional, afirma que la conciencia es (como) una *ilusión del usuario, y el resultado de una evolución o de un desarrollo*. Una ilusión del usuario igual que lo es una carpeta amarilla para archivos de Windows, por ejemplo, a la que ponemos un nombre y asignamos un lugar en la pantalla del ordenador, una ilusión que le sirve de mucho al usuario del programa. La imagen manifiesta del hombre en el mundo con todo lo que ella contiene, pero en especial con el conjunto de las entidades postuladas por la psicología del sentido común, las creencias, emociones y deseos de los que somos o podemos llegar a ser conscientes, es declarada en su totalidad una *ilusión del usuario*. Pero en el bien entendido de que debemos recogerla, a esta ilusión, como *parte integrante de nuestra ontología*, puesto que se trataría del único acceso de que disponemos a los procesos neuro-computacionales efectivos. Además tenemos que tener en cuenta que, para el "usuario", darle toda la importancia que merece a esta imagen manifiesta del mundo en que existe sería «cuestión de vida o muerte» (DENNETT, 2017, p. 368-70).

En todo esto más bien contrasta Dennett con Nietzsche, quien al final de su carrera lúcida se iba a mostrar bastante más radical a la hora de rebajar la realidad efectiva de los procesos conscientes, al considerarla tan solo una efectividad "de quinto orden". Por mucho que nos habiliten, como reconoce de buen grado el filósofo alemán, para vivir en un mundo humano donde nuestra relativa persistencia sea posible. En cualquier caso, como iba a señalar Habermas (1968a y b), Nietzsche es un gran virtuoso de la reflexión consciente que se dedica a aplicar su destreza en la negación inmisericorde de las posibilidades epistémicas de la reflexión consciente. Es decir, el pensamiento nietzscheano revela un manejo verdaderamente magistral de todos los registros y matices de la hoy llamada *teoría de la mente* (o bien *mind-reading*), al mismo tiempo que emite declaraciones que

25 *Cf*. sobre este importante extremo Rodríguez González (2018, p. 33-46).

exhiben su autoconciencia de los límites de la actitud intencional en cuanto a su rendimiento cognoscitivo.

Dennett no llega a prodigarse tanto como el pensador alemán en este ejercicio en apariencia tan paradójico, si bien, al igual que Nietzsche, apuntará a otro ámbito diferente del consciente como hogar propio de la más efectiva causación[26]: el ámbito de los circuitos cerebrales, en su caso, y por supuesto computacionalmente concebidos. Por eso nos asegura el americano que la única manera de desarrollar una genuina ciencia natural de la conciencia, o sea, una que por fin consiga trascender la imagen manifiesta, es mediante el establecimiento de correspondencias entre contenidos mentales, o significados, y estructuras y eventos informacionales de carácter subpersonal. Estas estructuras y eventos subpersonales serían los responsables causales de los tan ricos detalles de la ilusión de usuario en que nosotros vivimos (DENNETT, 2017, p. 367). Pero los detalles con contenido mental de la ilusión de la conciencia en el "usuario" serían, como hemos dicho, los únicos instrumentos que tenemos para aislar las estructuras subpersonales, y así llegar hasta ellas por vía científica.

Claro está que esta solución dennettiana del problema parece valer sólo para la conciencia tenida por "fácil", conciencia de acceso o en definitiva funcional, pero no para la difícil o dura, la conciencia fenoménica. Una objeción que para Dennett no supone sino el intento de reinstaurar el dualismo esgrimiendo la cuestión de los *qualia*, el *what it is like* famoso del murciélago de Nagel. Como sabemos, ningún autor contemporáneo de la filosofía de la mente habría luchado con mayor encono que Dennett, pero asimismo con más sentido del humor, contra esas "dulces criaturas de ensueño", los *qualia inefables e inanalizables,* cuya persistencia lo único que atestiguaría, a su juicio, es la potencia del «pellizco de la gravedad cartesiana» (DENNETT, 2017, p. 338), o sea, lo difícil que resulta, incluso hoy, salirse de la órbita del planteamiento dualista o materialista *cartesianos*. Serían los *qualia* meros artefactos de la imaginación, lo que contrasta con el partido metafísico u oscurantista que se les ha pretendido sacar. Serían propiamente, se nos asegura, *proyecciones subjetivas en la imagen manifiesta de las reacciones del cerebro a los estímulos objetivos*. Unas "proyecciones" consistentes en la confirmación de las expectativas bayesianas del cerebro, expectativas que son subjetivas, claro, pero solo en el sentido de que dependen de las reacciones de nuestro sistema nervioso.

26 Aunque hay que tener en cuenta que en la ontología dennettiana se aceptan diferentes niveles de realidad, relativizados convenientemente a las posturas posibles del observador, con lo que se defiende la virtud productiva de los procesos mentales reconocidos en la postura intencional. Desde ella se seleccionarían *patrones reales* de información (THOMPSON, 2009).

Aquí ocurre lo mismo que con nuestra supuesta experiencia de la causalidad como relación real y objetiva de producción o potencia, esa que iba a desmontar Hume en su otra "extraña inversión del razonamiento", anterior a las de Darwin y Turing. Lo que hizo el filósofo escocés con la relación causal lo quiere hacer Dennett ahora con cosas tan ontológicamente "enigmáticas" como la dulzura del azúcar o la ternura del bebé[27] (DENNETT, 2017, p. 356). En definitiva, estos *qualia* deberán ser «puestos entre paréntesis» sin contemplaciones si lo que queremos es pasar de un tratamiento de la conciencia en la imagen manifiesta a su estudio científico (DENNETT, 2017, p. 365).

En la época de Nietzsche también se planteaba este problema como el *problema de la sensación*. Un reto que algunos juzgaban imposible de abordar ya que daría fe del límite superior del conocimiento humano (DU BOIS REYMOND, 1884). Mientras que a veces el filósofo alemán se limitaba a apuntar con displicencia que a él este problema no le quitaba el sueño, otras, en cambio, divagaba con el extravagante pan-psiquismo como única solución viable (PARKES, 1994). Una posición que por supuesto ha vuelto a aparecer en nuestros días.

En cualquier caso, para Nietzsche, como después para Dennett, el problema *real* de la conciencia no sería éste de las sensaciones y los sentimientos construidos como artefactos de la imaginación dualista, sino el de la conciencia específicamente humana, lingüística o de contenido conceptualmente articulado, si lo queremos poner en la expresión de Katsafanas (2005).

Desmintiendo de nuevo a Descartes, tanto Nietzsche como Dennett subrayan que la conciencia específicamente humana es algo *devenido*, surgido en concreto a partir de la necesidad de la comunicación, que sería tan acuciante para la especie animal más vulnerable. La «heurística nietzscheana de la necesidad», en expresión de Stegmaier (2012), aplicada a la génesis de nuestra conciencia lingüística, habría sido hoy del todo confirmada por la siguiente conclusión de la psicología cognitiva, que Dennett gusta por su parte de resaltar: *la comunicación es la única conducta que exige la auto-monitorización del pro-*

27 Consideremos la ilusión cromática denominada *the neon-colour spreading phenomenon*. Hay quien sigue insistiendo en que, por mucho que esté de acuerdo en que el anillo azul que se forma en el centro de la imagen es solo una apariencia, porque sabemos que no hay allí nada realmente azul, y mucho menos un anillo, todavía tenemos que preguntarnos en qué consiste esa apariencia recalcitrante, porque se trata de algo que *parece ser* de manera muy real. Lo que le responde Dennett es ni más ni menos que ciertamente se trata *solo* de una apariencia, y entonces, por definición, no es nada real. Es decir, no hay ninguna diferencia entre parecer azul y parecer *realmente* azul. El anillo azul es una apariencia y punto. Cfr. la conversación imaginaria de Dennett con el qualófilo Otto en Boden (2016, p. 129-31).

pio sistema de control (DENNETT, 2017, p. 390). La conciencia nos sirve, la conciencia nos es necesaria, porque estamos obligados a comunicarnos con el otro humano para poder vivir. Tenemos que cooperar con el otro lo queramos o no, necesitamos de él, pero para conseguir su colaboración hay que hacerle saber cómo estamos, qué nos ocurre. Y por lo tanto tenemos que saber nosotros antes qué nos pasa, qué necesitamos exactamente: era forzoso haber llegado a ser animales auto-conscientes. Nietzsche y Dennett coinciden por tanto en afirmar que no tendríamos conciencia si no la necesitáramos[28].

Pero, además, de esa misma necesidad vital, la misma urgencia de la vida social, procedería asimismo la *sensación del yo*: nos es preciso aprender a *administrar nuestra transparencia*, o sea, hay que saber distinguir aquellos pensamientos que podemos comunicar sin correr peligro de aquellos que no. Así se iría trazando cada vez con más nitidez la frontera sentida entre el yo y los otros, a partir de la reconocida importancia que tiene para nosotros «no ser (como) patos sentados», tan fáciles de matar. Es la necesidad de minimizar el peligro que entraña el otro para nosotros, para poder compaginarlo con la acuciante necesidad que sin embargo tendríamos de su colaboración y de su ayuda, aquello que va a explicar el hecho de que una de las actividades más propiamente humanas sea la de dar explicaciones, al otro y a nosotros mismos, algo que hacemos de manera incesante, a cada paso que damos. Introducirse en el espacio de las razones sería exactamente lo mismo, desde este punto de vista social que es el esencial para el tema de la conciencia humana o lingüística, que constituirse como animal auto-consciente.

Además de esto, sorprende el parecido con el tratamiento nietzscheano del tema de la voluntad (RODRÍGUEZ GONZÁLEZ, 2018, p. 66-7) de lo que Dennett nos dice acerca de nuestra experiencia consciente de la voluntad, al hilo de su lectura del conocido estudio de Wegner (2002, p. 96), del que no nos resistimos a citar lo siguiente: «The experience of will, then, is the way our minds portray their operations to us, not their actual operation». La verdadera causación discurre, entonces, a otro nivel diferente del de la voluntad consciente, el nivel de los circuitos neuronales y las partes sub-personales de nuestro pensar[29], si lo decimos con Dennett; o el nivel de las pulsiones inconscientes que integrarían el cuerpo nietzscheano, esa «inteligencia más elevada» o más abarcadora.

28 «If we didn't have to be able to talk to each other about our current thoughts and projects, and our memories of how things were, and so forth, our brains wouldn't waste the time, energy, and gray matter on an edited digest of current activities, which is what our stream of consciousness is» (DENNETT, 2017, 344).

29 «All this subpersonal, neural-level activity is where the actual causal interactions happen that provide your cognitive powers, but all 'you' have access to is the results» (DENNETT 2017, 348).

Y entonces, la amenaza del epifenomenalismo, que parece volver a caer sobre nosotros, y con él la del irrealismo de lo mental, podría quedar conjurada con la tesis de Nietzsche de la encarnación o corporalización (*Einverleibung*) de las creencias y los deseos, su hacerse instintivos o convertirse en nuestra carne y nuestra sangre: circulación "dionisíaca", se podría decir, entre fondo y superficie.

En fin, en vez de centrarnos en el Yo como Gran Significador o Constituyente del Sentido, ese Yo director y único espectador del Teatro Cartesiano, tendríamos que centrarnos ahora en la nueva visión de la subjetividad como estructura dominante de pulsiones, por eso devenidas conscientes en cuanto resultado final y momentáneo, en el caso de Nietzsche; o bien tendríamos que centrarnos en las «agencias menores distribuidas por el cerebro». Es decir, en los que Dennett (2017, p. 353) denomina *subpersonal blurts* que esbozarían sin cesar narraciones parciales que son revisadas y vueltas a narrar en competencia las unas con las otras[30]. Vamos contando cuentos biográficos hasta que por fin emergería algo así como un yo-personaje o incluso protagonista, una especie de yo en cualquier caso provisional, transitorio, resultado ocasional de esas narraciones en lucha, un yo que nos sirve de presentación social, ante los otros y ante nosotros mismos: el organismo humano genera su Yo igual que *la araña teje su tela*[31].

Con todo esto piensa Dennett que habríamos conseguido cerrar finalmente la brecha cartesiana, porque si ya habíamos mostrado cómo son posibles las competencias sin comprensión, y luego la emergencia de la comprensión misma, ahora podemos añadir que ya sabemos cuál sería la función biológica y social de la conciencia y de la actitud intencional, dando así respuesta a la pregunta nietzscheana, formulada en *La gaya ciencia* 354, sobre el enigmático para qué de la conciencia.

IV. Conclusión.

Se puede afirmar que todo lo hasta aquí expuesto *resuelve* el problema mente-cuerpo en tanto problema *filosófico*, en el sentido de que *cerraría por fin la* «*herida dualista*». Porque no cabe duda de que el naturalismo continuista y gradualista de Nietzsche y Den-

30 "Blurts": del verbo inglés *to blurt out*: hablar sin pensar: convertir compulsivamente contenidos de pensamiento en palabras.

31 La imagen de la araña que teje su tela expresa la necesidad con que se produce o se hace algo, impremeditadamente, y así la usan tanto Nietzsche (1873) como Dennett (1991, Pos 7035).

nett se *puede comprender*[32], lo que para nosotros significa que, desde esta concepción del antiesencialismo, el ser humano habría dejado de ser un "milagro" en el universo físico. Por este flanco concreto de la comprensión naturalizada, por lo tanto, cesaría esa perplejidad o "admiración" que siempre provocaba la relación psicofísica como misterio filosófico paradigmático o, en suma, problema no resoluble. Definitivamente, una posición naturalista como la que hemos desarrollado sí que se puede comprender, sin tener que enredarnos en la sospecha de que solo *creeríamos* comprenderla, en contra de lo que han afirmado del neodarwinismo autores como Fodor, Nagel y por supuesto Plantinga[33].

La investigación científica actual apoyaría sin duda esta idea de que ya no hay problema de la mente y el cuerpo *en el sentido filosófico del término "problema"*. Si nos centramos por ejemplo en el tradicional enigma del carácter unificado el campo de la conciencia, el "irresoluble" problema de la conciencia en tanto *una* conciencia centralizada y coherente—en definitiva, el célebre *binding problem*[34]—, veremos que éste se habría solventado hoy por vía experimental con el descubrimiento de que la percepción consciente tiene lugar, a diferencia del procesamiento inconsciente de los estímulos visuales, que por supuesto es asimismo algo real, justo en el momento en que ciertas áreas ampliamente distribuidas del córtex cerebral se involucran momentáneamente en oscilaciones sincronizadas de alta frecuencia. De modo que esa sincronización vendría a ser el equivalente fisiológico de la unificación semántica, explicándose así el hecho de que los contenidos de conciencia *solo* puedan ser coherentes. Con lo que se va a poder aislar la evidencia causal necesaria para consolidar una relación bien determinada entre la actividad neuronal y los contenidos de la conciencia fenoménica.

Naturalmente, con explicaciones como ésta se evidencia que lo que aún nos quedaría por resolver es el problema *científico* de la conciencia, o incluso muchos problemas difíciles y confusos relativos a ella y sus *mecanismos relevantes*. Pero, en cualquier caso,

---

32 Es muy importante para el objetivo de este trabajo recordar, en este punto, la afirmación wittgensteiniana de que «la *verdad* de mis afirmaciones es la prueba de mi comprensión de esas afirmaciones». Solo comprendo algo, verdaderamente, si ese algo es verdadero, porque si es falso entonces no lo puedo comprender. Como mínimo, «si hago ciertas afirmaciones falsas, se hace dudoso que las entienda» (WITTGENSTEIN, 1974, §§ 80-81, 12). Por lo demás, aquello que cuenta como prueba adecuada de una afirmación «pertenece a la lógica», o sea, a la descripción del juego de lenguaje. Nietzsche, por su parte, habría decidido en un momento de su madurez entregarse filosóficamente solo a lo que para él fuese posible comprender.

33 Es impresionante el modo en que Dennett (2009; p. 10062-3) hace trizas el argumento básico del doblemente creacionista Plantinga.

34 Cuyo carácter inexplicable al premio nobel John Eccles le hacía apostar muy fuerte por la tesis del dualismo psicofísico a la altura de los setenta del siglo pasado.

ninguno de ellos tendría carácter filosófico. Sobre todo, está claro, llegar a revelar el modo preciso en que los patrones de activación neuronal dan surgimiento a sentimientos subjetivos, ese es el objetivo de la ciencia de la conciencia. Casi seguro que esta cuestión permanecerá embrollada durante mucho tiempo todavía[35]. Pero si eliminamos todo lo relativo a la ilusión de los *qualia,* como entidades auténticamente metafísicas, átomos de materia mental, entonces no seremos forzados a ver en este problema científico ningún enigma insalvable, o misterio filosófico, sino un asunto difícil más a resolver por la tecnociencia.

Tampoco haría falta ponerse del lado del Husserl de *La crisis de las ciencias europeas* para conceder que la imagen científica del hombre en el mundo se basa en la imagen manifiesta, habiéndose desarrollado precisamente para que podamos manejarnos mejor con ésta. Pero esto no quita la verdad materialista, hoy atestiguada desde todos los frentes de la investigación, de que la imagen de la experiencia subjetiva y personal es hecha posible y generada, es *producida* y mantenida, en cualquier caso, por la actividad neurofisiológica del sistema nervioso del animal humano como integrante de la imagen científica. *Es la conciencia algo que el cerebro hace porque resulta una necesidad biológica*, y si debe ser incluida en nuestra ontología es precisamente por esa razón, porque la hace el cerebro. A este respecto iba a mostrar Nietzsche al final sus reticencias. Muy significativo es este texto suyo:

> También sobre esto nosotros hemos reflexionado mejor: el llegar a ser conscientes, el 'espíritu', nosotros lo consideramos precisamente como síntoma de una relativa imperfección del organismo, como un ensayar, un tantear, un desacertar, como una fatigosa labor en la que innecesariamente se consume mucha energía nerviosa, —nosotros negamos que se pueda hacer cualquier cosa perfecta mientras se la haga todavía de manera consciente. El 'espíritu puro' es una pura estupidez: si descontamos el sistema nervioso y los sentidos, la 'envoltura mortal', *entonces cometemos un error de cálculo*--¡y nada más!... (NIETZSCHE, 2016, p. 714-5).

De lo que se trataría, en cambio, en el caso de Sellars (1963b), es de probar a incorporar la imagen científica en nuestra concepción de nosotros mismos como personas, o sea, como animales con intenciones y lenguaje. Una concepción personal a su juicio ineludible, que completará la imagen científica pero no reconciliándose con ella, pues pretenderlo carece de sentido, sino simplemente uniéndose a ella.

35 Tal es la fundamentada previsión del neurocientífico W. Singer en su conversación con Th. Metzinger (2009).

Este naturalismo evolutivo sí que se puede comprender, pero el sobrenaturalismo del Diseño Inteligente mucho menos, porque la presunta explicación teológica es más oscura en sí misma que lo que se pretende aclarar con ella. Lo que quiso Nietzsche (2016, p. 390-2), ya sabemos, fue «retrotraducir» el humano a la naturaleza, fue naturalizarnos completamente a los hombres en un cierto sentido, para que ya no volviésemos a caer en la tentación de hacer caso a los «pájaros cantores» que nos siguen susurrando al oído que somos de otra naturaleza y de otro origen diferentes del terreno. La mística peculiar de este naturalismo nietzscheano es la antiquísima del dionisismo[36]. Y por nuestra parte somos de la convicción, siguiendo a Nietzsche, de que la asimilación de la imagen científica destruye inevitablemente el moralismo judeocristiano, restituyéndonos con ello el plano de la pura inmanencia. Ser fieles al sentido de la Tierra, en una palabra, con la ayuda del método científico, que esencialmente es para Nietzsche el arte del bien leer (la virtud de la *Redlichkeit*)[37].

Referencias

ABEL, G. (2001): "Bewusstsein-Sprache-Natur: Nietzsches Philosophie des Geistes", **Nietzsche-Studien** 30, 1-43, pp. 1-3. En inglés con revisiones: "Consciousness, Language, and Nature. Nietzsche's Philosophy of Mind and Nature". In: M. DRIES & P.J.E. KAIL (Eds.). **Nietzsche on Mind & Nature**. Oxford University Press, 2015, p. 142-163.

ANDRESEN, J. Nietzsche *contra* Dennett. **The Journal of Nietzsche Studies**, Volume 46, Issue 1, p. 120-131, 2015.

BODEN, M. **AI:** Its nature and future. Oxford: Oxford University Press, 2016. [Después se cambiaría este título por el de *AI: A very short introduction*].

36 Hay también una mística en Dennett y en Dawkins, como se puede comprobar en su conversación sobre el evolucionismo, recogida en un vídeo sobre el genio de Charles Darwin: https://www.youtube.com/watch?v=5lfTPTFN94.

37 En este asunto del bien leer y el sentido de los hechos es esencial el parágrafo 52 de *El anticristo*.

DAWKINS, R. **El gen egoísta**. Las bases biológicas de nuestra conducta. Barcelona: Labor, 1979.

DEACON, T. W. **Incomplete Nature***:* How mind emerged from matter. New York / London: W.W. Norton & Company, 2011.

DENNETT, D. C. **Consciousness explained**. New York, Little & Brown and Company, 1991.

DENNETT, D. C. **Darwin's dangerous idea**. Evolution and the meanings of life. New York: Penguin Books, 1995.

DENNETT, D. C. True Believers. In: HAUGELAND, J. (Ed.): **Mind Design II**, Cambridge, Massachusetts, The MIT Press, 1997, pp. 57-79 [1981].

DENNETT, D. C. Darwin's 'Strange Inversion of Reasoning'. **Proceedings of National Academy of Sciences** vol. 106, suppl.1, June 16, pp. 10061-10065, 2009.

DENNETT, D. C. **From bacteria to Bach and back**. The evolution of minds. Nueva York y Londres: W.W. Norton & Company; Penguin Books, Kindle Edition, 2017.

Du BOIS REYMOND, É. Über die Grenzen des Naturerkennens: die sieben Welträthsel. Leipzig, Von Veit & Comp., 1884.

FEIGL, H. **The 'Mental' and the 'Physical'**. The essay and a postcript. Minneapolis – Minnesota: University of Minnesota Press, 1967.

KATSAFANAS, P. Nietzsche's Theory of Mind. Consciousness and Conceptualization. **European Journal of Philosophy**, 13: 1, pp. 1-31, 2005.

HABERMAS, J. **Conocimiento e interés.** Madrid, Taurus, 1982 [1968b].

HABERMAS, J. **La lógica de las ciencias sociales**. Madrid, Tecnos, 1988 [1968a].

METZINGER, Th. **The ego tunnel**. The science of the mind and the myth of the self. New York, Basic Books, 2009.

MINSKY, M. **The society of mind.** New York/London, Simon & Schuster, 1988 [1986].

NIETZSCHE, F. **Fragmentos póstumos II**. Madrid, Tecnos, 2008.

NIETZSCHE, F. **Fragmentos póstumos III**. Madrid, Tecnos, 2010.

NIETZSCHE, F. **Obras completas III**, I. Madrid, Tecnos, 2014.

NIETZSCHE, F. **Obras completas IV**, II. Madrid, Tecnos, 2016.

PARKES, G. **Composing the soul.** Reaches of Nietzsche's psychology. Chicago: The University of Chicago Press, 1994.

POPPER, K. R., y ECCLES, J. **El yo y su cerebro**. Barcelona, Labor, 1980.
POPPER, K. R. **Conocimiento objetivo.** Un enfoque evolucionista. Madrid, Tecnos, 1974 [1972].

POPPER, K. R. **El cuerpo y la mente**. Barcelona, Paidós, 1997.

RATHKOPF, Ch. A. Mental Evolution: A Review of Daniel Dennett's *From Bacteria to Bach and Back*. **Biol Philos**, 32: 1355-1368, 2017.

RODRÍGUEZ GONZÁLEZ, M. Popper, mente y cultura, In PERONA A., Jiménez, (ed.): **Contrastando a Popper**. Madrid: Biblioteca Nueva, 2008, p. 197-226.

RODRÍGUEZ GONZÁLEZ, M. **Más allá del rebaño**. Nietzsche, filósofo de la mente. Madrid, Avarigani, 2018.

RORTY, R. **Philosophy and the mirror of nature**. Princeton, New Jersey: Princeton University Press, 1979.

SEARLE, J. R. ¿Pueden los ordenadores pensar? In: SEARLE, J. R. **Mentes, cerebros y ciencia**, Madrid, Cátedra, 1985. [*The 1984 Reid Lecture*]

SELLARS, W. Empiricism and the Philosophy of Mind, In: SELLARS, W. **Science, perception and reality**. London, Routledge and Kegan Paul, 1963a. p. 127-197.

SELLARS, W. Philosophy and the Scientific Image of Man. In: SELLARS, W. **Science, perception and reality**. London, Routledge and Kegan Paul, 1963b, p. 1-41.

STEGMAIER, W. **Nietzsches Befreiung der Philosophie.** Kontextuelle Interpretation des V. Busch der Fröhlichen Wissenschaft. Berlín / Boston, De Gruyter, 2012.
THOMPSON, D. L. **Daniel Dennett.** Contemporary american thinkers. New York-London, Continuum International Publishing, 2009.

WEGNER, D.M. **The illusion of conscious will**. Cambridge Ma.: The MIT Press, 2002.

WITTGENSTEIN, L. **On Certainty/Über Gewissheit**. Anscombe, G.E.M. & von Wright, G.H. (Eds.). Oxford, Basil Blackwell, 1974.

WITTGENSTEIN, L. **Investigaciones filosóficas**. Barcelona, México, UNAM/ Crítica, 1988.

A CONTRIBUIÇÃO DE SCHOPENHAUER PARA A IDEIA DE "MENTE CORPORIFICADA"

André Henrique Mendes Viana de Oliveira

Resumo:

Neste artigo pretendemos mostrar como a filosofia de Schopenhauer traz contribuições para a ideia de *mente corporificada*. A princípio, apresentamos a crítica de Schopenhauer ao dualismo, que concebe mente e corpo como realidades distintas. Em seguida, indicamos como sua filosofia sugere uma materialização do sujeito transcendental. Por fim, discutimos como a noção de mente integrada ao corpo pode abrir um novo horizonte temático ao possibilitar a relação entre a filosofia da mente e o campo da filosofia prática.

Palavras-chave:

Corpo; intelecto; mente corporificada; sujeito transcendental; vontade.

Abstract:

In this paper we intend to show how Schopenhauer's philosophy brings contributions to the ideia of a *embodied mind*. At first, we present Schopenhauer's criticism to the dualistic thesis, which conceives mind and body as different realities, then we point out how his philosophy suggests a materialization of the transcendental subject. Finally, we discuss how the notion of mind integrated to body can opens a new theoretic horizon by means of allowing to relate philosophy of mind and the pratical philosophy.

Keywords:

Body, intellect, embodied mind, transcendental subjetc, will.

Introdução

A problemática aberta com a postulação do *cogito* cartesiano é, muitas vezes, indicada como a mais importante raiz moderna da investigação filosófica a respeito da mente. Em que pese as especulações da filosofia clássica (e também da medieval) acerca da alma (*ψυχή*), as investigações de cunho empírico-experimental que floresceram na modernidade parecem ter aberto e fortalecido a via que legou ao mundo contemporâneo a concepção de que a *subjetividade* não seria um fenômeno de caráter transcendente, mas, sim, um

produto da *mente*, cujas propriedades constituiriam um conjunto complexo de processos intrinsecamente ligados ao corpo[1].

Uma vez estabelecida essa concepção, o problema central passaria a ser o de caracterizar com mais consistência qual a natureza desse fenômeno a que chamamos *mente*, e qual a natureza da relação desta com o corpo. Podemos dizer que, por muito tempo, essas questões foram marcadas por uma contraposição ontológica entre o que é material (o corpo) e o que é imaterial (a alma, a mente, o pensamento); contraposição que também tem em Descartes uma de suas raízes mais fortes. Quando, por exemplo, o filósofo francês escreve: "compreendi por aí que era uma substância pensante cuja essência ou natureza consiste apenas no pensar, e que, para ser, não necessita de nenhum lugar, nem depende de qualquer coisa material" (DESCARTES, 1987, p. 46), notamos que a postulação desse dualismo sugere a necessidade de se explicar como ocorre a interação entre o que é material e o que é (supostamente) imaterial.

Ainda na modernidade, porém, essa dicotomia corpo-alma passaria a ser amplamente questionada, na medida em que se encontravam indícios científicos sobre a importância do corpo (organismo) para os chamados processos mentais; e seria mesmo uma injustiça não reconhecermos que Descartes, ainda que de modo incipiente, tentou pensar uma unidade entre corpo e alma no quadro de sua filosofia. Jaquet (2011) faz notar isso de forma clara em seu livro *A unidade do corpo e da mente: afetos, ações e paixões em Espinosa*, atribuindo a Descartes a tentativa de elaborar um discurso psicofísico que pensasse corpo e alma como um todo integrado. De fato, no artigo 34 de *As paixões da alma,* podemos ler a seguinte afirmação de Descartes sobre a glândula pineal:

> Concebamos, pois, que a alma tem a sua sede principal na pequena glândula que existe no meio do cérebro, de onde irradia para todo o resto do corpo, por intermédio dos espíritos, dos nervos e mesmo do sangue, que, participando das impressões dos espíritos, podem levá-los pelas artérias a todos os membros (DESCARTES, 1987, p. 230).

No entanto, ainda conforme Jaquet (2011), a ideia de que a união entre corpo e alma dependesse prioritariamente de uma pequena porção de matéria (a glândula pineal), que seria a sede da alma[2], tornou-se uma ideia alvo da crítica de Espinosa e de muitos

1 A este respeito é pertinente ver, por exemplo, o verbete "filosofia da mente" em ABBAGNANO, 2007, p. 762.

2 O que hoje sabemos sobre a glândula pineal é que ela faz parte do nosso sistema endócrino e se encontra na parte central do cérebro, sendo atribuídas a ela algumas funções importantes, como a produção de melatonina e a regulação do ritmo circadiano (ciclo sono-vigília).

outros filósofos que lhe foram posteriores ao longo de todo o período moderno (vindo, a mesma crítica, a reverberar na filosofia contemporânea).

Contudo, nos parece particularmente digno de nota que o nome de Arthur Schopenhauer quase não apareça nos estudos filosóficos sobre essa temática. Devido a isso é que propomos abordar, neste artigo, o problema da relação entre mente e corpo, a partir da filosofia de Schopenhauer, cuja obra nos parece repleta de contribuições relevantes para a noção de *mente corporificada*; formulação importante para o quadro que compõe a discussão contemporânea em filosofia da mente.

O problema mente-corpo: de Schopenhauer à neurociência

A crítica de Schopenhauer tem como alvo central a tese do dualismo de substâncias, já que foi com base nela que Descartes instituiu a suposta diferença radical entre substância pensante (*res cogitans*) e substância corpórea ou material (*res extensa*). O filósofo alemão destaca no primeiro volume de *Parerga e Paralipomena*, especificamente no § 12 do capítulo "Fragmentos para a história da Filosofia", que os sistemas filosóficos da filosofia moderna "são cálculos que não funcionam (*aufgehn*): deixam um resto, ou se se prefere um exemplo da química, um sedimento insolúvel"[3] (SCHOPENHAUER, 2009, p. 101). Tal sedimento insolúvel do sistema cartesiano teria surgido justamente da hipóstase de um dualismo substancial:

> Ele [Descartes], contudo, admitiu dois tipos de substância: a pensante e a extensa. Estas deveriam ser mutuamente afetadas por *influxus physicus*, o que logo se revelou como seu resto. Tal influxo, de fato, se produziria não só de fora para dentro, na representação do mundo corpóreo, mas também de dentro para fora, entre a vontade (que sem nenhum cuidado foi adicionada ao pensamento) e as ações corporais. A relação íntima entre esses dois tipos de substância tornou-se o principal problema (SCHOPENHAUER, 2009, p. 102).

Além de censurar o caráter não-científico das ideias de alma imaterial e de substância pensante, Schopenhauer considera-as fruto de um paralogismo da razão dialética. Da descoberta de um "Eu" que se manteria fixo em meio à passagem e variação das múltiplas representações do sujeito, o teria hipostasiado como uma substância, já que se via ali o atributo da permanência, tal como na matéria. E, como se trataria de algo não extenso, essa substância jamais poderia ser dada no espaço, do que teriam concluído sua não espacialidade e sua não temporalidade, posto que no tempo não há permanência.

3 Os trechos da obra *Parerga* e *Paralipomena* citados neste artigo foram traduzidos por nós a partir de duas edições: a edição original em alemão, *Sämtliche Werke in fünf Bänden*, e a edição em espanhol indicada nas referências bibliográficas.

Importante lembrar que, segundo a doutrina de Schopenhauer, aquele "Eu", sujeito cognoscente, é um subproduto da Vontade, e ela se objetiva no organismo, no qual brota o cérebro, com todas as suas funções e disposições cognitivas. Essa tese de Schopenhauer revela a total dependência do sujeito do conhecimento em relação à constituição física do organismo. Nas palavras do próprio filósofo:

> Em meu pensamento, o sujeito do conhecer, como o corpo no qual ele se apresenta objetivamente como função cerebral (*Gehirnfunktion*), é fenômeno da vontade, a qual, enquanto única coisa em si, constitui aqui o substrato do correlato de todos os fenômenos, isto é, do sujeito do conhecimento (SCHOPENHAUER, 2009, p. 133).

Já no segundo volume de sua principal obra, *O mundo como vontade e como representação,* vemos o filósofo afirmar, por exemplo, que "a cada um a própria experiência demonstrou abundantemente a contínua e total dependência da consciência que conhece, do cérebro, e é mais fácil acreditar numa digestão sem estômago, que numa consciência sem cérebro" (SCHOPENHAUER, 2015, p. 241). Assim, a concepção schopenhaueriana de *sujeito cognoscente* se direciona de modo explícito para uma *materialização do transcendental*[4]; aquilo que entendemos como *a mente* possui um profundo enraizamento em nossa estrutura orgânica, e não pode ser tomada como independente do aporte material que é o próprio organismo onde ela se insere e funciona.

Como consequência disso, surge a tese de que nosso conhecimento do mundo possui um alcance não só limitado (em termos epistêmicos) como direcionado (em termos pragmáticos), pois o nosso corpo, e o intelecto a ele ligado, é o que determina as condições de possibilidade de conhecimento do mundo. De um ponto de vista objetivo, segundo Schopenhauer, isso ocorre porque o cérebro é dependente das funções mais vitais, o que o colocaria numa situação subalterna em relação ao restante do organismo. Assim, o cérebro,

4 A direção que Schopenhauer toma a partir da filosofia crítica de Kant fica explícita em várias passagens de sua obra. Por exemplo, no texto *Crítica da filosofia kantiana* (apêndice de *O mundo como vontade e como representação*), Schopenhauer assim direciona o criticismo contra a chamada "metafísica dogmática", indicando, ao fim da passagem, aquilo que aqui compreendemos como *materialização do transcendental*: "o fato de que a filosofia crítica, para chegar a esse resultado, teve de ir além das *veritates aeternae* sobre as quais estava baseado todo o dogmatismo de até então, e assim fazer de tais verdades mesmas o objeto de sua investigação, tornou-se filosofia *transcendental*. Daí resulta, ademais, que o mundo objetivo, como o conhecemos, não pertence à essência das coisas em si mesmas, mas é seu mero fenômeno, condicionado exatamente por aquelas mesmas formas que se encontram *a priori* no intelecto humano (isto é, o cérebro)" (SCHOPENHAUER, 2005, p. 530-1).

> junto com os nervos e a medula espinhal a ele anexados, é um mero fruto, um produto, sim, em verdade um parasita do restante do organismo, na medida em que não intervém (*eingreift*) diretamente em sua maquinaria interna, mas tão somente (*blo**ß**e*) serve ao fim (*Zweck*) da autoconservação, regulando a relação do organismo com o mundo exterior (SCHOPENHAUER, 2015, p. 243).

A submissão dos processos e mecanismos da racionalidade a uma base biológica, além de revelar a limitação natural do conhecimento humano, porquanto só nos é possível conhecer o mundo segundo os valores e necessidades do nosso organismo, mostra que é preciso entendermos a racionalidade como algo essencialmente integrado ao funcionamento do corpo. Ou seja, ela não consiste num processo desvinculado das outras funções, por mais distantes que estas pareçam estar. Nenhum processo consciente ou inconsciente prescinde das atividades que ocorrem nos níveis mais básicos do organismo. Morgenstern (2013, p. 27) nos lembra que essa tese do pensador alemão "alcançou um grande significado no desenvolvimento posterior da filosofia moderna, especialmente com Nietzsche e Bergson, mas também no pragmatismo e na teoria evolutiva do conhecimento"[5], reiterando a atualidade do pensamento de Schopenhauer para o tema, como propomos.

Por outro lado, Birnbacher (2005) destaca algo que pode nos servir para uma crítica até mesmo da própria concepção schopenhaueriana a respeito do organismo, mais especificamente a respeito da interação entre os níveis mais básicos do organismo e aqueles processos que atribuímos aos mecanismos mais complexos da atividade cerebral, pois é preciso lembrar que, apesar de submeter tanto os processos conscientes quanto os inconscientes ao pulso da vontade (que se manifesta no corpo como um todo), Schopenhauer não acredita que o cérebro tome parte sobre as atividades que ocorrem fora de uma dimensão consciente. Para ele, os atos do corpo que não são *motivados*[6], ou seja, que não são dependentes de representações, ocorrem sem influência do cérebro. Tais atos, escreve ele, "seguem-se imediatamente dos estímulos, a maioria internos, e constituem os movimentos reflexos que partem da mera medula espinhal, como os espasmos e convulsões, nos quais a vontade faz efeito sem a participação do cérebro" (SCHOPENHAUER, 2015, p. 305).

---

5 As traduções do texto de Morgenstern são de nossa responsabilidade.

6 Para Schopenhauer, todos os eventos do mundo são conhecidos como representações de um sujeito, e todas essas representações estão submetidas a alguma das formas que regem a causalidade fenomênica. Assim, as ocorrências físicas estão submetidas à causalidade material; os juízos e raciocínios estão submetidos às razões (que operam neles como causas); as mudanças de ordem vegetativa estão submetidas às excitações; e toda a dimensão do agir humano é regida por *motivos*, sendo "motivo" definido como "um estímulo externo, cuja ocasião gera uma *imagem no cérebro* (*Bild im Gehirn*), sob cuja mediação a vontade realiza o efeito propriamente dito, a ação corporal." (SCHOPENHAUER, 2013, p. 68, grifo do autor). Neste sentido, Schopenhauer nomeia a causalidade que rege as ações humanas de *lei de motivação*.

Porém, como Birnbacher (2005, p. 142) sublinha, já ficou demonstrado pela moderna neurologia que algumas decisões "que são tomadas sem cálculo racional e, por assim dizer, 'a partir do ventre', muitas vezes se mostram (até) como as 'melhores decisões'. Nessas decisões 'instintivas' determinadas partes do cérebro estão completamente envolvidas."[7] O que isto indica é que entre os níveis mais básicos e os mais complexos do organismo há uma integração que Schopenhauer parece não ter considerado devidamente.

Essas constatações comentadas por Birnbacher advém do estudo de casos com pacientes que sofreram lesões cerebrais que danificaram sua capacidade de perceber um "sentimento intuitivo" capaz de direcionar suas ações de modo mais razoável e favorável para si. Em determinadas situações, estes pacientes tomam "decisões mais desfavoráveis para si do que as pessoas normais, não *embora*, mas *porque* decidem exclusivamente de forma racional. A eles falta o 'sentimento intuitivo', de que eles estão tomando a decisão errada" (BIRNBACHER, 2005, p. 142).

Por "decisões razoáveis" entenda-se aqui simplesmente aquelas que contribuem para a integridade global do organismo, isto é, aquelas que não tendam a prejudicá-lo em qualquer sentido. Assim, dado que tanto a racionalidade quanto as determinações instintivas trabalhem para a conservação e integridade do organismo, o desequilíbrio entre aquelas duas dimensões parece levar a uma desvantagem, do que deduzimos a necessidade de entendê-las como funções que se integram e se harmonizam num conjunto. Alguns aspectos da psicologia de Schopenhauer, porém, tendem a distanciar aquelas duas dimensões do organismo (vontade e intelecto), por exemplo: quando ele atribui a uma delas determinado conjunto de tarefas sobre as quais a outra dimensão não exerceria qualquer influência; e quando discrimina a natureza moral dos indivíduos, isto é, o seu caráter, negando ao cérebro qualquer papel sobre este[8].

Neste sentido, tomemos por exemplo as afirmações de Schopenhauer (2015, p. 299-305) de que "o cérebro controla as relações com o mundo exterior: este é o seu único ministério, e com ele paga a sua dívida com o organismo que o alimenta"; ou que "a vontade opera na vida orgânica por meio de estímulos nervosos que não vêm do cérebro". E ainda, sobre a frenologia desenvolvida por Franz J. Gall:

---

7 As traduções do texto de Birnbacher são de nossa responsabilidade.

8 Discutimos ambos os problemas nos subcapítulos 3.2 e 3.3 de nosso livro *Materialismo agônico: corpo, mente e matéria na filosofia de Schopenhauer*, ao qual remetemos o leitor. Os subcapítulos se intitulam "Preeminência da vontade sobre o intelecto" e "O problema da separação entre intelecto e vontade", respectivamente.

> O maior erro da frenologia de Gall é que ele estabeleceu órgãos do cérebro também para as características morais – Ferimentos na cabeça com perda de substância cerebral via de regra fazem efeitos muito prejudiciais ao intelecto (...). Ao contrário, nunca lemos que, após uma infelicidade desse tipo, o caráter tenha sofrido uma mutação, que o indivíduo teria se tornado moralmente pior ou melhor (...) Porque a vontade não possui sua sede no cérebro, e, ademais, ela, como o metafísico, é o *prius* do cérebro, como de todo o corpo, por conseguinte, não sofre mutações por ferimentos do cérebro (SCHOPENHAUER, 2015, p. 298).

Obviamente, não pretendemos aqui resgatar a crítica à famigerada teoria frenológica do anatomista alemão. Antes, trata-se somente de indicarmos, através da crítica que Schopenhauer dirige a Gall, como o filósofo se posiciona a respeito do papel desempenhado pela vontade e pelo intelecto em sua concepção geral sobre o corpo. Se, por um lado, ele formula a tese de uma objetivação da vontade (metafísica) no corpo, por outro, não avança muito na construção de uma concepção mais integrada de organismo, isso porque a vontade mesma não é considerada algo físico, figurando como incessível ao mundo material. A identificação da vontade com o caráter inteligível[9] tende a tornar a relação vontade - intelecto uma via de mão única, onde a primeira engendra o segundo e depois se guarda inacessível numa dimensão metafísica apartada da matéria.

Como bem destaca Rodrigues (2014, p. 51), "A ideia do organismo humano como um sistema integrado composto por vários subsistemas — tais como os sistemas límbico, neural, digestivo, respiratório e reprodutor — é relativamente nova na história da filosofia e da ciência". Nesse sentido, o mais provável é que no interior de tal discussão, na qual a filosofia de Schopenhauer toma parte, ainda não houvesse se fortalecido uma perspectiva mais "interacionista" das diferentes funções intrínsecas ao organismo vivo. Talvez a influência da concepção cartesiana sobre a relação entre alma e corpo ainda pesasse de forma a dificultar o vislumbre de uma relação organicamente integrada da atividade cognitiva, inserida no mecanismo biológico como um todo e em sua intrínseca relação com o ambiente[10].

---

9 As noções de *caráter inteligível*, *caráter empírico* e *caráter adquirido* compõem a base central da psicologia de Schopenhauer, tal como desenvolvida na obra *Sobre o fundamento da moral* (2001). O caráter *inteligível* designa a vontade particularizada em cada indivíduo, o núcleo mais profundo de seu ser singular; o caráter *empírico* se refere à manifestação daquele primeiro em ações objetivas no mundo; e o caráter *adquirido* se refere ao modo como um indivíduo direciona aquelas ações segundo o conhecimento que passa a ter de si mesmo ao longo de sua experiência de vida.

10 Na filosofia contemporânea, a crítica à tradição que pode ser chamada de "cognitivista" é feita em especial pelo enativismo, em obras como *The embodied mind: cognitive science and human experience* (Varela, Thompson e Rosch) e *Mind in life: biology, phenomenology and sciences of mind.* (Evan Thompson). A esse respeito ver nosso artigo *Corpo, cognição e vontade: aproximação e distanciamento entre Schopenhauer e a*

Nota-se um reforço dessa hipótese quando Damásio (2000, p. 61), com quem Birnbacher (2005) dialoga em seu artigo, chama a atenção para o fato de que houve, e em certo sentido ainda há:

> a ausência notável de uma noção de *organismo* na ciência cognitiva e na neurociência. A mente permaneceu ligada ao cérebro em uma relação um tanto equívoca, e o cérebro foi constantemente separado do corpo em vez de ser visto como parte de um organismo vivo e complexo. A concepção de um organismo integrado – a ideia de um conjunto composto de um corpo propriamente dito e de um sistema nervoso – já aprecia na obra de pensadores como Ludwig von Bertalanffy, Kurt Goldstein e Paul Weiss, mas teve pouco impacto na formação das concepções tradicionais de mente e de cérebro.

Com efeito, para citar um desses pensadores apontados por Damásio, temos em Bertalanffy a elaboração de uma concepção de organismo que em muitos aspectos indica a necessidade de uma reorientação de nossa visão clássica de ciência. Ao introduzir uma noção de "sistema" como conceito chave para a observação e interpretação dos fenômenos, Bertalanffy (2015, p. 15) reclama por um novo paradigma científico que está, em suas próprias palavras, "em contraste com o paradigma analítico, mecanicista, causal numa só direção da ciência clássica". O biólogo austríaco entende os organismos como sistemas abertos, ou seja, sistemas que trocam (importam e exportam) matéria com o ambiente. Isso faz com que eles não possuam uma configuração estática, mas dinâmica. A rigor, cada organismo se constitui mais propriamente de "uma ordem hierárquica de sistemas abertos", o que implica que há no interior deles um processo de interação, ou fluxo constante de troca no qual o que aparece em certo nível como estrutura estável se mantém na verdade devido à troca permanente dos componentes de nível imediatamente inferior.

Neste sentido, explica Bertalanffy (2015, p. 206), "o organismo multicelular mantém-se mediante a troca das células, a célula conserva-se pela troca das estruturas celulares, estas por sua vez pela troca dos compostos químicos que as constituem"; o que para o biólogo "é uma boa ilustração do fluxo heraclítico, graças ao qual o organismo vivo se mantém".

Notemos, portanto, que nem sempre é apropriado tomarmos os organismos como sistemas em equilíbrio. A aparente estabilidade deles se configura, na verdade, como um pseudo-equilíbrio dinâmico; o "fluxo heraclítico" que os regula impede que os analisemos sob os mesmos parâmetros dos sistemas fechados. Ou seja, trata-se aqui do reconheci-

---

teoria enativista (OLIVEIRA, 2016, p. 141-152), e o de Baptista e Aldana, *Arthur Schopenhauer and the embodied mind* (BAPTISTA & ALDANA, 2018, p. 153-181).

mento de um nível ou grau dos seres cuja condição de possibilidade é o próprio desequilíbrio, entendido como um movimento contínuo de destruição e regeneração. Em sua hierarquia de sistemas pode haver sistemas em equilíbrio, mas, ainda de acordo com Bertalanffy (2015, p. 162), "o organismo enquanto tal não pode ser considerado um sistema em equilíbrio".

Se a partir disso ampliarmos um pouco nosso escopo, notaremos que, desde o (macro) nível cosmológico até o nível do organismo individual, é possível vislumbrar a ideia schopenhaueriana de uma "luta universal" entre todos os seres e produtos da matéria; pois, segundo o pensador alemão, "do conflito entre fenômenos mais baixos resultam os mais elevados, que devoram a todos, porém efetivando o esforço de todos em grau mais elevado"[11] (SCHOPENHAUER, 2005, p. 209). Assim, o corpo de cada organismo torna-se um palco onde se apresenta a dialética entre equilíbrio e desequilíbrio que subjaz àquele "fluxo heraclítico"; dinâmica na qual intelecto, vontade e ambiente interagem na luta pela permanência do organismo em meio à sua inevitável transformação.

(Para além das) Considerações finais

O reconhecimento do corpo como base fundamental de significação do mundo, tal como lemos no §18 de *O mundo como vontade e como representação*, é uma tese cuja validade é reforçada pela neurociência mais recente. Naquele parágrafo, Schopenhauer assevera que a "busca para a significação do mundo" jamais seria possível se o sujeito que conhece fosse algo como "uma cabeça de anjo alada destituída de corpo". Porém, na medida em que o *sujeito* é entendido como um *indivíduo*, depreende-se daí que aquele conhecimento (processos mentais) provém de um corpo/organismo dotado de vontade. Assim, para usarmos os termos do próprio Schopenhauer, o conhecimento, que é fruto de nossa atividade cerebral, "é no todo intermediado por um corpo, cujas afecções, como se mostrou, são para o entendimento o ponto de partida para a intuição do mundo". Coaduna-se com esta tese schopenhaueriana observações como a de Damásio (2012, p. 20), quando este afirma, por exemplo, que:

11 "Constantemente a matéria que subsiste tem de mudar de forma, na medida em que, pelo fio condutor da causalidade, fenômenos mecânicos, químicos, anseiam avidamente por entrar em cena e assim arrebatam uns aos outros a matéria" (SCHOPENHAUER, 2005, p. 211).

> O nosso próprio organismo, e não uma realidade externa absoluta, é utilizado como referência de base para as interpretações que fazemos do mundo que nos rodeia e para a construção do permanente sentido de subjetividade que é parte essencial de nossas experiências (...). A mente existe dentro de um organismo integrado e para ele; as nossas mentes não seriam o que são se não existisse uma interação entre o corpo e o cérebro durante o processo evolutivo (...). A mente teve primeiro de se ocupar do corpo, ou nunca teria existido.

Ao nos determos no significado da paráfrase que Damásio (2012, p. 184) faz do famoso adágio de Pascal, quando escreve que "o organismo tem algumas razões que a razão tem de utilizar", podemos perceber quão rica em capacidade explicativa (e interpretativa) pode ser uma filosofia do corpo em relação ao seu antípoda, isto é, à antiga postulação e ênfase numa racionalidade descorporificada. Se impulsos biológicos podem atuar como critérios enraizados, isso implica que o organismo tem suas razões; e mais: implica a necessidade de um questionamento da própria concepção tradicional de racionalidade e de conhecimento, bem como das consequências que uma concepção corporificada de conhecimento pode acarretar para compreendermos a dimensão do agir humano.

É certo que, a partir do que, atualmente, sabemos com base na biologia evolutiva, é um ponto pacífico que qualquer investigação filosófica ou científica a respeito da mente deva levar em conta a dependência fundamental que os processos mentais guardam em relação às funções mais básicas do cérebro e do organismo em geral. Contudo, apesar dessa considerável concordância no plano teórico, aquela tradicional concepção "intelectualista" da mente ainda não foi amplamente questionada quanto às suas implicações para o âmbito da filosofia prática, sobretudo para o da política, entendida como conjunto de ações concretas realizadas por agentes dotados de consciência e de vontade.

Se todo o mundo do agir político constitui a realização do que se intenciona a partir de consciências, e se estas são o produto da interação "orgânica" entre vontades e ambiente, isto é, entre vontades e o conjunto de estímulos e circunstâncias de um determinado contexto, então pode e deve haver uma profícua articulação crítica, ainda por se estabelecer, entre uma consistente explicação biológica do organismo e as implicações que esse conhecimento biológico pode trazer para a compreensão do agir político dos indivíduos.

Como bem destaca Rodrigues (2014, p. 51), "Sem o ambiente não há organismo. Sem organismo não há consciência. E sem consciência não há mundo". Isto porque, se podemos falar em "mundo", este só pode ser concebido como o mundo *conhecido*, *intencionado*, ou seja, justamente este mundo em que vivemos; o mesmo que construímos e transformamos continuamente a partir de nossas escolhas e ações. Trata-se de um mundo

construído por indivíduos dotados de vontade; consciências que *querem* de uma forma, e não de outra. Em suma, trata-se de organismos que buscam suas "razões" e as expressam no modo como querem moldar o mundo da vida.

Neste sentido é que queremos destacar que a ideia de uma mente *corporificada* pode abrir horizonte para uma compreensão mais biologicamente fundamentada do agir moral e político dos indivíduos, na medida em que passarmos a entender que suas ações são (também) os produtos de organismos que *querem*; e esse seu *querer* é a expressão de determinadas "razões" (ou valores) construídas ao longo de toda a trajetória multifatorial (genética, ambiente, educação etc.) que constitui a existência de cada organismo e de cada grupo de organismos. Concordamos, assim, com Maturana (1998, p. 33) quando afirma que as raízes do nosso ser cognitivo se estendem até sua própria base biológica, e que, devido a esse fato, "não há dúvida de que ele se manifesta em todas as ações da vida social humana nas quais costuma ser evidente, como no caso dos valores e das preferências. Não há descontinuidade entre o social, o humano e suas raízes biológicas".

Por fim, tomando aqui de empréstimo uma expressão ainda do texto de Maturana (que não se refere diretamente ao filósofo da Vontade, mas que, assim pensamos, pode ser usada no contexto do problema que discutimos ao longo deste artigo); a nosso ver, talvez a mais importante contribuição deixada por Schopenhauer para todo esse escopo de pesquisa, e para a compreensão da relação entre corpo e mente, tenha sido a indicação essencial de que existe, na dimensão prática de nossa existência, um "fundamento não-racional do racional".

Referências

ABBAGNANO, N. **Dicionário de Filosofia**. Tradução: Alfredo Bosi e Ivone Castilho Benedetti. 5. ed. São Paulo: Martins Fontes, 2007.

BAPTISTA, T. & ALDANA, E. Arthur Schopenhauer and the embodied mind. **Ludus Vitalis**, vol. XXVI, n. 49, pp. 153-181, 2018.

BERTALANFFY, L. **Teoria geral dos sistemas**: fundamentos, desenvolvimentos e aplicações. Trad. Francisco M. Guimarães. 8. ed. Petrópolis: Vozes, 2015.

BIRNBACHER, D. Schopenhauer und die moderne Neurophilosophie. **Schopenhauer-Jahrbuch** 86, Frankfurt am Main, 2005.

DAMÁSIO, A. R. **O erro de Descartes**: emoção, razão e cérebro humano. Trad. Dora Vicente e Georgina Segurado. São Paulo: Companhia das Letras, 2012.

DAMÁSIO, A. R. **O mistério da consciência**: do corpo e das emoções ao conhecimento de si. Trad. Laura T. Motta. São Paulo: Companhia das Letras, 2000.

DESCARTES, R. **Discurso do método; As paixões da alma**. Trad. J. Guinsburg e Bento Prado Júnior. 4 ed. São Paulo: Nova Cultural, 1987. (Os Pensadores).

JAQUET, C. **A unidade do corpo e da mente**: afetos, ações e paixões em Espinosa. Trad. Marcos F. de Paula e Luís C. G. Oliva. Belo Horizonte: Autêntica editora, 2011.

MATURANA, H. **Da biologia à psicologia**. Trad. Juan A. Lorens. – 3 ed. – Porto Alegre: Artes Médicas, 1998.

MORGENSTERN, M. Schopenhauers Kritik des Materialismus. **Schopenhauer-Jahrbuch** Frankfurt am Main, 94, 2013.

OLIVEIRA, A. H. M. V. de. Corpo, cognição e vontade: aproximação e distanciamento entre Schopenhauer e a teoria enativista. **Revista Voluntas**: estudos sobre Schopenhauer, [s.l.], v. 7, n. 2, p. 141-52, 2º sem. 2016.

OLIVEIRA, A. H. M. V. de. **Materialismo agônico:** corpo, mente e matéria na filosofia de Schopenhauer. Teresina: Ed. IFPI, 2022. DOI: 10.51361/978-65-86592-54-2.

RODRIGUES, E. G. O organismo: um sistema que integra consciência, mente, corpo e meio ambiente. **Revista Simbio-Logias**, Vol. 7, n. 10, Dez/ 2014, pp. 49-61.

SCHOPENHAUER, A. **Sämtliche Werke in fünf Bänden**. Textkritik bearbeitet und hrsg. Von Wolfgang Frhr. von Löhneysen. Stuttgart/Frankfurt am Main: Suhrkamp taschenbusch wissenchaft, 1989.

SCHOPENHAUER, A. **Sobre o fundamento da moral**. Trad. Maria L. M. Cacciola. 2 ed. São Paulo: Martins Fontes, 2001.

SCHOPENHAUER, A. **O mundo como vontade e como representação**. Tomos I e II. Trad. Jair Barboza. São Paulo: Unesp, 2005/2015.

SCHOPENHAUER, A. **Parerga y paralipómena I y II**. Traducción, introducción y notas de Pilar López de Santa María. Madrid: Editorial Trotta, 2009.

SCHOPENHAUER, A. **Sobre a vontade na natureza**. Trad. Gabriel Valladão Silva. Porto Alegre, RS: L&PM, 2013

THOMPSON, E. **Mind in life:** biology, phenomenology, and the sciences of mind**.** Cambridge: The Belknap Press of Harvard University Press, 2007.

VARELA, F. J., THOMPSON, E. & ROSCH, E. **The embodied mind:** cognitive science and human experience. Cambridge: MIT Press, 1993.


DIAS

RESENHA DO LIVRO *GALILEO'S ERROR: FOUNDATIONS FOR A NEW SCIENCE OF CONSCIOUSNESS* (VINTAGE BOOKS, 2020), DE PHILIP GOFF

João Paulo M. Araujo[1]

Qual o lugar da consciência no mundo material? Como integrá-la em nossa história científica do universo? O livro de Philip Goff objetiva responder a essas e outras questões. Composto por cinco capítulos, dentre os quais considero os mais importantes o primeiro e o quarto, Goff consegue oferecer um bom estado da arte da filosofia da mente contemporânea, tomando sempre como pano de fundo uma boa dose de metafísica e filosofia da ciência. No primeiro capítulo, Goff faz uma análise descritiva de como Galileo criou o problema da consciência ao separar o mundo em qualidades primárias e secundárias. Passando pelo segundo e terceiro capítulos, temos todo um *background* descritivo em torno das teorias dualistas e materialistas, cumprindo, portanto, aquela função propedêutica de situar o leitor nos debates atuais da filosofia da mente. Em seu quarto capítulo, ele apresenta sua proposta pampsiquista para solução do problema da consciência, mostrando, para tanto, seus *insights* e as dificuldades de sua postura. Por fim, no quinto e último capítulo, Goff introduz uma discussão de ordem moral, estética e até mesmo espiritual (embora naturalizada) de como incorporar o pampsiquismo em nossas vidas, transcendendo, por assim dizer, a via filosófica puramente racional de compreensão. Para não correr o risco de me repetir acerca das visões dualistas e materialistas, em minha resenha focarei apenas em como Goff apresenta o problema a partir de Galileu para depois introduzir a sua concepção pampsiquista da realidade.

É a partir da reconstituição do caminho de Galileu na construção das bases da ciência moderna que Philip Goff delineia aquilo que ele acredita ser uma resposta promissora para

1 Doutor em Filosofia pelo programa integrado de Pós-Graduação em Filosofia UFPB-UFPE-UFRN. Mestre em Filosofia pela Universidade Federal de Pernambuco com período sanduíche na Universidad de Buenos Aires pelo Programa Capes PPCP-Mercosul. Professor horista no colegiado de filosofia da Universidade Estadual de Roraima. Membro do grupo de pesquisa Escola Amazônica de Filosofia – EAF.

o difícil problema da consciência. Esta resposta não viria nem por uma via dualista, nem por uma via materialista. O que ele irá propor é uma visão pampsiquista da consciência. Mas no que consiste essa visão? Goff acentua que o pampsiquismo estaria em acordo de que há elementos de verdade, tanto no dualismo quanto no materialismo. Apesar das dificuldades de adequar essa visão ao nosso mundo objetivo, pampsiquistas acreditam que a consciência é uma característica onipresente e fundamental no mundo físico. Como coloca Goff (2020, p. 5), "nada é mais certo do que a consciência, e ainda, nada é mais difícil de incorporar em nossa imagem científica do mundo". Apesar do conhecimento alcançado sobre o funcionamento do cérebro nas últimas décadas, ainda nos é obscura a questão de como o cérebro produz a consciência. Mesmo com todo o progresso funcional e descritivo das neurociências, a ciência ainda não logrou em prover uma explicação decisiva da consciência. Mas por que tanta dificuldade?

De acordo com Goff (2020, p. 6), "a ciência tem um histórico sombrio na explicação da consciência. Mas o histórico da ciência física em explicar praticamente todo o resto é impressionante". É a partir de críticas a teorias como o eliminativismo, Goff faz a sua mais profunda afirmação da realidade da consciência. Mesmo se adotarmos uma postura redutiva ou eliminativista da consciência, a própria atividade científica depende da realidade da consciência. Numa clara analogia, ele afirma que "a ciência não poderia provar que a consciência não existe, assim como a astronomia não poderia provar que não existem telescópios" (Goff, 2020, p. 10). Ao remeter às histórias de vida de Newton e Einstein, Goff (2020, p. 13) chama atenção para o papel da imaginação no desenvolvimento da ciência, isto é, "o fato de que muitos momentos importantes no progresso científico envolveram sonhar com novas possibilidades, evocar na imaginação novas formas de pensar sobre o universo".

De modo similar, Goff acredita que o progresso em nossa compreensão da consciência será feito não somente através da observação do cérebro, mas também de um radical modo de reimaginação da mente e do cérebro. Segundo Goff (2020, p. 13), "há uma boa razão para pensar que explicar a consciência exigirá uma mudança tão fundamental e abrangente quanto a que ocorreu no início da revolução científica". Goff chama atenção para a ideia de que a consciência foi removida do domínio da investigação científica nos primórdios da revolução e que, para resolver o problema, é preciso encontrar um meio de colocá-la de volta.

Quando olhamos de perto a história da ciência, percebemos o quão imbuída de filosofia ela está. Galileu é um desses exemplos marcantes em nossa História Ocidental.

Como sabemos, além de cientista, Galileu também era filósofo. De acordo com Goff, há ao menos dois aspectos no qual a contribuição de Galileu foi muito mais filosófica do que científica. Galileu rejeitou duas características centrais da física de Aristóteles. A primeira delas foi a visão ptolemaica de mundo e a segunda, a teleologia, ou seja, a ideia de que os objetos inanimados possuíam finalidades próprias na natureza.

Mas o ponto que Goff (2020, p. 15) mais chama à atenção é que "Galileu conseguiu rejeitar uma crucial parcela da física de Aristóteles não através da observação ou experimento, mas através de um puro argumento filosófico". Como bem conhecemos a história, Galileu provou que a doutrina aristotélica da queda dos corpos era logicamente incoerente e, nesse sentido, "a reimaginação mais fundamental da natureza realizada por Galileu (...) nunca foi justificada por observação ou experimentos. Ela foi, e continua sendo, uma peça de especulação filosófica" (Goff, 2020, p. 15).

A questão é que antes de Galileu não havia uma linguagem matemática para descrever a natureza. Era de comum acordo entre muitos filósofos uma compreensão do mundo de que as coisas eram dotadas de qualidades sensoriais como cores, cheiros, sons etc. Consequentemente, tais qualidades não poderiam ser apreendidas por uma linguagem matemática. Assim, para resolver esse problema, Galileu propôs uma reimaginação do mundo material na qual os objetos não teriam qualidades sensoriais, mas apenas tamanho, forma, localização e movimento. Para ilustrar esse ponto, imagine uma experiência perceptual (em todas as dimensões sensoriais) de uma maçã. Esta maçã não é algo de fato subjetivamente vermelha, macia, adocicada etc.; ela é apenas um objeto que possui qualidades quantitativas e objetivas, como, por exemplo, tamanho e forma. Ao desconsiderar essas qualidades subjetivas da maçã, Galileu poderia apreender os objetos em linguagem matemática. Mas como ficariam as qualidades sensoriais se elas não existem nos objetos? Neste ponto, seguindo a tradição aristotélica, Galileu afirmou que as qualidades só existem na alma do sujeito que as percebe. Como observa Goff (2020, p. 18), "Galileu transformou as qualidades sensoriais das características de coisas no mundo (...) em formas de consciência na alma dos seres humanos"

É a partir desta divisão que Goff afirma que foi Galileu quem criou o problema da consciência. O ponto crucial é que Galileu não pretendia que a sua física matemática fosse uma descrição completa do mundo, mas apenas dos objetos materiais quantitativos. É aí que entra o erro de Galileu, que "foi comprometer-nos com uma teoria da natureza que implicava que a consciência era essencialmente e inevitavelmente misteriosa" (Goff, 2020, p. 21-22). Em seu livro, Goff considera três possibilidades de correção desse erro.

A primeira delas é o dualismo naturalista, o qual aceita o dualismo de Galileu, mas nega que a mente seja algo misterioso; eles tomam a mente como sendo parte da ordem natural. A segunda é o materialismo, que discorda da visão de Galileu de que a consciência é um fenômeno que resiste a uma explicação física; alguns materialistas mais radicais, como, por exemplo, Keith Frankish[2], argumentam que a consciência é uma ilusão. A terceira é o pampsiquismo, que, de acordo com Goff, pode ser nossa melhor esperança para solucionar o problema da consciência. Quando Galileu decidiu que a ciência não deveria se ocupar da consciência, deu-se início a uma corrida para descobrir qual o lugar dela no universo. Portanto, segundo Goff (2020, p. 23), "para resolver o problema precisamos de alguma forma encontrar uma maneira de tornar a consciência, uma vez mais, uma questão de ciência".

Durante todo o quarto capítulo de seu livro, esse é um dos objetivos de Goff, tornar a consciência uma questão de ciência. Ao definir o pampsiquismo como "a visão de que a consciência é uma característica fundamental e onipresente da realidade física" (Goff, 2020, p. 113), o autor tenta espantar os estereótipos em torno dessa visão filosófica, que, para ele, é muito mal compreendida. Esses estereótipos ocorrem porque nos prendemos ao significado literal da palavra ("pan" = tudo; "psique" = mente). Isso termina acarretando algumas confusões, criando-se o imaginário de que pampsiquistas comungam da ideia de que objetos inanimados possuem vidas conscientes tão ricas quanto as nossas. O seu esforço também se caracteriza por desconstruir essa visão.

De acordo com Goff, existem dois aspectos equivocados em torno desse modo de compreender o pampsiquismo. O primeiro deles é que os pampsiquistas não pensam que tudo é literalmente consciente. Nas palavras de Goff (2020, p. 113), o que eles acreditam é que "os constituintes fundamentais do mundo físico são conscientes, mas não precisam acreditar que todo arranjo aleatório de partículas conscientes resulta em algo que é consciente por si só". Em outras palavras, o pampsiquista negará que suas roupas ou qualquer outro objeto inanimado do seu entorno sejam conscientes, mas, em contrapartida, eles afirmam "que elas são, em última análise, compostas de coisas que são conscientes" (GOFF, 2020, p. 113). Isso nos leva ao segundo aspecto no qual os pampsiquistas não acreditam que a consciência de um ponto de vista humano esteja em toda parte. Nossa forma de consciência, afirma Goff, é o resultado de milhões de anos de evolução pela seleção natural. Portanto, nada com esse tipo de configuração é encontrado em partículas

2 Ver: Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies*, 23, (2016), p. 11-39.

individuais, "se os elétrons têm experiência, então eles têm uma forma inimaginavelmente simples" (GOFF, 2020, p. 113). Do fato de que em nós a consciência é algo extremamente sofisticado, isso não significa que ela não possa existir em formas e escalas mais simples. Uma breve espiada na natureza e podemos perceber muitas outras formas de consciência em outras espécies.

Para Goff, qualquer teoria geral da realidade que não tem lugar para a consciência não pode ser uma teoria verdadeira. A sua esperança é que o pampsiquismo pode ser uma forma de integrar a consciência em nossa imagem científica de mundo, evitando, portanto, os problemas do dualismo e do materialismo. No que concerne ao dualismo, o pampsiquismo evita seus problemas, pois não postula a consciência fora do mundo físico e, portanto, se esquiva do desafio de explicar a interação entre a mente e o cérebro. No que concerne ao materialismo, o pampsiquismo concorda que a consciência está alocada no cérebro, mas não pretende explicar a consciência em termos de processos cerebrais inconscientes. Dessa forma, "o pampsiquismo não oferece uma explicação redutiva da consciência, ou seja, não explica a consciência em termos de algo mais fundamental do que a consciência" (GOFF, 2020, p. 115-116).

Apesar de Goff estar endossando uma visão pampsiquista da realidade, ele não vai buscar diretamente seu referencial teórico em autores pampsiquistas. Como podemos observar em seu livro, ele deixa transparecer que o monismo neutro[3] de Bertrand Russell (1927) e a recepção de sua teoria por Thomas Eddington (1928) no campo da física seriam boas respostas que terminaram contribuindo para o pampsiquismo. A especulação de Russell e Eddington orbitava fora dos debates polarizados entre dualistas e materialistas e, em vista disso, naquele tempo foram completamente descartadas do cenário filosófico dominante. O ponto é que para Goff, a física não nos diz nada sobre a natureza física da realidade. Dito de outra maneira, a física seria apenas uma ferramenta de predição, ela "não nos diz o que é a matéria, mas apenas o que ela faz" (GOFF, 2020, p. 125).

A partir disso, Goff (2020, p. 126) ardilosamente introduz o problema da natureza intrínseca das coisas: "A ciência física limita-se a fornecer informações sobre o comportamento das coisas de que fala – partículas, campos, espaço-tempo – e nada nos diz sobre sua natureza intrínseca". Para ele, o problema das naturezas intrínsecas não é algo presente apenas em questões mais fundamentais da física moderna, mas também surge quando

3 Podemos caracterizar essa postura de Russell como um tipo de metafísica, uma vez que sua visão sustenta que a realidade última das coisas é de um único tipo. Dessa forma, o monismo neutro defende que a natureza intrínseca da realidade não é nem mental nem material, mas, sim, neutra.

discutimos química e neurociências. Vale ressaltar que não é papel da física postular naturezas intrínsecas, isso é algo inerente à nossa tradição metafísica filosófica; o papel da física é prever comportamentos, e é daí que as ciências obtêm seu sucesso. Todavia, Goff (2020, p. 129) afirma que:

> [...] não devemos confundir essa utilidade prática com a aspiração ontológica de prover uma teoria completa da realidade. Sem ir além das informações fornecidas pela ciência física, seremos incapazes de atingir o objetivo final da ciência: uma Teoria de Tudo.

A questão sobre a natureza intrínseca das coisas é algo que acompanha o pensamento filosófico desde o seu alvorecer. Todavia, há um problema maior que os pampsiquistas enfrentam em sua defesa do lugar da consciência em nosso universo físico. Assim como o dualismo e o materialismo, possuem dificuldades ao tratarem do problema mente-corpo. No pampsiquismo, seu maior desafio é conhecido como o problema da combinação. Se um dos objetivos é compreender a natureza intrínseca das coisas a partir de partículas fundamentais nas quais a consciência existe de modo mais simples e elementar, Goff (2020, p. 144) formula a questão e descreve o problema do seguinte modo:

> Como você passa de pequenas coisas conscientes, como partículas fundamentais, para grandes coisas conscientes, como cérebros humanos? Entendemos como tijolos formam uma parede, ou peças mecânicas compõem um motor de carro em funcionamento. Mas não conseguimos entender como pequenas mentes podem de alguma forma se combinar para formar uma grande mente (GOFF, 2020, p. 144).

Para Goff é crucial resolver o problema da combinação, uma vez que isso colocaria o pampsiquismo na frente de outras teorias no que diz respeito à melhor explicação. Ainda mais, Goff considera que o problema da combinação está num outro patamar quando comparado com os problemas enfrentados pelo dualismo e materialismo, pois, ao contrário de seus rivais, o problema seria mais tratável em termos explicativos. A ideia é que teorias materialistas e dualistas teriam lacunas a serem fechadas entre coisas de natureza diferente, enquanto que o pampsiquismo a partir do problema da combinação estaria tentando preencher uma lacuna entre coisas essencialmente do mesmo tipo.

Goff oferece duas respostas ao problema da combinação, as quais seriam promissoras no que diz respeito à realidade do pampsiquismo. Focarei apenas na primeira delas, que surge a partir dos experimentos empíricos com o cérebro dividido. Sabemos que

desde os experimentos de Roger Sperry[4] (1984) é de comum acordo que nosso senso de identidade é produzido a partir de funções centrais realizadas em regiões muito diferentes do cérebro. Com a divisão do corpo caloso, não dividimos apenas o cérebro, mas também a mente, que, por seu turno, leva à existência de duas mentes conscientes localizadas dentro do mesmo cérebro. Um dos efeitos disso é que a "'mente' localizada no hemisfério esquerdo é responsável pela fala, mas é incapaz de reconhecer rostos; a 'mente' localizada no hemisfério direito tem reconhecimento facial, mas é efetivamente muda" (Goff, 2020, p. 151).

Seguindo os passos de Luke Roelofs[5] (2019), Goff argumenta que os casos de cérebro dividido podem ajudar na construção de uma boa resposta para a questão da combinação mental. Mas a estratégia neste caso é a partir do caminho de volta, isto é, de uma descombinação mental. Se antes do corte do corpo caloso tínhamos um indivíduo com uma mente unificada, agora surgem dois indivíduos conscientes separados. Goff acredita que os casos de cérebro dividido podem nos oferecer um controle empírico da combinação mental, ou seja, "se estiver certo, então podemos, imaginando o inverso do que levou à descombinação, inferir o que é necessário para a combinação mental" (GOFF, 2020, p. 152).

Todo o esforço de Goff ao longo de seu livro converge para um momento apoteótico de afirmação do pampsiquismo, isto é, para aquilo que ele chamou de "Um manifesto para uma ciência da consciência Pós-Galileana". Num curto resumo, seu manifesto visa consolidar quatro teses que foram exploradas ao longo de seu livro. A primeira delas é o *realismo sobre a consciência*, ou seja, nossa consciência subjetiva é um dado básico que possui o mesmo valor e *status* dos dados de observacionais e experimentais. O segundo é o *empirismo*, de acordo com o qual a dimensão dos dados quantitativos observacionais e experimentais são tão fundamentais em valor e *status* quanto os dados qualitativos da consciência. Em terceiro, o *anti-dualismo*; nele Goff afirma que a consciência não está separada do mundo físico, mas, sim, presente como uma característica de natureza intrínseca do mundo físico. Em quarto e último, uma *metodologia pampsiquista*. Nesta metodologia, Goff (2020, p. 174) defende que devemos procurar explicar a consciência humana e animal em termos de formas mais básicas de consciência, ou seja, "formas básicas de consciência que se postula existirem como propriedades básicas da matéria".

---

4 Para uma descrição pormenorizada, ver o artigo de Sperry intitulado: Consciousness, personal identity and the divided brain. *Neuropsychologia*, Vol. 22, No. 6, 1984, pp. 661-73.

5 ROELOFS, Luke. *Combining Minds*: How to Think about Composite Subjectivity. New York. Oxford. 2019.

Considero o livro de Philip Goff uma boa obra introdutória ao pampsiquismo em filosofia da mente. Apesar de todo o seu esforço e aposta na ideia de que o pampsiquismo pode oferecer uma solução para o problema da consciência, sabemos que as coisas não são tão simples assim, as questões existem e estão presentes em seu texto. De toda forma, dentre tantas possibilidades - e isso levando em consideração a predominância do materialismo no cenário filosófico atual -, talvez, num futuro próximo, quem sabe, não possamos ter o pampsiquismo figurando como uma tendência na filosofia da mente?

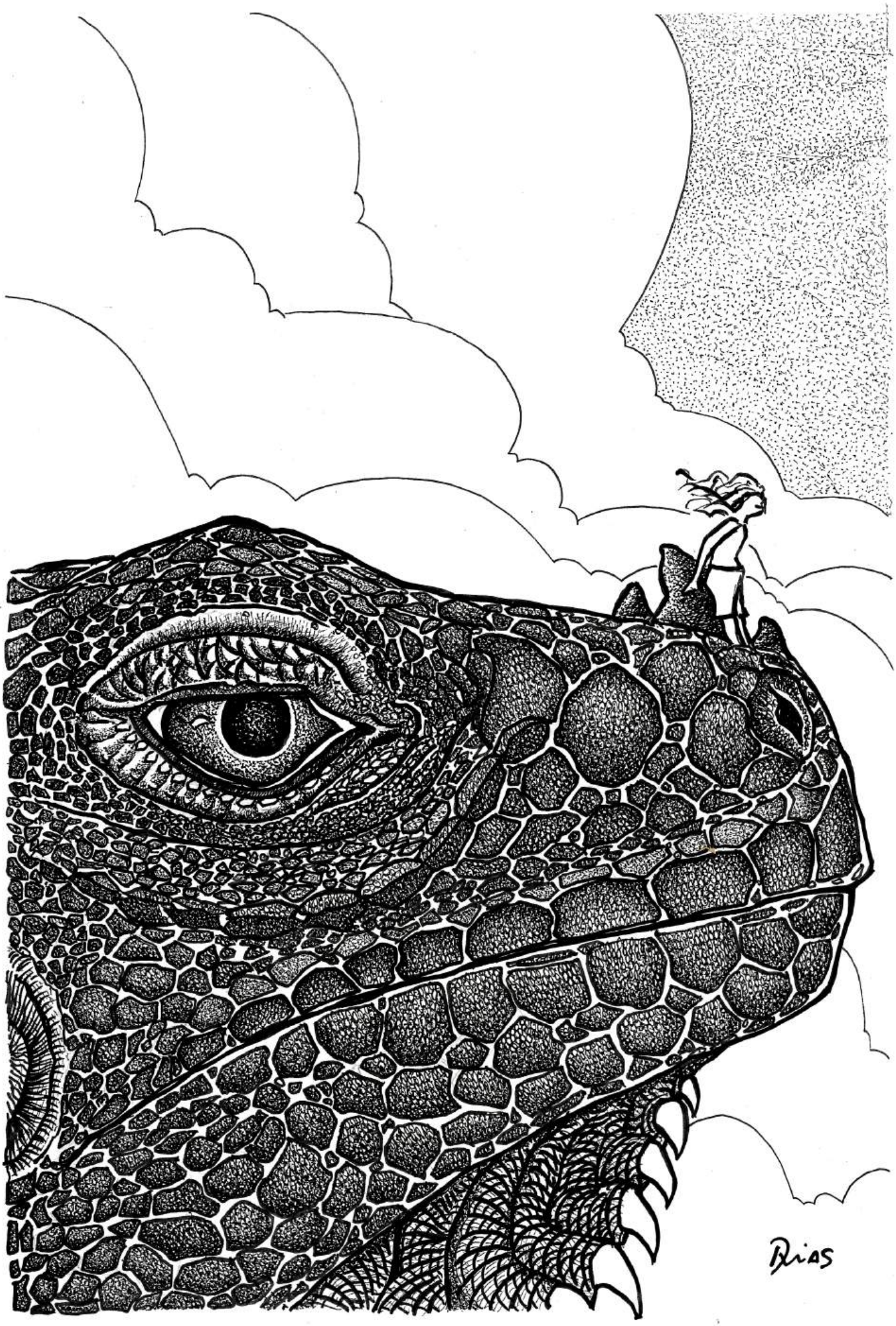
Referências

GOFF, Philip**. Galileo's error:** foundations for a new science of consciousness. New York. Vintage Books, 2020.

FRANKISH, Keith. Illusionism as a Theory of Consciousness. **Journal of Consciousness Studies**, 23, (2016), pp. 11-39.

ROELOFS, Luke. **Combining minds**: How to think about composite subjectivity. New York. Oxford, 2019.

SPERRY, Roger. Consciousness, personal identity and the divided brain. **Neuropsychologia**, Vol. 22, No. 6, 1984, pp. 661-73.


DIAS

PREFÁCIO DO LIVRO *EMERGENT EVOLUTION (THE GIFFORD LECTURES)*

Tradutora: Cristiane Xerez Barroso

Há meio século, com o decorrer dos anos, um estudante foi chamado para ocupar a cátedra em um jantar relacionado com a *Royal School of Mines*. Estavam presentes membros da equipe. E o afortunado jovem foi homenageado com o apoio do Professor Huxley.

"Qual das linhas da ciência que você seguiu despertou mais seu interesse?"

Seguindo o fio da minha resposta, ele extraiu de mim a confissão de que um interesse pela filosofia, e pelo esquema geral das coisas, era mais profundo do que o meu interesse nas aplicações práticas da ciência para o que então pretendia ser o meu treinamento básico. Com uma gentileza compreensiva que logo dissipou meu medo dele, ele me levou a falar mais livremente, a contar-lhe como isso aconteceu, o que eu havia lido e assim por diante. Que tal homem se importasse em saber o que Berkeley e Hume fizeram por mim; o que obtive do Discurso de Descartes; como eu estava naquele momento "enredado em dificuldades" por causa de Spinoza; me encheu de uma feliz surpresa. Seus comentários foram tão maduros; e eles foram feitos para *me* ajudar! "Não importa o que você faça", disse ele, "mantenha essa luz acesa. Mas lembre-se de que a biologia fornece um iluminante novo e poderoso." Então começaram os discursos. Suas palavras de despedida foram: "Quando você atingir o objetivo do seu curso, por que não passar um ano conosco em *South Kensington*?"

Assim, quando obtive o diploma do qual tão pouco uso direto seria feito, e quando minha necessidade do iluminante, e minha falta de conhecimento íntimo dos fatos sobre os quais a nova lâmpada lançava luz foi devidamente impressa em mim durante uma visita à América do Norte e ao Brasil, eu segui seus conselhos, assisti a suas palestras e trabalhei em seu laboratório.

Numa das ocasiões memoráveis em que me chamou para ir à sua sala privada, ele falou da *Genesis of Species*, de *St*. George Mivart. Eu havia feito algumas perguntas a ele alguns dias antes, às quais ele estava ocupado demais para responder; e ele me deu a oportunidade de repeti-las. Mivart disse: "Se então tais poderes inatos devem ser atribuí-

dos a átomos químicos, a espécies minerais, a gêmulas e a unidades fisiológicas, é apenas razoável atribuí-los a cada organismo individual" (p. 260). Perguntei por que motivos esta linha de abordagem era irracional; pois, até então, havia escondido dentro de mim algum toque de "heresia pelagiana" em questões evolutivas. Longe de esnobar um jovem herege, ele tratou-o gentilmente. A questão, disse ele, estava aberta à discussão; mas ele pensava que a posição de Mivart se baseava em outras considerações que não científicas. Qualquer analogia entre o crescimento de um cristal e o desenvolvimento de um organismo era de validade muito duvidosa. "Sim, senhor", eu disse, "exceto que ambos nos convidam a distinguir entre um fator interno e a incidência de condições externas". Ele então perguntou o que eu entendia por "poderes inatos", dizendo que para Mivart eram as "formas substanciais" da tradição escolástica. Aventurei-me a sugerir que os escolásticos e os seus discípulos modernos estavam tentando explicar o que os homens da ciência devem talvez apenas aceitar com base nas evidências. E perguntei se "um poder inato" no organismo poderia ser substituído pelo que ele nos ensinou a denominar de "uma tendência metamórfica interna" que deve ser "tão distintamente reconhecida quanto a de uma tendência conservadora interna" (H.E. ii. p. 116). "Claro que pode, desde que considere isso apenas como uma expressão de certos fatos atualmente inexplicáveis." Perguntei então se era nesse sentido que se deveria aceitar a sua afirmação de que a natureza dá saltos (ii. pp. 77, 97) e, se fosse assim, se a diferença na qual Mivart colocou tanta ênfase – aquela entre as capacidades mentais dos animais e dos homens – não poderia ser considerada um salto natural no progresso evolutivo.

Este era o ponto para o qual eu estava conduzindo. Não me lembro claramente de tudo o que Huxley disse. Minhas anotações, escritas infelizmente não naquela época, mas um ano depois, mostram: "Estresse na fala e na linguagem: nenhuma evidência de *salto* na estrutura laríngea, na boca ou no cérebro: a criança passa *continuamente* do estágio animal para o estágio humano: neuroses e psicoses."

O que ele mais se preocupou em enfatizar ao lidar com Mivart foi que – quer houvesse ou não saltos naturais – sempre houve uma correlação estrita de neuroses e psicoses (ii. pp. 158, 164), que deve ser aceita pela ciência como o resultado natural da evolução do cérebro e da mente. Acreditando que ele cortejou ao invés de se ressentir com uma expressão franca daquilo que ele sentia como uma dificuldade, perguntei por que motivos ele falava da neurose como *antecedente* (i. 238) da psicose; e por que, se elas fossem correlacionadas como concomitantes, não poderíamos seguir Spinoza ao considerar *cada* uma delas como causal dentro do seu atributo e, portanto, *ambas* desempenhando seus papéis na causalidade natural. Ele tinha dúvidas sobre se o tratamento metafísico de Spinoza era

útil na interpretação científica, mas deu-lhe crédito por tentar aprofundar *more suo* as questões fundamentais.

Em conclusão, ao responder a uma batida na porta, ele dispensou um mero neófito com as palavras encorajadoras: "Você poderia muito bem fazer de tudo isso um campo especial de investigação".

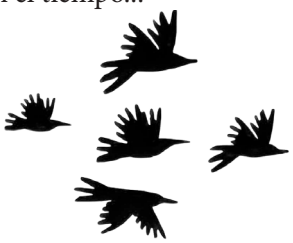
Isso, entre outras coisas, tentei fazer desde então. O fato de o *Senatus* da Universidade de St. Andrews ter me considerado digno de apresentar, como *Gifford Lecturer*, as conclusões às quais fui levado é uma honra pela qual estou profundamente emocionado.

O resultado é um esquema construtivo que Huxley não aceitaria – e isso em mais de um aspecto. Ele não era, contudo, intolerante com conclusões divergentes das suas (embora pudesse sentir-se chamado a combatê-las), se elas fossem honestamente obtidas. E assim, prestando homenagem ao que ele fez por mim há cinquenta anos e posteriormente, digo dele o que o professor Alexander disse delicadamente sobre Spinoza: "Um grande homem não existe para ser seguido servilmente, e pode ser mais honrado pela divergência do que pela obediência."

C. LLOYD MORGAN.

Bristol, fevereiro de 1923.

VIAJE MENTAL EN EL TIEMPO Y FILOSOFÍA DE LA MEMORIA



André Sant'Anna[1]
rosolemandre@gmail.com

Tradutor: Juan F. Álvarez

Sobre la presente traducción

Los estudios sobre la memoria en la filosofía analítica han aumentado considerablemente desde la segunda mitad del siglo XX. Uno de los resultados de esta proliferación de trabajos sobre la memoria ha sido la consolidación de las teorías causales, las cuales han provisto la explicación hegemónica de la memoria. Sin embargo, recientes investigaciones en las ciencias cognitivas han puesto en tela de juicio la plausibilidad de tal explicación y han conducido a construir teorías que prescinden de criterios causales. El presente texto de André Sant'Anna es una reseña de los resultados de la intersección entre ciencias cognitivas y filosofía de la memoria. El autor se centra particularmente en el marco conceptual del viaje mental en el tiempo y avizora líneas de investigación promisorias.

Si bien los cuestionamientos por la naturaleza de los recuerdos se remontan a la Antigua Grecia y han sido tema de discusión en momentos históricos distintos, el desarrollo de los estudios filosóficos de la memoria a partir de los años sesenta y el reciente debate entre causalistas y anti-causalistas han sido factores determinantes para la consolidación de la filosofía de la memoria como disciplina independiente (BERNECKER Y MICHAELIAN, 2017; MICHAELIAN, DEBUS Y PERRIN, 2018). La literatura sobre la filosofía de la memoria se halla predominantemente en lengua inglesa. Aunque los estudios tanto sociales como literarios acerca de la memoria son abundantes en español, las preguntas sobre la definición de la memoria, la naturaleza de sus objetos y sus implica-

1 Durante la publicación del artículo original, el autor estaba afiliado a la Universidad de Otago (Nueva Zelanda). En el momento en que se realiza esta traducción, Sant'Anna es investigador postdoctoral de la Fundación Alexander von Humboldt en la Universidad de Colonia (Alemania).

ciones epistemológicas y metafísicas permanecen desatendidas en la comunidad hispanoparlante. Dado que Sant'Anna se enfoca aquí en estos aspectos, se espera que la presente traducción motive la discusión en español de las preguntas filosóficas sobre la memoria.

Cabe resaltar que algunas publicaciones recientes han introducido estos debates en la comunidad hispanoparlante. Sin embargo, Felipe De Brigard (2018) ha publica la sola introducción a la filosofía analítica de la memoria en español, la cual aparece en la *Enciclopedia de la Sociedad Española de Filosofía Analítica*. Otros autores han publicado reflexiones sobre la memoria desde un enfoque histórico. Por ejemplo, Díaz Quiroz (2021) propone un estudio de la memoria intelectual en la obra de Descartes, mientras que Fierro (2021) examina la relación entre memoria y amor en la obra de Platón. Otros filósofos han emprendido estudios de interés científico. Guerrero-Velázquez (2021), por ejemplo, examina la relación entre memoria y percepción en las entrevistas autobiográficas llevadas a cabo en contextos experimentales.

Cabe resaltar igualmente que Marina Trakas es probablemente la filósofa de la memoria más activa en la comunidad hispanoparlante. La autora ha propuesto análisis en los que se coordinan enfoques científicos, artísticos y filosóficos para estudiar la memoria (véase, por ejemplo, TRAKAS 2017, 2021a, 2021b, 2022). Recientemente, Trakas (2022) publicó un artículo sobre el viaje mental en el tiempo de especial pertinencia para el texto de Sant'Anna. En dicho artículo, la autora propone que la noción metafórica de "viaje mental en el tiempo" carece de claridad y genera tensiones conceptuales tanto en las ciencias cognitivas como en la filosofía y, por consiguiente, debería ser abandonada. Sant'Anna, en cambio, no se ocupa de la clarificación de la metáfora. El autor se concentra en las preguntas filosóficas que puede suscitar la investigación científica sobre el viaje mental en el tiempo. Se invita al lector, entonces, a mantener en mente que, si bien Trakas y Sant'Anna proponen análisis en niveles explicativos distintos, el viaje mental en el tiempo es un tema complejo que sigue alimentando debates en múltiples comunidades académicas.

El objetivo de la presente traducción comparte el mismo espíritu introductorio del texto de De Brigard. Asimismo, esta traducción pretende nutrir los estudios en español en los que se profundizan las discusiones filosóficas sobre la memoria a la vez que se trascienden los límites de diferentes disciplinas. A un nivel práctico, se espera que esta traducción pueda ser utilizada en cursos introductorios a la filosofía analítica de la memoria.

Finalmente, permítaseme hacer un comentario sobre un aspecto de la traducción. He conservado la abreviatura estándar en inglés "www" (*what, when, where*) para hacer referencia a dos cosas: primero, a la perspectiva según la cual la memoria episódica invo-

lucra información sobre el evento ocurrido (*what*), su localización temporal (*when*) y su ubicación espacial (*where*) y, segundo, a este tipo de información. Así pues, las expresiones "perspectiva www" e "información www" serán utilizadas en aras a la simplicidad.

Agradezco a André Sant'Anna y a la revista *Filosofia Unisinos – Unisinos Journal of Philosophy* por concederme el permiso para realizar esta traducción. En el volumen 19 (número 1) de esta revista puede consultarse el artículo original en inglés.

*
* *

VIAJE MENTAL EN EL TIEMPO Y FILOSOFÍA DE LA MEMORIA

Abstract:

The idea that episodic memory is a form of mental time travel has played an important role in the development of memory research in the last couple of decades. Despite its growing importance in psychology, philosophers have only begun to develop an interest in philosophical questions pertaining to the relationship between memory and mental time travel. Thus, this paper proposes a more systematic discussion of the relationship between memory and mental time travel from the point of view of philosophy. I start by discussing some of the motivations to take memory to be a form of mental time travel. I call the resulting view of memory the *mental time travel view*. I then proceed to consider important philosophical questions pertaining to memory and develop them in the context of the mental time travel view. I conclude by suggesting that the intersection of the philosophy of memory and re-search on mental time travel not only provides new perspectives to think about traditional philosophical questions, but also new questions that have not been explored before.

Keywords:

Mental time travel; memory; episodic memory; philosophy of memory.

Resumo:

A ideia de que a memória episódica é uma forma de viagem no tempo mental ("mental time travel") desempenhou um papel importante no desenvolvimento da pesquisa sobre memória nas últimas duas décadas. Apesar de sua crescente importância na psicologia, apenas recentemente os filósofos começaram a interessar-se por questões filosóficas relativas à relação entre memória e viagem no tempo mental. Assim, este artigo propõe uma discussão mais sistemática da relação entre memória e viagem no tempo mental do ponto de vista da filosofia. Começo discutindo algumas das motivações para considerar a memória uma forma de viagem no tempo mental. Chamo a visão resultante sobre a memória de visão "mental time travel". Procuro então considerar importantes questões filosóficas relativas à memória e desenvolvê-las no contexto da visão "mental time travel". Concluo sugerindo que a intersecção entre a filosofia da memória e a pesquisa sobre viagem no tempo mental não apenas fornece novas perspectivas para pensar sobre questões filosóficas tradicionais, mas também novas questões que não foram exploradas anteriormente.

Palavras-chave:

Viagem no tempo mental; memória; memória episódica; filosofia da memória.

Resumen:

La idea de que la memoria episódica es una forma de viaje mental en el tiempo ha desempeñado un rol importante en el desarrollo de las investigaciones sobre la memoria en el último par de décadas. A pesar de su creciente importancia en la psicología, los filósofos apenas han comenzado a desarrollar un interés en las cuestiones filosóficas relativas a la relación entre la memoria y el viaje mental en el tiempo. Este artículo propone una discusión más sistemática de la relación entre memoria y viaje mental en el tiempo desde el punto de vista de la filosofía. Con este objetivo en mente, comienzo discutiendo algunas de las razones que motivan la concepción de la memoria como una forma de viaje mental en el tiempo. Al punto de vista resultante lo denomino la *perspectiva del viaje mental en el tiempo*. Posteriormente, procedo a considerar cuestiones filosóficas importantes sobre la memoria y las desarrollo en el contexto de dicha perspectiva. Concluyo sugiriendo que la intersección de la filosofía de la memoria y la investigación sobre el viaje mental en el tiempo no solamente provee nuevas perspectivas para pensar cuestiones filosóficas tradicionales, sino también nuevas preguntas que no han sido exploradas anteriormente.

Palabras clave:

Viaje mental en el tiempo; memoria; memoria episódica; filosofía de la memoria.

Introducción

La idea de que la memoria episódica es una forma de viaje mental en el tiempo ha desempeñado un rol importante en el desarrollo de las investigaciones sobre la memoria en el último par de décadas. El viaje mental en el tiempo, de acuerdo con Suddendorf y Corballis (1997, p. 133), "abarca la reconstrucción mental de eventos personales del pasado (memoria episódica) y la construcción mental de eventos posibles en el futuro". "La importancia real del viaje mental en el tiempo", agregan, "radica en viajar al futuro en vez de viajar al pasado; así, en el presente permanecemos en mayor medida enfrentando el futuro que observando el pasado" (SUDDENDORF y CORBALLIS, 1997, p. 147).

La memoria se ha considerado tradicionalmente como una capacidad relacionada con el pasado, en el sentido de que nos permite evocar eventos que sucedieron previamente. No obstante, la sugerencia según la cual la memoria episódica es sólo una forma de viaje mental en el tiempo desafía esta idea, pues "la función primordial del viaje mental en el tiempo hacia el pasado consiste en proveer la materia prima a partir de la cual se construyen e imaginan futuros posibles" (SUDDENDORF y CORBALLIS, 2007, p. 302). Estas consideraciones dan lugar a preguntas filosóficas importantes. Una de estas preguntas pone en duda si la memoria requiere o no una conexión causal apropiada con los eventos o experiencias del pasado. Desde el artículo de Martin y Deutscher (1966), la posición estándar ha sido asumir que para recordar se requiere una conexión de este tipo (véase, por ejemplo, BERNECKER, 2008; DEBUS, 2008; MICHAELIAN, 2011; ROBINS, 2016b). Una segunda pregunta relevante es si la memoria episódica puede ser una fuente de conocimiento sobre el pasado (véase, por ejemplo, DEBUS, 2014; MICHAELIAN, 2016b). Dado que el viaje mental en el tiempo hacia el pasado, o memoria episódica, sirve para proveer la materia prima con miras a simular escenarios futuros, no es claro en qué condiciones nos puede proveer información fiable de los acontecimientos pasados. Una tercera pregunta más general cuestiona la relación entre la memoria y otras formas de viaje mental en el tiempo, tal como la imaginación de eventos futuros. Dado que ambas son el resultado de capacidades cognitivas similares, la pregunta de si pertenecen al mismo tipo metafísico toma un lugar central en filosofía (véase PERRIN y MICHAELIAN, 2017).

Estas y otras preguntas han atraído la atención de los filósofos interesados en la memoria (véase, por ejemplo, DE BRIGARD, 2014A; DEBUS, 2014; MICHAELIAN, 2016B; PERRIN, 2016). En este artículo exploraré algunas de las implicaciones que la *perspectiva del viaje mental en el tiempo sobre la memoria*, como me referiré a esta perspectiva, tiene para la filosofía de la memoria. Comenzaré discutiendo algunas de las razones que motivan la consideración de la memoria episódica como una forma de viaje mental en el tiempo. Luego exploraré las implicaciones de esta idea para la filosofía de la memoria.

La memoria episódica y el viaje mental en el tiempo

Antes de discutir la relación entre la memoria y el viaje mental en el tiempo, será útil clarificar primero qué *es la memoria episódica*. Este término fue inicialmente introducido por Endel Tulving (1972, p. 385) y, a grandes rasgos, corresponde al sistema de memoria responsable de recibir y almacenar "información sobre episodios o eventos temporalmente datados y sobre relaciones espaciotemporales entre estos eventos".[2] Así, cuando un sujeto recuerda episódicamente un evento, su memoria contiene información asociada con el evento en cuestión sobre el *qué*, el *dónde* y el *cuándo*. Esta es la perspectiva qué-dónde-cuándo de la memoria episódica o, simplemente, la *perspectiva www* (por sus siglas en inglés). En la formulación inicial de Tulving, los recuerdos episódicos contrastan con los recuerdos semánticos, los cuales se refieren a los recuerdos sobre hechos generales que no fueron necesariamente experimentados. Por ejemplo, cuando recuerdo que la Segunda Guerra Mundial finalizó en 1945, estoy recordando un hecho semánticamente por medio del lenguaje. El sistema de memoria semántica, dice Tulving (1972, p. 386), se refiere al "conocimiento organizado que posee una persona sobre las palabras y otros símbolos verbales, sus significados y referentes, sobre las relaciones entre ellos, sobre reglas, fórmulas y algoritmos para manipular aquellos símbolos, conceptos y relaciones". Por ende, a diferencia de los recuerdos episódicos, los recuerdos semánticos no requieren la experiencia previa de eventos.

2 El término *memoria* es ambiguo y podría referirse a diferentes cosas, tales como, la *capacidad* de un sujeto para recordar (por ejemplo, "John tiene una buena memoria"), el *sistema cognitivo* responsable de producir recuerdos (por ejemplo, "tu memoria no está funcionando bien"), o los *outputs* de ese sistema cognitivo, tales como, los *estados mentales* que podemos llamar "memorias" (por ejemplo, "guardo memoria de la fiesta de mi décimo cumpleaños"). Para mis propósitos, uso el término con miras a hacer referencia tanto al sistema cognitivo responsable de producir recuerdos, como a los estados mentales individuales producidos por tal sistema.

El elemento importante para destacar en esta definición de la memoria episódica es que está primordialmente basada en el tipo de información que es procesada y almacenada. A causa de ello, la definición afronta importantes problemas. Uno de tales problemas tiene que ver con el hecho de que algunos recuerdos semánticos poseen información "www"; por ejemplo, mi recuerdo de que la batalla de Waterloo se libró en 1815 ♠ [3]. Por tanto, no es enteramente claro si los recuerdos episódicos y los recuerdos semánticos pueden ser distinguidos solamente en función de la información contenida en ellos. Otro problema se refiere a la dimensión fenomenológica de los recuerdos episódicos. Recordar un evento particular que fue previamente experimentado parece involucrar más que la recuperación de información. Recordar episódicamente parece tener una fenomenología distintiva, que involucra un "sentimiento del pasado" (RUSSELL, 1921, p. 161-2) y un "sentimiento de calidez e intimidad" (JAMES, 1890). En otras palabras, además de la información contenida, los recuerdos episódicos parecen hacer referencia al pasado ("sentimiento del pasado") y pertenecer a los sujetos de una manera única ("sentimiento de calidez e intimidad"). Por ejemplo, cuando recuerdo mi décima fiesta de cumpleaños, la memoria no solamente presenta el evento como habiendo ocurrido en el pasado, sino también como "mío", en el sentido de que parezco *poseer* el recuerdo.

Estas y otras dificultades condujeron a Tulving a reformular su primera caracterización de la memoria episódica. Posteriormente, Tulving propuso una definición que tomaba en cuenta los aspectos fenomenológicos descritos más arriba. De acuerdo con el autor, aparte de contener información "www", los recuerdos episódicos incluyen un tipo único de conciencia, que él llamó *conciencia autonoética* o *autonoesis* (TULVING, 1985, 2005). La autonoesis, afirma Tulving (2005, p. 15), "se refiere al tipo de conciencia que caracteriza la recolección consciente de sucesos personales", es decir, es lo que hace a los sujetos "conscientes de que la experiencia presente está relacionada con la experiencia pasada de una manera en la que ningún otro tipo de experiencia lo está" (p. 15)[4].

♠ 3 Como he especificado más arriba, la expresión "información www" se refiere a la información sobre qué (*what*), dónde (*where*) y cuándo (*when*). Así, el ejemplo del recuerdo semántico que propone el autor sobre la batalla de Waterloo de 1815 posee la información que es supuestamente característica de los recuerdos episódicos. ¿Qué se recuerda? la batalla, ¿dónde ocurrió la batalla? en Waterloo, ¿cuándo? en 1815 (nota del traductor).

4 Aunque inicialmente caracterizada en términos fenomenológicos, no hay acuerdo entre los autores sobre qué es exactamente la autonoesis. Algunos han afirmado, por ejemplo, que la autonoesis tiene un valor epistémico importante. Dokic (2001, 2014) sostiene que la memoria episódica conlleva un "sentimiento de conocimiento", en la medida en que indica a los sujetos que los recuerdos episódicos se originan en sus experiencias pasadas. Fernández (2016) defiende una perspectiva similar, pero incluye la autonoesis en el contenido de la memoria antes que en su fenomenología. Hace poco Mahr y Csibra (2018) propusieron una

La definición de la memoria episódica que involucra la autonoesis es muy importante. Dado que "el acto de recordar […] es caracterizado por una conciencia peculiar y única de re-experimentar aquí y ahora algo ocurrido con anterioridad, en otro tiempo y en otro lugar" (TULVING, 1993, p. 68), la memoria hace que los sujetos "sean capaces de *viajar mentalmente en el tiempo*: […] una persona puede transportarse a voluntad hacia el pasado personal y hacia el futuro" (TULVING, 1993, p. 67, énfasis añadido). Así, además de ser la responsable de la sensación única asociada con los recuerdos episódicos, la autonoesis dota a los sujetos con la capacidad más general de "viajar" en el tiempo subjetivo. No es difícil proponer esta idea a partir de fundamentos fenomenológicos. Como señala Klein (2015, p. 21), hay una "simetría temporal percibida entre los movimientos hacia adelante (futuro) y hacia atrás (pasado) desde el presente". A modo de ilustración, considere un sujeto que está pensando en las vacaciones en la playa del próximo año. De manera similar a los recuerdos episódicos, el sujeto tiene la sensación de que el pensamiento es poseído por él, en el sentido de que las vacaciones son suyas y no de otro individuo. Sin embargo, a causa de que el evento es algo que *puede* ocurrir, tal evento le es presentado como algo "futuro" en su pensamiento presente. Parece, entonces, que podemos "reubicarnos" en el futuro de la misma manera en que podemos hacerlo en relación con el pasado.

La capacidad que nos confiere la autonoesis de viajar al tiempo pasado subjetivo y al tiempo futuro subjetivo constituye un incentivo importante para concebir la memoria episódica solamente como una forma — entre otras — de *viaje mental en el tiempo*. Pese al énfasis hecho en las consideraciones fenomenológicas anteriores, también hay buenas razones empíricas para respaldar esta perspectiva. En una investigación reciente, Perrin y Michaelian (2017) discuten las similitudes entre la memoria episódica y el viaje mental al futuro encontradas en diferentes dominios. En estudios sobre psicología del desarrollo, por ejemplo, se ha mostrado que la capacidad de los niños para recordar el pasado e imaginar el futuro surge aproximadamente al mismo tiempo (SUDDENDORF Y BUSBY, 2005; ATANCE, 2008; FIVUSH, 2011). En estudios sobre pacientes con trastornos en la memoria, se ha encontrado que las deficiencias en la memoria incurren en deficiencias similares en la habilidad de pensar sobre escenarios futuros (KLEIN *et al*., 2002; ROSENBAUM *et al.*, 2005; HASSABIS *et al*., 2007). Además, ciertos estudios con técnicas de imagenología cerebral muestran también que hay un fuerte solapamiento entre las regiones cerebrales

---

explicación "comunicativa" de la función de la memoria episódica, en la cual la autonoesis es concebida como responsable de "delinear cuál de nuestras afirmaciones sobre el pasado podemos sostener con autoridad epistémica". A pesar de estos desarrollos importantes, tomaré por sentado la idea más estándar de que la autonoesis es principalmente una característica fenomenológica de la memoria episódica.

asociadas con la memoria episódica y con el viaje mental al futuro (ADDIS *et al*., 2007; SCHACTER *et al.*, 2007, 2012).

No intentaré revisar la literatura relevante aquí[5]. Señalaré, más bien, un desarrollo importante de *la perspectiva del viaje mental en el tiempo sobre la memoria*. Más recientemente, algunos investigadores han sugerido que la función primordial del viaje mental en el tiempo no es permitirnos recordar el pasado. Suddendorf y Corballis (1997, p. 147), por ejemplo, afirman que "la importancia real del viaje mental en el tiempo radica más en viajar al futuro que al pasado; así, en el presente permanecemos en mayor medida enfrentando el futuro que observando el pasado". En esta misma línea, De Brigard (2014a, p. 158, énfasis añadido) dice que "recordar es una operación particular de un sistema cognitivo que permite la recombinación flexible de diferentes componentes de las huellas codificadas en representaciones de eventos pasados posibles […] *al servicio de* la construcción de eventos futuros posibles" ♠ [6]. Y, más recientemente, Michaelian (2016b, p. 103) afirma que "recordar no es un tipo de proceso diferente al de otros procesos constructivo-episódicos"; de modo que, "recordar es imaginar un episodio perteneciente al pasado personal" (p. 111).

La idea de que la función principal del viaje mental en el tiempo no es recordar el pasado, sino imaginar el futuro tiene consecuencias importantes. Una de esas consecuencias es que nuestra concepción del sentido común sobre la memoria, de acuerdo con la cual la función de la memoria es almacenar información de lo sucedido en el pasado, parece estar en peligro. Que nuestras representaciones del pasado sean inexactas siempre y cuando sean beneficiosas para futuras acciones es compatible con la perspectiva del viaje mental en el tiempo. De manera que, como indica De Brigard (2014a, p. 158), "muchos

5 Para una revisión detallada referida a cuestiones filosóficas, véase Perrin y Michaelian (2017).

♠ 6 El término metafórico "huella" que utiliza De Brigard en esta cita hace referencia al término técnico "huella mnémica" frecuentemente usado en la filosofía y neurociencia de la memoria. Durante el proceso de codificación de información se fortalecen ciertas redes neuronales gracias a la coactivación de determinadas regiones cerebrales—el lóbulo temporal medio, el córtex sensorial, el córtex prefrontal lateral y el córtex parietal superior. Según De Brigard (2014a), las huellas mnémicas son las propiedades disposicionales de tales redes, las cuales permiten su reactivación con aproximadamente el mismo patrón de activación que tuvieron durante la codificación. Dichas propiedades pueden alterarse con el paso del tiempo a causa de distintos factores y por tanto modificar el contenido de los recuerdos episódicos. Hasta qué punto puede modificarse el contenido de los recuerdos es una pregunta que permanece abierta en ciencias cognitivas y es parte importante del debate epistemológico entre preservacionistas (DUMMETT, 1994; AUDI, 1995; BURGE, 1997; SENOR 2017) y generacionistas (LACKEY, 2005; FERNÁNDEZ 2016; MICHAELIAN, 2016b). Para un examen de la relación entre filosofía, neurociencia y psicología cognitiva en torno a la pregunta por las huellas mnémicas, véase De Brigard (2014b). Para una discusión del rol de las huellas mnémicas en las teorías filosóficas de la memoria, véase Robins (2017b) (nota del traductor).

casos ordinarios de recuerdos defectuosos *no deben* ser vistos como instancias del mal funcionamiento de la memoria". Esto suscita una pregunta adicional que es de particular interés para los filósofos sobre si la memoria proporciona conocimiento del pasado y, de ser así, cómo lo haría. Debido a que la función primordial de la memoria no es recuperar información sobre el pasado, necesitamos una explicación adecuada acerca de cómo el conocimiento puede formarse a partir de la memoria. Asimismo, la perspectiva del viaje mental en el tiempo plantea cuestiones relevantes pertenecientes a la relación entre los recuerdos y los eventos pasados. La teoría causal de la memoria, la cual ha sido predominante en la filosofía en las pasadas cuatro décadas, estipula que recordar requiere de la preservación de una conexión causal apropiada con los eventos del pasado. Sin embargo, si la memoria es una forma de viaje mental en tiempo de la misma manera que lo es la imaginación, y "si la imaginación no necesita basarse en la información almacenada en última instancia originada en la experiencia del episodio relevante" (MICHAELIAN, 2016b, p. 111), entonces no hay una razón de principio para decir que tal requisito aplique para la memoria.

En suma, la perspectiva del viaje mental en el tiempo sobre el recuerdo suscita una gran cantidad de preguntas y problemas importantes para los filósofos interesados en la memoria. En la siguiente sección, con la intención de desarrollar aquellos problemas, consideraré algunas de las implicaciones que tiene dicha perspectiva para la filosofía de la memoria.

El viaje mental en el tiempo y la filosofía de la memoria

La perspectiva del viaje mental en el tiempo sobre la memoria no sólo desafía importantes concepciones tradicionales, sino que también ofrece prospectos para investigaciones futuras en el tema. En esta sección, consideraré algunos temas relacionados con la perspectiva del viaje mental en el tiempo que son de interés potencial para los filósofos de la memoria. No obstante, a causa de que el interés de los filósofos en estos temas es aún muy reciente, no hay muchos trabajos que realicen un tratamiento sistemático de las cuestiones que discutiré más abajo. Por esta razón, en lugar de intentar ofrecer un sondeo del debate, trataré de motivar algunos problemas de interés potencial.

La teoría causal de la memoria

Luego de la publicación del artículo seminal de Martin y Deutscher (1966), "Remembering", los filósofos de la tradición analítica comenzaron a desarrollar un creciente interés en las cuestiones filosóficas referidas a la memoria. Martin y Deutscher propusieron lo que es conocido ahora como la *teoría causal de la memoria* (TCM). La TCM ha sido considerablemente influyente y todavía moldea en gran medida la manera en que los filósofos piensan la memoria hoy.[7] Sin embargo, el punto de vista del viaje mental en el tiempo plantea importantes inquietudes sobre esta teoría.

La TCM provee un conjunto de criterios para determinar si un estado mental determinado es o no un recuerdo. Para la TCM, un sujeto *S* recuerda un evento *e* si y sólo si:

> (1) *S* representó *e* en el pasado (*condición de la representación pasada*);
> (2) *S* tiene una representación mental presente de *e* (*condición de la representación presente*);
> (3) El contenido de la representación mental presente de *e* es suficientemente similar al contenido de la representación pasada de *e* (*condición del contenido*);
> (4) Hay una conexión causal apropiada entre la representación presente de *e* y la representación pasada de *e* (*condición de la conexión causal*).[8]

Para clarificar estos criterios, considérese mi recuerdo aparente de mi décima fiesta de cumpleaños. Con miras a que mi recuerdo cuente como un recuerdo genuino del evento, necesito haber experimentado previamente dicho evento. Esta es la *condición de la representación pasada*. Adicionalmente, necesito ser capaz de representar el mismo evento en el presente. Esta es la *condición de la representación presente*. Sin embargo, mis representaciones pasada y presente pueden ser representaciones del mismo evento únicamente si sus contenidos son suficientemente similares (esta es la *condición del contenido*); por ejemplo, si los contenidos de ambas representaciones contienen amigos, miembros de mi

7 Michaelian y Robins (2018) ofrecen un análisis reciente y exhaustivo de la TCM en relación con los desarrollos recientes en la filosofía de la memoria.

8 Esta discusión es adaptada a partir de Bernecker (2010, capítulo 1). Véase también Bernecker (2015, p. 302).

familia, una torta de chocolate, etc.[9] Finalmente, recordar requiere que mi representación actual de mi décima fiesta de cumpleaños sea causada de una manera apropiada por mi representación pasada del mismo evento (esta es la *condición de la conexión causal*). El requisito de la conexión causal es la principal novedad de la TCM. Además, dado que es también fuente de problemas que surgen en el contexto de la perspectiva del viaje mental en el tiempo sobre la memoria, me centraré en este requisito más detenidamente.

Se supone que la condición causal descarta casos que, intuitivamente, no cuentan como recuerdos, pero que son permitidos por (1)-(3) ♠[10]. Con miras a ilustrar esto, considérese el caso de Kent descrito por Martin y Deutscher (1966, p. 174):

> Un hombre que llamaremos Kent presencia un accidente automovilístico y observa detalles particulares del accidente a causa de su posición especial. Más tarde, Kent está involucrado en otro accidente en el cual recibe un severo golpe en la cabeza y, como resultado, olvida cierta sección de su propia historia, incluyendo el primer accidente. Kent ya no puede cumplir el primer criterio para recordar el primer accidente. Algún tiempo después de este segundo accidente, un hipnotizador popular y más bien irresponsable ofrece un espectáculo. Él hipnotiza un gran número de personas y les sugiere que ellos creerán que han estado en un accidente automovilístico en una hora y lugar determinados. El hipnotizador nunca ha oído nada sobre Kent ni sobre los detalles del accidente de Kent y es por pura coincidencia que la hora, el lugar y los detalles que proporciona son exactamente como los del primer accidente de Kent. Kent es uno de los miembros del grupo que ha sido hipnotizado. La sugestión funciona y […] [Kent] cree firmemente que ha estado en un accidente. El accidente, como Kent cree que es, es justo como el primero en el que estuvo realmente involucrado.

Este caso satisface (1) y (2), en la medida en que Kent tiene una representación pasada del accidente automovilístico y una representación actual del mismo evento. El caso

9 Martin y Deutscher (1966) conciben esta similitud en términos de una analogía estructural entre la representación pasada y la representación actual. Los autores afirman que "la experiencia pasada debe constituir un análogo estructural de la cosa recordada en la medida en que [el sujeto] pueda representar la cosa exactamente" (p. 191). No es totalmente claro, sin embargo, dónde se encuentra la analogía estructural. La interpretación más natural parece ser que el contenido de la representación pasada debe tener el mismo tipo de estructura que el contenido de la representación presente, pero los autores no se pronuncian sobre la estructura de aquellos contenidos. Otro problema es que no es claro qué tanta "analogía estructural" es requerida para que *S* esté recordando. Si bien no queremos que se requiera que el contenido de la representación pasada sea el mismo que el contenido de la representación presente, es difícil encontrar una manera para determinar en principio cuánta similitud es requerida. Para mis propósitos, dejaré a un lado estas preocupaciones. Para una discusión relacionada, véase Michaelian (2011 y 2016b, p. 90).

♠ 10 Sant'Anna usará la expresión "condición causal" para referirse de forma abreviada a la condición (4), que más arriba denominó la "condición de la conexión causal" (nota del traductor).

satisface además (3), pues la representación actual de Kent es suficientemente similar a su representación pasada. No obstante, no parece correcto decir que Kent está recordando genuinamente. La razón para afirmar esto consiste en que su representación actual no preserva el tipo correcto de conexión causal con su representación pasada. Para usar el término de Martin y Deutscher (1966), la representación pasada no es "operativa" al producir la representación actual. En el caso de Kent, la causa operativa, por así decirlo, es el hipnotizador. Para la TCM, entonces, recordar no es sólo cuestión de acertar los detalles de una experiencia pasada de un evento, sino también de permanecer en una relación causal apropiada con esa experiencia.

Además de ofrecer una forma de excluir casos no contemplados por (1)-(3), la condición de la conexión causal ha sido usada para proveer una taxonomía de la memoria. En su versión actual, la TCM es una respuesta a la pregunta general sobre qué se precisa para que un sujeto recuerde. No obstante, hay más de una manera en la que uno puede recordar algo exitosa o no exitosamente, lo cual requiere una explicación de aquellas diferencias. Por ejemplo, es consistente con el recuerdo de mi décima fiesta de cumpleaños que me equivoque con algunos de sus detalles[11]. Puedo recordar correctamente que toda mi familia estaba allí y que la fiesta tomó lugar en cierta ubicación, pero puedo recordar simultánea e incorrectamente que comí torta de fresa. En este caso, podemos decir que estoy *recordando mal* (*misremembering*) mi décima fiesta de cumpleaños. Así, Robins (2016b) ha afirmado recientemente que, dado el carácter constructivo de la memoria (véase BARTLETT, 1995; SCHACTER *et al.*, 2007, 2012; MICHAELIAN, 2011; DE BRIGARD, 2014a), necesitamos apelar a la conexión causal entre las representaciones pasadas y actuales para distinguir el recuerdo del recuerdo falso[12]. En esta misma línea, Bernecker (2017) ha sugerido que uno solamente puede distinguir el recuerdo exitoso de las confabulaciones (véase HIRSTEIN, 2005) si uno exige que el primero, pero no las segundas, preserve una conexión causal con experiencias pasadas (véase también ROBINS, 2016b, 2017a). Por tanto, la conexión causal es importante para proveer un análisis adecuado no sólo del recuerdo, sino también de los diferentes tipos de recuerdo exitoso y no exitoso.

La perspectiva del viaje mental en el tiempo sobre la memoria desafía el estatus central otorgado a la condición de la conexión causal en una teoría de la memoria. Como

11 Aunque, de nuevo, no es del todo claro qué grado de inexactitud es compatible con el recuerdo. Véase Michaelian (2011) y la nota al pie 9.

12 En Michaelian (2016a) se halla tanto una crítica de la propuesta de Robins como un intento por ofrecer una taxonomía de la memoria que abandona completamente la conexión causal.

se discutió más arriba, según esta perspectiva, la función primordial de la memoria no es recordar el pasado (véase SUDDENDORF Y CORBALLIS, 1997; DE BRIGARD, 2014A; MICHAELIAN, 2016B). Pero, si tal es el caso, entonces es difícil entender la razón por la cual deberíamos respaldar la TCM. Hay múltiples razones para pensar de este modo. Como plantea Michaelian (2016b, p. 111), una de esas razones consiste en que, dado que otras formas de viaje mental en el tiempo no necesitan tener una conexión causal con las experiencias pasadas, no hay en principio ninguna manera de requerirla en el caso de la memoria. Esto no significa, por supuesto, que no pueda haber una conexión tal, sino únicamente que no es necesaria.

Otra razón que pone en duda la TCM es que, desde la perspectiva del viaje mental en el tiempo, instancias sencillas del recuerdo serían descartadas por la TCM. La conexión causal nos permite preservar la intuición según la cual, en casos como el de Kent, los sujetos no están recordando. Sin embargo, intuitivamente parece que no requerimos que todas las instancias del recuerdo conserven una conexión causal apropiada con los eventos pasados. Considere el siguiente caso: imagine que yo experimenté en el pasado mi décima fiesta de cumpleaños y que tengo ahora el presunto recuerdo de ello. Recuerdo a mis amigos y a mi familia estando allí y me recuerdo comiendo pastel de chocolate; no obstante, suponga que mi representación actual no está siendo causada por mi representación previa de mi décima fiesta de cumpleaños, sino por dos experiencias diferentes que involucraron los elementos relevantes de mi representación actual. En este caso, el contenido de mi representación actual se deriva, en parte, de mi experiencia de mi décima fiesta de cumpleaños — a la que asistieron los mismos individuos — y también se deriva, en parte, de mi experiencia de otra fiesta a la que asistí — donde hubo una torta de chocolate. En el presente caso, no hay conexión causal del tipo requerido por la TCM, pero parece demasiado restrictivo afirmar que el sujeto no está recordando el evento relevante solamente porque el contenido de su representación actual no se deriva del contenido de la experiencia original[13].

13 Se podría afirmar aquí que, intuitivamente, el caso no cuenta como una simple instancia del recuerdo justamente porque no hay conexión causal. No pretendo poner en disputa las intuiciones de las personas sobre este y otros casos similares, pero, en la medida en que queremos que nuestras intuiciones sean compatibles con lo que la investigación empírica nos dice sobre la memoria, esta parece la forma más plausible de describirlas. En otras palabras, dado el carácter constructivo de la memoria (véase, por ejemplo, BARTLETT, 1995; SCHACTER *et al*., 2007, 2012; MICHAELIAN, 2011; DE BRIGARD, 2014a), no es improbable que casos como el descrito puedan ocurrir.

La tercera razón por la cual la perspectiva del viaje mental en el tiempo desafía la TCM es que tal teoría es incompatible con el carácter constructivo del viaje mental en el tiempo. Debido a que el viaje mental en el tiempo está al servicio de la simulación de eventos para asistir a los sujetos en futuras interacciones con el entorno, parece demasiado restrictivo exigir que nuestras representaciones del pasado tengan que basarse en el contenido proveniente de una única fuente particular. Por ejemplo, al pensar sobre cómo debo actuar en mi entrevista de trabajo de la próxima semana, mi representación actual del pasado será más beneficiosa si se basa en diferentes experiencias pasadas de entrevistas de trabajo, que si se basa únicamente en una experiencia particular[14].

En suma, la TCM ha ocupado una posición central en la teorización filosófica sobre la memoria durante los últimos cincuenta años. Además de ofrecer un análisis de la memoria que explica un amplio rango de casos, tal teoría proporciona un principio útil para concebir una taxonomía de la memoria. Sin embargo, si la perspectiva del viaje mental en el tiempo es correcta, la centralidad de la TCM podría no estar justificada.

El viaje mental en el tiempo y nuestro conocimiento del pasado

Una consecuencia directa del abandono de la condición causal puede observarse en la epistemología de la memoria. Dado que tal condición ya no es necesaria para recordar, no hay garantía de que el contenido de nuestras representaciones presentes se derive del contenido de nuestras representaciones pasadas. Siendo este el caso, la pregunta que se plantea por sí misma es si podemos formar conocimiento de lo ocurrido en el pasado sobre la base de nuestras representaciones presentes y, de ser así, cómo lo haríamos. ¿Es el viaje mental en el tiempo capaz de proporcionar tal conocimiento? Antes de ocuparme de esta pregunta, es importante distinguir los dos sentidos en que puede formularse. Por una parte, podemos formular la pregunta pragmática sobre la posibilidad de que la memoria nos provea información que, en contextos prácticos, posibilite inferencias útiles sobre cómo fueron las cosas en el pasado. Este cuestionamiento lo denominaré *la pregunta epistémica pragmática*. Por otra parte, podemos preguntar si la memoria provee realmente conocimiento del pasado, en el sentido de que sirve como fundamento de nuestras creencias justificadas sobre el pasado. Llamaré este cuestionamiento la *pregunta epistémica estricta*.

14 Véase, sin embargo, Sutton (1998) y Michaelian (2011) para diferentes intentos de proveer una perspectiva causal compatible con el carácter constructivo de la memoria. Para una discusión relacionada, véase Robins (2016a y 2016b).

Esta distinción es importante porque una respuesta afirmativa a la pregunta epistémica pragmática no nos da necesariamente una respuesta afirmativa a la pregunta epistémica estricta. Puede darse el caso de que el contenido de mi recuerdo de mi décima fiesta de cumpleaños sea el mismo o sea muy similar al contenido de los recuerdos que otras personas tienen de este evento, de modo que yo pueda realizar inferencias útiles sobre este evento en contextos pertinentes; pero, de esto no se sigue que mi memoria me permite conocer algo sobre el evento en cuestión. En contraste, una respuesta a la pregunta epistémica estricta requiere de la identificación de los elementos que permiten que nuestros recuerdos presentes sirvan como fundamentos de nuestras creencias justificadas sobre el pasado.

La condición causal provee una respuesta a la pregunta epistémica estricta. Dado que el contenido de mi representación presente de un evento está causado por mi representación pasada de tal evento, la conexión causal hace posible que la memoria justifique mi conocimiento del pasado. En otras palabras, las creencias que formamos con base en la memoria están justificadas porque hay una conexión causal apropiada entre recuerdos y eventos pasados. No obstante, si, como sugiere la perspectiva del viaje mental en el tiempo, esta condición no es necesaria para recordar, ¿cómo podemos explicar la relación entre el contenido de nuestras representaciones pasadas y presentes?

No es del todo claro qué alternativas hay aquí para los defensores de la perspectiva del viaje mental en el tiempo. De hecho, debido a que es el crítico más sistemático de la condición causal, Michaelian (2016b) ha sido el único hasta ahora en proporcionar un tratamiento explícito a esta pregunta. Su enfoque adopta un marco conceptual fiabilista amplio en epistemología, de acuerdo con el cual "el estatus epistémico de una creencia está determinado por la fiabilidad del proceso que la produce" (MICHAELIAN, 2016b, p. 39; véase también GOLDMAN, 2012). A grandes rasgos, la idea consiste en que uno está justificado al sostener una creencia determinada si un proceso fiable produce la creencia en cuestión. De acuerdo con la propuesta de Michaelian, podemos explicar por qué la memoria sirve como fundamento para formar conocimiento del pasado en términos de la fiabilidad de sus procesos subyacentes. Esta solución, sin embargo, no resultará atractiva para quienes no están inclinados hacia alguna forma de fiabilismo. La razón de esto es que, como Michaelian (2016b, p. 40) reconoce, la solución toma al fiabilismo como punto de partida y luego procede a explicar *cómo* la memoria es fiable. No obstante, si se es escéptico de la idea de que la fiabilidad por sí misma puede proporcionar una explicación de la justificación epistémica, una explicación de la manera en que la memoria es fiable no será suficiente para hacer frente a la pregunta epistémica estricta.

La pregunta acerca de si el fiabilismo es o no una buena explicación de la justificación epistémica está más allá de los límites de este texto. Sin embargo, hecha explicita la pregunta sobre cómo la memoria puede formar conocimiento sobre el pasado, podría ser útil explorar otras alternativas. Un acercamiento posible a la pregunta podría ser adoptar una visión *eternalista* de los eventos (BERNECKER, 2008). De acuerdo con el eternalismo, los eventos no dejan de existir cuando se convierten en eventos pasados. El eternalismo es promisorio porque permite afirmar que los eventos pasados son partes constitutivas de los recuerdos. Para ilustrar esto, considere una analogía con la percepción. Los relacionistas de la percepción sostienen que los objetos de tamaño mediano son partes constitutivas de la percepción, en el sentido de que yo no podría tener una experiencia visual de la silla en mi oficina si este objeto no estuviese ahí (véase, por ejemplo, CAMPBELL, 2007; MARTIN, 2004; BREWER, 2007; FISH, 2009). Una motivación importante para reconocer el rol constitutivo desempeñado por los objetos en la percepción es que esto permite explicar cómo los objetos fundamentan nuestro conocimiento del mundo (para una discusión reciente véase SCHELLENBERG, 2016). De manera similar, podría afirmarse que reconocer el rol constitutivo desempeñado por los eventos pasados en los recuerdos nos permite explicar cómo tales eventos fundamentan nuestro conocimiento del pasado[15].

Sin embargo, el eternalismo enfrenta problemas importantes. Por ejemplo, no es obvio cómo nuestros recuerdos pueden estar constituidos por eventos localizados en una ubicación espaciotemporal diferente. Si bien el eternalismo deja espacio, por lo menos en principio, para que la relación entre eventos pasados y recuerdos tome lugar mediante el reconocimiento de la existencia de tales eventos, aún se requiere una explicación de la manera en que estos se relacionan con nuestras representaciones mentales presentes. El problema consiste en que es difícil entrever cómo sería tal explicación. Otro problema de esta perspectiva es que nos exige pagar un alto precio metafísico para explicar cómo la memoria fundamenta nuestro conocimiento del pasado. Dado que estamos obligados a postular la existencia de eventos pasados, algunos pueden ver esta solución con escepticismo (por ejemplo, MICHAELIAN, 2016b, p. 63).

15 Debus (2008, p. 406-7) hace la misma observación cuando afirma que "la explicación relacional [de la memoria] debe ser verdadera si aceptamos (como deberíamos) que las personas pueden obtener algunas veces conocimiento sobre el pasado a partir de sus [recuerdos]". Sin embargo, su explicación de la memoria requiere la postulación de una separación fundamental entre la memoria y otras formas de viaje mental en el tiempo, lo que hace que su visión sea poco prometedora aquí (véase, DEBUS, 2014 y la sección "The metaphysics of mental time travel").

Otra alternativa, que llamaré la *solución pragmática*, consiste en negar que la pregunta epistémica pragmática es diferente de la pregunta epistémica estricta. Desde este punto de vista, tener conocimiento del pasado es simplemente una cuestión de realizar inferencias útiles sobre cómo eran las cosas en aquel entonces. Si tenemos o no conocimiento del pasado, dirá el pragmatista, depende de la manera en que nuestros recuerdos puedan informar nuestro comportamiento futuro. Si los recuerdos posibilitan comportamientos que conducen a la acción coordinada con otros individuos en situaciones pertinentes—tales como, discutir quién asistió a la fiesta de cumpleaños o, más primitivamente, discutir dónde es posible encontrar comida—entonces eso es todo lo que se requiere para decir que tenemos conocimiento del pasado. En consecuencia, el pragmatista negará que necesariamente debe haber una conexión causal con las representaciones pasadas, siempre y cuando las representaciones actuales permitan inferencias útiles sobre el pasado.

La solución pragmatista también enfrenta problemas importantes. El primero es similar al problema del fiabilismo planteado anteriormente. En otras palabras, la solución solamente será atractiva para aquellos que ya están inclinados hacia una visión pragmatista en epistemología. El segundo problema es que la solución pragmatista parece arbitraria, en la medida en que parece implicar que nuestro conocimiento del pasado depende de lo que ciertos individuos "acuerdan" que sea el caso. Empero, no es claro quiénes son los individuos pertinentes en cada situación o incluso si hay en principio una manera de identificarlos. Más aún, el énfasis en la utilidad podría conducir a resultados contraintuitivos, pues un recuerdo podría ser útil para guiar la conducta actual de diferentes individuos sin ser verdadero respecto al pasado. Dicho de otra manera, es completamente plausible que los sujetos puedan recordar mal algunos o todos los detalles de un evento de manera similar—así como su memoria también puede reportar concordancias con la de los otros—pero aun así fallar en describir efectivamente lo ocurrido.

En conclusión, parece que una explicación del modo en que formamos conocimiento del pasado de acuerdo con la perspectiva del viaje mental en el tiempo podría requerir algunos compromisos controversiales. Si bien estos compromisos pueden tomar lugar en diferentes dominios — por ejemplo, en metafísica, como el caso de la solución eternalista, o en epistemología, como los casos de las soluciones fiabilista y pragmatista — una respuesta convincente a esta pregunta requerirá inevitablemente una motivación adecuada de aquellos compromisos.

Los objetos del viaje mental en el tiempo

La perspectiva del viaje mental en el tiempo sobre la memoria también suscita preguntas importantes acerca de los objetos del viaje mental en el tiempo ♠ [16]. Si la memoria es solamente una forma de viaje mental en el tiempo, entonces una explicación de los objetos de la memoria dependerá inevitablemente de una explicación más general de los objetos del viaje mental en el tiempo. Tradicionalmente, los filósofos han abordado el cuestionamiento de los objetos de la memoria con bastante detalle. Inspiradas en Hume (2011) y Locke (1975), las visiones *representacionales* o *realistas indirectas* sostienen que los objetos de la memoria son representaciones internas de eventos (véase RUSSELL, 1921; BYRNE, 2010). En contraste, las visiones *relacionales* o *realistas directas* plantean que los objetos de la memoria son los eventos pasados mismos (véase REID, 2000; LAIRD, 2014; RUSSELL, 2001; DEBUS 2008). Dado este marco teórico, una sugerencia natural para abordar la pregunta de los objetos del viaje mental en el tiempo sería tomar la explicación predilecta de los objetos de la memoria y aplicarla al viaje mental en el tiempo. Sin embargo, esto parece poner las cosas al revés. Desde la perspectiva del viaje mental en el tiempo, la categoría de "viaje mental en el tiempo" es más básica que la categoría de "memoria", así que necesitamos primero una explicación de los objetos del viaje mental en el tiempo, la cual informará luego nuestra explicación de los objetos de la memoria.

La pregunta por los objetos del viaje mental en el tiempo no ha sido abordada hasta ahora en la literatura, por lo que no hay puntos de vista establecidos sobre ella ♠ [17]. No obstante, esto no debe prevenirnos de pensar sobre cómo podría lucir una respuesta a la pregunta. Una forma de comenzar a hacerle frente consiste en distinguir entre diferentes formas de viaje mental en el tiempo. Si bien esto no siempre se hace explícito en las discusiones sobre el tema, hay más de una forma en la que puede ocurrir el viaje mental hacia el pasado y hacia el futuro. Además de la memoria episódica — la cual se refiere al viaje mental en el tiempo hacia eventos que ocurrieron en el pasado — y del pensamiento futuro episódico — que se refiere al viaje mental en el tiempo a eventos que podrían ocurrir — también podemos pensar sobre los eventos contrafactuales localizados en el tiempo

♠ 16 La noción de "objeto" del viaje mental en el tiempo hace referencia a la cuestión ontológica sobre la naturaleza de los elementos con los que se relacionan los sujetos en primera instancia cuando recuerdan, imaginan o piensan contrafácticamente los eventos (nota del traductor).

♠ 17 En efecto, al momento de la publicación del texto original no había publicaciones que abordaran la pregunta por los objetos del viaje mental en el tiempo. Sin embargo, Sant'Anna y Michaelian (2019) abordaron posteriormente este asunto y propusieron una perspectiva particular al respeto (nota del traductor).

subjetivo (véase DE BRIGARD, 2014a). Por ejemplo, yo puedo pensar sobre cómo sería mi vida en este momento si no hubiera ido a la universidad. En este caso, estoy pensando sobre un evento que podría haber ocurrido en el pasado y que influenciaría el presente, pero que ya no es posible. Similarmente, puedo pensar sobre cómo sería mi vida en diez años si no hubiera ido a la universidad. En este caso, estoy pensando sobre un evento que tomaría lugar en el futuro si algún otro evento en mi pasado hubiera sido diferente. En ambos casos, entonces, estoy considerando pensamientos sobre situaciones contrafactuales orientadas al pasado y al futuro.

Lo anterior sugiere que una explicación de los objetos del viaje mental en el tiempo necesita tomar en cuenta no solamente la memoria episódica y el pensamiento futuro episódico, sino también las formas del pensamiento contrafactual episódico (véase DE BRIGARD, 2014a) dirigido al pasado y al futuro. Esto hace significativamente más difícil la pregunta inicial, pues ahora tenemos que explicar cómo las cosas que ya no pueden ser el caso pueden ser, de alguna manera, los objetos de nuestros pensamientos. Una línea de investigación prometedora podría ser apelar a la noción de *objetos intencionales*. Como originalmente propuso Brentano (2014), los objetos intencionales son inexistentes y son los objetos directos de conciencia de la mente. Aunque esta es una línea prometedora, nadie la ha desarrollado sistemáticamente hasta ahora[18].

Otra alternativa podría ser echar una mirada a las explicaciones tradicionales de los objetos de la memoria como puntos de partida. Aunque las perspectivas relacionales han sido defendidas más consistentemente en el contexto de la memoria, estas no parecen ofrecer prospectos para una explicación más general de los objetos del viaje mental en el tiempo. La razón es que los objetos del viaje mental en el tiempo — excepto discutiblemente los objetos de la memoria — no existen y es imposible estar relacionado con algo inexistente. Así, a menos que se esté dispuesto a comprometerse con perspectivas metafísicas más controversiales — tales como el punto de vista según el cual existen objetos intencionales (CRANE, 2001, 2013) o alguna forma de realismo modal (LEWIS, 1986) – no es claro si las visiones relacionales pueden ser sostenidas coherentemente. En contraste, las visiones representacionales podrían ser más prometedoras. Dado que los objetos que son representados por el viaje mental en el tiempo no necesitan existir para ser representados, no hay necesidad de preocuparse por el estatus metafísico de aquellos eventos. Lo que es relevante para explicar cómo somos conscientes de los eventos pertinentes es la

18 Véase, sin embargo, Crane (2001, 2013) para discusiones potencialmente útiles sobre los objetos intencionales en filosofía de la mente.

existencia de representaciones, que servirían como sustitutos de los eventos. No es claro, sin embargo, qué problemas habría para una explicación representacional de los objetos del viaje mental en el tiempo. A causa de que esta pregunta no ha sido explorada con suficiente detalle, queda por ver si el representacionalismo puede resistir un análisis más detallado [19].

La metafísica del viaje mental en el tiempo

La consideración de los cuestionamientos anteriores nos conduce, finalmente, a examinar una pregunta más general sobre la metafísica del viaje mental en el tiempo. Como vimos, la perspectiva del viaje mental en el tiempo sobre la memoria plantea muchos problemas diferentes respecto a la epistemología y la metafísica de la memoria. Pero qué tan apremiantes sean aquellas preguntas dependerá de la manera en que se conciba la categoría de memoria en relación con la categoría más amplia de viaje mental en el tiempo. Hasta ahora, he tomado por sentado que hay buenas razones para aceptar que la memoria es solamente otra ocurrencia del viaje mental en el tiempo. Empero, algunos filósofos se han resistido a esta perspectiva. Debus (2014), por ejemplo, sostiene que la memoria y el viaje mental en el tiempo orientado hacia el futuro—o lo que la autora llama imaginación sensorial — son fenómenos de diferentes clases, porque hay importantes disimilitudes metafísicas entre ellos.

El debate sobre la metafísica del viaje mental en el tiempo es aún muy reciente y, como ocurre con algunas de las preguntas anteriores, no hay perspectivas bien establecidas en la literatura. A pesar de este hecho, seguiré aquí a Perrin y Michaelian (2017) y distinguiré entre perspectivas metafísicas del viaje mental en el tiempo *continuistas* y *discontinuistas*. La primera acepta que las similitudes entre la memoria y otras formas de viaje mental en el tiempo apoyan la visión más general de que son fenómenos del mismo tipo. La segunda, en contraste, sostiene que aquellas similitudes no son suficientes para

[19] A pesar de que en este texto el autor sugiere otorgar un carácter promisorio al representacionalismo, Sant'Anna y Michaelian (2019) plantearon recientemente una crítica de acuerdo con la cual la perspectiva representacionalista no es viable para dar cuenta de los objetos del viaje mental en el tiempo, debido a que los estados mentales que contarían como instancias de este sistema cognitivo no podrían establecer sus propias condiciones de satisfacción. Adicionalmente, los autores proponen una perspectiva pragmatista — basada en el trabajo de Charles S. Peirce — que explicaría de mejor manera los objetos del viaje mental en el tiempo, recogería los aspectos relevantes tanto del representacionalismo como del relacionalismo y desecharía los elementos problemáticos de ambas posturas (nota del traductor).

decir que la memoria y otras formas de viaje mental en el tiempo son fenómenos del mismo tipo.

Las razones para respaldar el continuismo varían. La razón general, sin embargo, parece provenir de diferentes vertientes de la investigación en las ciencias empíricas ♠ [20]. Como discutí en la segunda sección, hay una gran variedad de trabajo empírico que resalta las importantes similitudes entre la memoria episódica y el viaje mental en el tiempo. Tal vez la razón más distintiva proviene del hecho de que el viaje mental al pasado y el viaje mental al futuro se basan en recursos cognitivos similares, lo cual sugiere que un mecanismo cognitivo común o "central" responsable del viaje mental en el tiempo será eventualmente identificado (ADDIS *et al.*, 2007; SCHACTER *et al.*, 2007, 2012). En términos más filosóficos, podemos concebir el continuismo como la perspectiva basada en una postura naturalista frente a la pregunta por la relación entre la memoria episódica y el viaje mental en el tiempo. Dicho de otro modo, según las visiones continuistas, dado que hay una cantidad considerable de evidencia empírica que sugiere que la memoria episódica es sólo otra instancia del viaje mental en el tiempo, debemos tomar esta evidencia seriamente cuando pensamos sobre la metafísica del viaje mental en el tiempo.

Las perspectivas discontinuistas, por el contrario, parecen estar motivadas por consideraciones *a priori* más generales sobre la metafísica del viaje mental en el tiempo. Esto no quiere decir, por supuesto, que el discontinuismo simplemente ignora la evidencia empírica sobre la que el continuismo se apoya[21]. Los discontinuistas, en cambio, creen que otras consideraciones — tal como la pregunta por la posibilidad de que el viaje mental en el tiempo establezca una conexión causal apropiada con los eventos en cuestión — son también importantes para proveer una imagen adecuada de la metafísica del viaje mental en el tiempo. Debus (2014), por ejemplo, afirma que la memoria episódica y otras formas de viaje mental en el tiempo son fenómenos de dos tipos fundamentalmente distintos[22]. En apoyo a esta tesis, la autora afirma que, a diferencia de la memoria episódica, otras formas de viaje mental en el tiempo no logran inducir a los sujetos en una relación experiencial con los eventos pertinentes. La noción técnica de la relación experiencial se refiere a la

♠ 20 Para una defensa reciente de la perspectiva discontinuista motivada por razones empíricas, véase Robins (2020) (nota del traductor).

21 Véase, por ejemplo, Perrin (2016) para una perspectiva discontinuista más modesta que toma en cuenta las similitudes subrayadas por la investigación empírica.

22 Debus (2014) discute únicamente la relación entre la memoria episódica y el pensamiento orientado al futuro, o imaginación sensorial. Sin embargo, dado que su argumento parece sugerir que otras formas de viaje mental en el tiempo son igualmente distintas de la memoria episódica, no haré esta distinción aquí.

relación causal y espaciotemporal que los sujetos tienen con los eventos involucrados en sus pensamientos. En la memoria episódica, esta relación se obtiene porque los eventos pertinentes ocurrieron y podemos trazar, al menos potencialmente, la conexión causal entre el recuerdo actual y el evento pasado. En otras formas de viaje mental en el tiempo, por el contrario, la relación no se obtiene puesto que los eventos pertinentes no existen.

Aparte de reflejar diferentes actitudes metafilosóficas hacia la misma pregunta, la disputa entre continuistas y discontinuistas refleja diferentes compromisos asumidos en relación con las cuestiones discutidas en secciones anteriores. Considérese la pregunta sobre si la memoria episódica requiere o no una conexión causal apropiada con los eventos pasados. Aunque el continuismo es compatible con la TCM, la perspectiva continuista no proporciona un lugar central a la condición de la conexión causal en su teorización metafísica del viaje mental en el tiempo. Para el continuismo, la presencia (o ausencia) de una conexión causal refleja, a lo sumo, únicamente una diferencia de grado entre la memoria episódica y otras instancias del viaje mental en el tiempo. Para los discontinuistas, sin embargo, esta pregunta es central en la metafísica del viaje mental en el tiempo y la presencia (o ausencia) de una conexión causal es suficiente para separar dos fenómenos mentales como dos tipos metafísicos diferentes.

Lo mismo aplica para la cuestión sobre nuestro conocimiento del pasado y los objetos del viaje mental en el tiempo. De acuerdo con los continuistas, como Michaelian (2016b), una explicación adecuada de la manera en que la memoria episódica provee conocimiento del pasado puede ofrecerse examinando la fiabilidad de los mecanismos que producen los recuerdos, lo cual no requiere conexiones causales con el pasado. Así, los elementos que probablemente hacen conscientes a los sujetos de los eventos pasados son las representaciones internas, que son separables de aquellos eventos. En este sentido, los continuistas podrían estar más inclinados a adoptar una perspectiva representacional de los objetos del viaje mental en el tiempo. Para los discontinuistas, en contraste, la memoria episódica es capaz de proporcionar conocimiento de una manera que otras formas de viaje mental en el tiempo no pueden. Esto es gracias a que la memoria episódica proporciona una relación con los eventos pasados, en la cual está involucrada necesariamente una conexión causal con tales eventos y que no es posible experimentar por medio de otras formas de viaje mental en el tiempo. Así pues, los discontinuistas podrían no estar satisfechos con una visión representacional de los objetos del viaje mental en el tiempo como representaciones de eventos que pueden ocurrir en la ausencia de conexiones causales con los eventos pertinentes. Una perspectiva realista directa o relacional de la memoria (DEBUS, 2008) parecerá ser, por tanto, más atractiva para los discontinuistas, la cual es reconocida por Debus (2014) como parte central de su explicación discontinuista.

Conclusión

La perspectiva según la cual la memoria es una forma de viaje mental en el tiempo ofrece apasionantes proyectos para nuevas investigaciones en el subcampo emergente de la filosofía de la memoria. Las visiones tradicionales de la memoria — tal como la teoría causal de la memoria — y las preguntas tradicionales acerca de la memoria — tales como la pregunta por la manera en que provee conocimiento del pasado y cuál es la naturaleza de sus objetos — necesitan ser reconsideradas en el marco conceptual más amplio del viaje mental en el tiempo. No obstante, estos cuestionamientos están interrelacionados con preguntas más generales y nuevas que surgen sólo en la investigación sobre el viaje mental en el tiempo, esto es, cuáles son los objetos del viaje mental en el tiempo y cuál es el estatus metafísico de aquellos estados mentales. De modo que la intersección entre la filosofía de la memoria y la investigación sobre el viaje mental en el tiempo no solamente provee nuevas perspectivas para pensar las preguntas tradicionales, sino también nuevos cuestionamientos que no han sido explorados anteriormente.

Agradecimientos

Agradezco a Kirk Michaelian por los comentarios sobre un borrador previo y por las diversas discusiones sobre temas relacionados con este artículo. Agradezco también a los participantes del seminario de filosofía de la memoria de 2017 en la Universidad de Otago y a los participantes del taller *Mental Time Travel e Agência Moral* en Unisinos, donde tuve la oportunidad de discutir algunos de estos temas con mayor detalle.

Referencias

ADDIS, D. R., Wong, A. T. & SCHACTER, D. L. Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. **Neuropsychologia**, *45*(7), 1363-77, 2007.

ATANCE, C. Future thinking in young children. **Current Directions in Psychological Science**, *17*(4), 295-8, 2008.

AUDI, R. Memorial justification. **Philosophical Topics**, *23*, 31-45, 1995.

BARTLETT, F. **Remembering***:* A study in experimental and social psychology. Cambridge, UK: Cambridge University Press, 1932.

BERNECKER, S. **The metaphysics of memory**. New York: Springer, 2008.

BERNECKER, S. **Memory**: A philosophical study. Oxford, UK: Oxford University Press, 2010.

BERNECKER, S. Memory in analytic philosophy. In: NIKULIN, D. (Ed.), **Memory***:* a history. Oxford, UK: Oxford University Press, 2015. p. 298-315

BERNECKER, S. A causal theory of mnemonic confabulation. **Frontiers in Psychology**, *8*, 1207, 2017.

BERNECKER, S. & MICHAELIAN, K. Editors' introduction: The philosophy of memory today. In S. BERNECKER & K. MICHAELIAN (Eds.), **The Routledge Handbook of philosophy of memory** (p. 1-3). Oxfordshire, UK: Routledge, 2017.

BRENTANO, F. **Psychology from an empirical standpoint**. Oxfordshire, UK: Routledge, 2014

BREWER, B. Perception and its object. **Philosophical Studies**, *132*(1), 87-97, 2007.

BURGE, T. Interlocution, perception, and memory. **Philosophical Studies**, *86*, 21-47, 1997.

BYRNE, A. Recollection, perception, imagination. **Philosophical Studies**, *148*(1), 15-26, 2010.

CAMPBELL, J. **Reference and consciousness**. Oxford, UK: Oxford University Press, 2002.

CRANE, T. Intentional objects. **Ratio**, *14*(4), 336-349, 2001.

CRANE, T. **The objects of thought**. Oxford, UK: Oxford University Press, 2013.

DE BRIGARD, F. Is memory for remembering? Recollection as a form of episodic hypothetical thinking. **Synthese**, *191*(2), 155-85, 2014a.

DE BRIGARD, F. The nature of memory traces. **Philosophy Compass**, *9*(6), 402-14, 2014b.

DE BRIGARD, F. Memoria. **Enciclopedia de la Sociedad Española de Filosofía Analítica**, 2018. http://www.sefaweb.es/memoria/.

DEBUS, D. Experiencing the past: A relational account of recollective memory. **Dialectica**, *62*(4), 405-32, 2008.

DEBUS, D. 'Mental time travel': Remembering the past, imagining the future, and the particularity of events. **Review of Philosophy and Psychology**, *5*(3), 333-50, 2014.

DÍAZ QUIROZ, D. F. Descartes y la memoria intelectual. **Estudios de Filosofía**, *64*, 123-38, 2021.

DOKIC, J. Is memory purely preservative? In: C. HOERL & T. MCCORMACK (Eds.). **Time and Memory**: Issues in philosophy and psychology. Oxford, UK: Oxford University Press, 2001, p. 213–32.

DOKIC, J. Feeling the past: A two-tiered account of episodic memory. **Review of Philosophy and Psychology**, *5*(3), 413-26, 2014.

DUMMETT, M. Memory and testimony. In: B. MATILAL & A. CHAKRABARTY (Eds.), **Knowing from words**: Western and indian philosophical analysis of understanding and testimony. New York: Springer, 1994. p. 251-72

FERNÁNDEZ, J. Epistemic generation in memory. **Philosophy and Phenomenological Research**, *92*(3), 620-44, 2016.

FIERRO, M. A. **Éros** el memorioso. **Revista de Psicología**, *20*(1), 243-55, 2020.

FISH, W. **Perception, hallucination, and illusion**. Oxford, UK: Oxford University Press, 2009.

FIVUSH, R. The development of autobiographical memory. **Annual Review of Psychology**, 62, 559-82, 2011.

GOLDMAN, A. **Reliabilism and contemporary epistemology**: Essays. Oxford, UK: Oxford University Press, 2012.

GUERRERO-VELAZQUEZ, C. A. Memoria y percepción en la entrevista autobiográfica: una simulación episódica que se adapta en tiempo real al contexto. **Estudios de Filosofía**, *64*, 21-45, 2021.

HASSABIS, D., KUMARAN, D., VANN, S. D. & MAGUIRE, E. A. Patients with hippocampal amnesia cannot imagine new experiences. **Proceedings of the National Academy of Sciences**, *104*(5), 1726-31, 2007.

HIRSTEIN, W. **Brain fiction**: Self-deception and the riddle of confabulation. Cambrigde, Ma: MIT Press, 2005.

HUME, D. **A treatise of human nature**. Oxford, UK: Clarendon Press, 2011.

JAMES, W. **The principles of psychology**. *Vol. I.* London : Macmillan, 1980.

KLEIN, S. B. What memory is. **Wiley Interdisciplinary Reviews***: Cognitive Science*, 6(1), 1–38, 2015.

KLEIN, S. B., LOFTUS, J. & KIHLSTROM, J. F. Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. **Social Cognition**, *20*(5), 353-79, 2002.

LACKEY, J. Memory as a generative epistemic source. **Philosophy and Phenomenological Research**, 70(3), 636-58, 2005.

LAIRD, J. **A Study in realism**. Cambridge, UK: Cambridge University Press, 2014.

LEWIS, D. **On the plurality of worlds**. Oxford, UK: Oxford University Press, 1986.

LOCKE, J. An Essay Concerning Human Understanding. In: P. NIDDITCH (Ed.). **The Clarendon edition of the works of John Locke:** An essay concerning human understanding. Oxford, UK: Oxford University Press, 1975.

MAHR, J. & CSIBRA, G. Why do we remember? The communicative function of episodic memory. **The Behavioral and Brain Sciences**, *41*, 1-93, 2018. Advance online publication.

MARTIN, C. B. & DEUTSCHER, M. Remembering. **Philosophical Review**, *75*, 161-96, 1966.

MARTIN, M. G. F. The limits of self-awareness. **Philosophical Studies,** *120*(1), 37–89, 2004.

MICHAELIAN, K. Generative memory. **Philosophical Psychology**, *24*(3), 23-342, 2011.

MICHAELIAN, K. Confabulating, misremembering, relearning: The simulation theory

of memory and unsuccessful remembering. **Frontiers in Psychology,** *7*, 1857, 2016a.

MICHAELIAN, K. **Mental time travel:** Episodic memory and our knowledge of the personal past. Cambrigde, Ma: MIT Press, 2016b.

MICHAELIAN, K., DEBUS, D. & PERRIN, D. The philosophy of memory today and tomorrow: Editors' introduction. In: K. Michaelian, D. Debus, & D. Perrin (Eds.). **New directions in the philosophy of memory**. Oxfordshire, UK: Routledge, 2018, p. 1–9.

MICHAELIAN, K. & ROBINS, S. K. Beyond the causal theory? Fifty years after Martin and Deutscher. In K. Michaelian, D. Debus, & D. Perrin (Eds.). **New Directions in the Philosophy of Memory**. Oxfordshire, UK: Routledge, 2018, p. 13-30.

PERRIN, D. Asymmetries in subjective time. In: K. Michaelian, S. Klein, & K. Szpunar (Eds.). **Seeing the future**: Theoretical perspectives on future-oriented mental time travel (p. 39-61). Oxford, UK: Oxford University Press, 2016.

PERRIN, D. & MICHAELIAN, K. Memory as mental time travel. In: BERNECKER, S. & K. MICHAELIAN, K. (Eds.). **The Routledge handbook of philosophy of memory**. Oxfordshire, UK: Routledge, 2017. p. 228-40

REID, T. **An inquiry into the human mind on the principles of common sense**. Philadelphia, PA: Pennsylvania State University Press, 2000 [1764].

ROBINS, S. K. Representing the past: Memory traces and the causal theory of memory". **Philosophical Studies**, *173*(11), 2993-3013, 2016a.

ROBINS, S. K. Misremembering. **Philosophical Psychology**, *29*(3), 432-47, 2016b.

ROBINS, S. K. Confabulation and constructive memory. **Synthese**, *196*, 2135-51, 2017a.

ROBINS, S. K. Memory traces. In BERNECKER, S. & MICHAELIAN, K. (Eds.). **The Routledge Handbook of Philosophy of Memory**. Oxfordshire, UK: Routledge, 2017b, p. 76-87.

ROBINS, S. Defending discontinuism, naturally. **Review of Philosophy and Psychology**, *11*(2), 469-86, 2020.

ROSENBAUM, R. S., KÖHLER, S., SCHACTER, D. L., MOSCOVITCH, M., WESTMACOTT, R., BLACK, S. E., ... & TULVING, E. The case of KC: contributions of a memory-impaired person to memory theory. **Neuropsychologia**, *43*(7), 989-1021, 2005.

RUSSELL, B. **The analysis of mind**. Wales, Australia: George Allen & Unwin, 1921.

RUSSELL, B. **The problems of philosophy**. Oxford University Press, 2001 [1921].

SANT'ANNA, A. & MICHAELIAN, K. Thinking about events: A pragmatist account of the objects of episodic hypothetical thought. **Review of Philosophy and Psychology**, *10*(1), 187-217, 2019.

SCHACTER, D. L., ADDIS, D. R. & BUCKNER, R. L. Remembering the past to imagine the future: the prospective brain. **Nature reviews neuroscience**, *8*(9), 657-61, 2007.

SCHACTER, D. L., ADDIS, D. R., HASSABIS, D., MARTIN, V. C., SPRENG, R. N. & SZPUNAR, K. K. The future of memory: remembering, imagining, and the brain. **Neuron**, *76*(4), 677-94, 2012.

SCHELLENBERG, S. Perceptual particularity. **Philosophy and Phenomenological Research**, 93(1), 25-54, 2016.

SENOR, T. Preservation and generation. In: BERNECKER, S. & K. MICHAELIAN, K. (Eds.). **The Routledge Handbook of Philosophy of Memory**. Oxfordshire, UK: Routledge, 2017, p. 323-34.

SUDDENDORF, T. & BUSBY, J. Making decisions with the future in mind: Developmental and comparative identification of mental time travel. **Learning and Motivation**, *36*(2), 110-25, 2005.

SUDDENDORF, T. & CORBALLIS, M. Mental time travel and the evolution of the human mind. **Genetic, Social, and General Psychology Monographs**, *123*(2), 133-67, 1997.

SUDDENDORF, T. & CORBALLIS, M. The evolution of foresight: What is mental time travel, and is it unique to humans? **Behavioral and Brain Sciences**, *30*(3), 299-313, 2007.

SUTTON, J. **Philosophy and memory traces**: Descartes to connectionism. Cambridge, UK: Cambridge University Press, 1998.

TRAKAS, M. *Rashômon*. La memoria y su conexión con el pasado. ***Ética y Cine***, *7*(3), 11-21, 2017.

TRAKAS, M. Dimensiones de análisis de los recuerdos personales como recuerdos afectivos. **Revista de Psicología**, *20*(1), 256-84, 2021a.

TRAKAS, M. Memoria y emoción: introducción al dossier. **Revista de Psicología**, *20*(1), 150-56, 2021b.

TRAKAS, M. El viaje mental en el tiempo en la filosofía y la ciencia cognitiva de la memoria. *RHV.* **An International Journal of Philosophy**, *20*, 141-63, 2022.

TULVING, E. Episodic and semantic memory. In: TULVING, E. & DONALDSON, W. (Eds.). **Organization of memory**. Cambridge, MA. Academic Press, 1972. p. 381-403

TULVING, E. **Elements of episodic memory**. Oxford, UK: Oxford University Press, 1985.

TULVING, E. What is episodic memory? **Current Directions in Psychological Science**, *2*, 67-70, 1993.

TULVING, E. Episodic memory and autonoesis: Uniquely human? In: TERRACE, H. S. & METCALFE, J. (Eds.). **The missing link in cognition:** Origins of self-reflective consciousnes. Oxford, UK: Oxford University Press, 2005.


DIAS