

Bem-vindos a uma viagem extraordinária a bordo de uma caravela filosófica, guiada pelo Autor, por "mares nunca dantes navegados". Esta Odisseia pelos mares tempestuosos da mente inclui a contribuição de 8 dos mais brilhantes especialistas internacionais sobre o cérebro e a consciência, como o Prémio Nobel da Física, Sir Roger Penrose.

Dividido em duas partes, o livro inicia com Steven S. Gouveia a traçar as coordenadas fundamentais da filosofia e da ciência da mente. A segunda parte ganha vida nos diálogos do Autor com os 8 pensadores, proporcionando uma experiência única de exploração intelectual. Prepare-se para zarpar rumo ao desconhecido, onde a ciência e a filosofia convergem numa jornada cativante pelos mistérios da mente consciente.



Steven S. Gouveia é licenciado em Filosofia. Aos 22 anos, iniciou um Doutoramento em Filosofia da Mente, com foco na Neurociência, na Universidade do Minho.

Aos 30 anos, foi nomeado Professor Honorário da Faculdade de Medicina Universidade Andrés Bello, Chile.

Actualmente, é Investigador Contratado no Mind, Language and Action Group, no Instituto de Filosofia, na Faculdade de Letras da Universidade do Porto, onde lidera um projecto de Ética da Inteligência Artificial na Medicina.

Foi investigador-visitante na Minds, Brain Imaging and Neuroethics Unit no Royal Institute of Mental Health, University of Ottawa, Canada.

Como Autor e Editor, publicou 14 livros académicos, sendo convidado para palestrar em conferências por todo o mundo.

Tem participado em inúmeros programas de televisão, rádio e podcasts. Mais informações: [www.stevensgouveia.weebly.com](http://www.stevensgouveia.weebly.com).

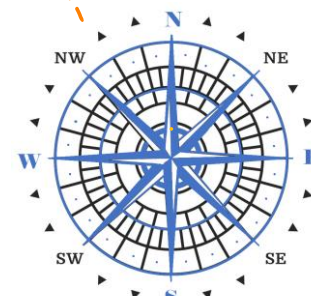
STEVEN S. GOUVEIA



David Chalmers



Nicholas Humphrey



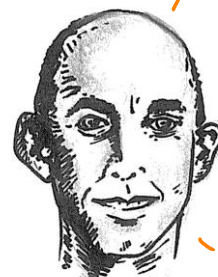
Susan Blackmore



Sir Roger Penrose

# A ODISSEIA DA MENTE

Diálogos sobre  
O  
CÉREBRO  
e a  
CONSCIÊNCIA



Anil Seth



Karl Friston



Joseph LeDoux



Christof Koch



STEVEN S. GOUVEIA

A ODISSEIA DA MENTE

# A ODISSEIA DA MENTE

Diálogos sobre

o

CÉREBRO

e a

CONSCIÊNCIA

**Steven S. Gouveia**

(autoria e organização)



2024

Steven S. Gouveia

MLAG | Instituto de Filosofia | Universidade do Porto  
Porto, Portugal

ISBN: 9798877871458

© O Autor e os Participantes, sob licença exclusiva.

Este trabalho está sujeito a direitos autorais e de tradução. Todos os direitos são únicos e exclusivamente licenciados pelo Autor e pelos Participantes, quer se trate da totalidade ou de parte do material, nomeadamente os direitos de tradução, reimpressão, reutilização de ilustrações, recitação, difusão, reprodução em microfilmes ou de qualquer outra forma física, e transmissão ou armazenamento e recuperação de informações, adaptação eletrónica, software de computador ou por metodologia semelhante ou diferente agora conhecida ou desenvolvida posteriormente.

O uso de nomes descritivos gerais, nomes registrados, marcas registradas, marcas de serviço, etc., nesta publicação não implica, mesmo na ausência de uma declaração específica, que tais nomes estejam isentos das leis e regulamentos de protecção relevantes e, portanto, gratuitos para uso geral. O Autor e os Participantes podem assumir com segurança que os conselhos e as informações contidas neste livro são considerados verdadeiros e precisos à data de publicação.

Nem o Autor nem os colaboradores oferecem garantia, expressa ou implícita, com relação ao material aqui contido ou por quaisquer erros ou omissões que possam ter sido cometidos. O Autor permanece neutro em relação às reivindicações jurisdicionais em mapas publicados e afiliações institucionais.

Capa: © Steven S. Gouveia.

Desenhos da Capa: © Ana Monteiro.

Transcrições: © Tássia Vianna.

Editora: Kindle Direct Publishing (Amazon).

# Índice

|          |   |
|----------|---|
| Prefácio | 5 |
|----------|---|

## Parte I | Consciência

|                              |    |
|------------------------------|----|
| I. Introdução: Consciência   | 11 |
| II. A Consciência Misteriosa | 27 |
| III. A Consciência Alterada  | 41 |
| IV. A Consciência Evolutiva  | 57 |
| V. A Consciência Quântica    | 67 |

## Diálogos I | Consciência

|                                     |     |
|-------------------------------------|-----|
| VI. Diálogo com David Chalmers      | 83  |
| VII. Diálogo com Susan Blackmore    | 99  |
| VIII. Diálogo com Nicholas Humphrey | 121 |
| IX. Diálogo com Sir Roger Penrose   | 145 |

## Parte II | Cérebro

|                         |     |
|-------------------------|-----|
| I. Introdução: Cérebro  | 167 |
| II. O Cérebro Preditivo | 191 |
| III. O Cérebro Livre    | 203 |
| IV. O Cérebro Integrado | 211 |
| V. O Cérebro Emocional  | 227 |

## Diálogos II | Cérebro

|                                      |     |
|--------------------------------------|-----|
| VI. Diálogo com Anil Seth_____       | 241 |
| VII. Diálogo com Karl Friston_____   | 261 |
| VIII. Diálogo com Christof Koch_____ | 281 |
| IX. Diálogo com Joseph LeDoux _____  | 299 |
| Conclusão_____                       | 319 |
| Agradecimentos_____                  | 323 |
| Biografia do Autor_____              | 329 |

# Prefácio

O objectivo deste livro é levar o leitor a bordo de uma caravela<sup>1</sup> filosófica acompanhado de alguns dos maiores e mais corajosos navegadores deste século XXI, numa valente expedição pelos mares tempestuosos do conhecimento sobre as complexas questões da mente humana.

Esta *Odisseia da Mente* levar-nos-á, inspirados pela bravura e coragem dos intrépidos exploradores dos Descobrimentos Marítimos, a desbravar as ondas e marés desconhecidas da consciência, onde cada pensamento é uma estrela que nos guia neste céu nocturno rodeado de mistérios que, talvez, nunca conseguiremos resolver.

Ao içarmos as velas da curiosidade, encontraremos certamente pelo menos alguma ilha onde poderemos pausar na viagem, e descobrir novas cores, novas

---

<sup>1</sup> A caravela portuguesa destacou-se por várias inovações, como a utilização da vela latina triangular que permitia uma maior capacidade de navegar contra o vento, ou o seu casco leve e ágil que lhe permitia uma alta versatilidade. Além disso, os portugueses também desenvolveram tecnologia de navegação mais avançada em relação à utilizada até então, através do uso da balestilha e do astrolábio, que lhes conferia a possibilidade de navegar em lugares do mundo que nunca tinham sido desbravados.

espécies, mais especiarias. Mas também, maior deslumbre e curiosidade, e, por isso, mais questões.

Talvez cada sinapse do nosso cérebro possa servir como uma bússola – como aquela no centro da capa deste livro – que aponta para direcções inexploradas e terras promissoras de maior compreensão.

Nesta busca incessante – nesta grande Odisseia do nosso século – enfrentamos um enigma central que permeia os mares da mente: o que confere à consciência a sua singularidade?

À medida que nos aventuramos por entre recifes de pensamentos e baías de reflexão, o mapa da consciência vai-se desenhando, assim, lentamente, diante dos nossos olhos.

Como os grandes navegadores que mapearam novas rotas por terras desconhecidas, enfrentamos as tormentas da incerteza e da complexidade, ao mesmo tempo que celebramos as descobertas que ampliam os nossos horizontes.

Em cada página deste livro, encontraremos ilustrações vivas de experiências e ideias que, espero, instiguem o leitor a criar a sua própria caravela e a explorar, por si mesmo, por “mares nunca dantes navegados”.<sup>2</sup>

---

<sup>2</sup> Referência à primeira estrofe do Canto I da obra *Os Lusíadas*, de Luís Vaz de Camões, poeta português que descreveu as aventuras dos Descobrimentos Marítimos pelos portugueses.

Em última instância, esta odisseia não é somente sobre desvendar o mistério da consciência nem a natureza do cérebro, mas, antes, sobre abraçar a jornada pelo desconhecido através do astrolábio da ciência e da balestilha da filosofia.

A mente, como o oceano, é vasta e inexplorada, e a nossa procura é guiada pela paixão da descoberta e pela sede de compreender um mistério incessante que parece atormentar todos os séculos da nossa humanidade.

Esperemos que, no fim desta viagem, tenhamos conseguido encontrar algum bom porto, que nos permita pelo menos um breve merecido descanso.

Para encerrar este prefácio, um apontamento menos poético: este livro tem uma estrutura específica, estando dividido em duas partes: a primeira parte, dedicada à consciência, e a segunda parte, dedicada ao cérebro.

Em cada parte, encontrará duas secções: a segunda contará com diálogos com 8 especialistas de renome internacional sobre os temas abordados. Já a primeira contará com introduções temáticas para que cada leitor possa apreciar os diálogos com maior facilidade, apesar da complexidade dos temas aqui apresentados.

Assim sendo, que comece esta *Odisseia da Mente*.

Steven S. Gouveia  
Porto, Portugal  
10 | 02 | 2024



# **PARTE I**

## **Consciência**

# **PARTE I**

## **Consciência**

## I. INTRODUÇÃO: CONSCIÊNCIA

Imagine que o leitor se encontra, neste preciso momento, numa das praias mais incríveis do planeta Terra: por exemplo, a maravilhosa praia de Copacabana do Rio de Janeiro, no Brasil.

Imagine ainda que está a beber uma saborosa caipirinha – a famosa bebida brasileira – aproveitando o dia de sol e a bela vista para o Morro do Pão de Açúcar.

Embora essa experiência seja provavelmente mágica o suficiente para a maioria de nós, há ainda outra magia a acontecer naquele momento. Qual seria esse facto? A magia de estar a saborear uma bebida específica, e experienciar alegria e felicidade, tudo isto acontecendo de um ponto de vista muito específico: o SEU ponto de vista!

Esta é a parte mágica desta história: ninguém parece poder realmente *explicar* como é que essas suas experiências conscientes estão a acontecer na sua mente. Este é um dos maiores mistérios do universo: de que forma é que um conjunto de neurónios inconscientes pode dar origem a um sentido unificado de experiência subjectiva, a uma Mente Consciente!

Por se tratar de um mistério tão grande, muitos cientistas de várias disciplinas estão a tentar aplicar todo o tipo de ferramentas e estratégias metodológicas para criar uma imagem clara de como essa sua mente é criada pelo seu cérebro, pelo seu corpo e pela sua conexão com o ambiente (o mundo!).

A consciência humana é, porventura, um dos últimos mistérios inexplicáveis do mundo actual. Claro, não é o único: muitos outros mistérios existiram no passado, como o mistério da origem do universo, o mistério da vida e da reprodução, do tempo, do espaço ou da gravidade.

Embora alguns destes mistérios possam ter em comum a inexistência de uma resposta final, podemos, apesar disso, reflectir sobre eles: esses mistérios não desapareceram, mas foram compreendidos, porque soubemos colocar as perguntas certas.

No entanto, em relação à consciência, não chegámos, aparentemente, a esse ponto: há uma terrível confusão, uma intensa luta de argumentos e contra-argumentos e diferentes teorias para explicar o mesmo fenómeno.

Recentemente, tive o privilégio de ser orador pela segunda vez numa das conferências académicas mais importantes do mundo sobre a consciência: a *Science of Consciousness Conference*.

Tendo participado anteriormente em 2019 na edição em Interlaken, na sublime Suíça, esta segunda participação realizou-se em maio de 2023, em Taormina, uma pequena e maravilhosa vila na Sicília italiana com vista para o vulcão Etna.

Ora, como orador dessa conferência, foi-me dado, no primeiro dia, um código específico e a indicação de um *website* onde eu poderia votar na teoria da consciência que acharia ser mais provável estar correcta.

Espante-se o leitor que, ao abrir a página online, me deparei com mais de 15 (!) opções para escolher: mais de 15 teorias da consciência estavam disponíveis para serem seleccionadas.

Como deve imaginar, se existem 15 teorias para explicar um fenómeno particular do mundo, então algo deve estar certamente equivocado.

Mas qual é, então, o mistério da consciência? Porque parece ser tão difícil – tanto para a ciência como para a filosofia – alcançar-se uma teoria plausível para explicar a sua existência?

Todos nós temos experiências subjectivas: percepções, sensações, dores e ideias. Como podem corpos físicos vivos num mundo físico produzir tal fenómeno?

Estas obscuridades ligadas à consciência fazem dela um dos problemas mais excepcionais e importantes da Era Contemporânea.

Alguns pensadores crêem que a consciência pode não ser assim tão misteriosa como parece à partida. Utilizando várias analogias da história da ciência, argumentam que, talvez, toda esta obscuridade seja um equívoco criado pela nossa linguagem e a forma como utilizamos os nossos conceitos mentais.<sup>3</sup>

Por exemplo, se olharmos para a forma como o conceito de “luz” foi considerado ao longo da história, talvez possamos aprender algo de relevante em relação à consciência.

A teoria corpuscular da luz, de origem grega e baseada no atomismo antigo, descreve a luz como sendo composta por partículas específicas de nome “corpúsculos”.

Esta teoria era defendida por ilustres pensadores como Sir Isaac Newton, que argumentava que todos os processos de reflexão e refração da luz só poderiam ser explicados se a luz fosse constituída de partículas, dado que a alternativa – a luz ser composta de ondas eletromagnéticas – não poderia obedecer a trajetórias em linha recta.<sup>4</sup>

Curiosamente, este argumento tinha uma forte base empírica: sabe-se que Newton teria feito cerca de 40

---

<sup>3</sup> Por exemplo: Churchland, P. (1986) *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, Cambridge: The MIT Press.

<sup>4</sup> Publicação original: Newton, I. (1704) *Opticks: or, A Treatise of the Reflexions, Refractions, Inflexions and Colours of Light*, London.

experiências que demonstravam a natureza corpuscular da luz.<sup>5</sup>

Esta teoria foi considerada verdadeira por séculos, até que o cientista Thomas Young,<sup>6</sup> tendo concebido a famosa Experiência da Dupla Fenda em 1810, vinha refutar a teoria corpuscular da luz em favor de uma teoria ondulatória, onde a luz foi considerada como tendo uma natureza igual à do som composta de ondas eletromagnéticas, o que poderia explicar muitos fenômenos – como a difração ou a interferência da luz – que a teoria anterior não poderia explicar.

Ora, neste caso, o conceito inicial de “luz” na teoria corpuscular foi, de certa forma, reduzido a um outro conceito de “luz” – radiação eletromagnética – da teoria ondulatória.

Esta redução tem dois sentidos específicos: por um lado, a teoria corpuscular no seu conjunto foi reduzida à teoria ondulatória; por outro lado, essa redução foi tão bem-sucedida que fomos capazes de reduzir o conceito de “luz” ao conceito de radiação eletromagnética.

---

<sup>5</sup> Publicação original: P. Rowlands, P. (2017) *Newton and Modern Physics*, London: World Scientific.

<sup>6</sup> Publicação original: Young, T. (2007) *Miscellaneous Works of the Late Thomas Young: Including His Scientific Memoirs*, Montana: Kessinger Publishing.

Será que só encontramos este exemplo de redução na história da ciência, em que acreditávamos que faria todo o sentido pensar a existência de determinados conceitos, mas que, afinal, descobrimos depois serem pseudoconceitos?

Segundo aqueles que acreditam que o conceito de luz da teoria corpuscular está ao mesmo nível que o conceito de consciência, há mais exemplos que podem ser oferecidos.

A teoria do flogisto foi desenvolvida no âmbito da química por Georg Stahl no século XII, que defendia que todos os corpos combustíveis teriam na sua composição um elemento de nome “flogisto” que era libertado para o ar durante os processos de combustão.<sup>7</sup>

Esta teoria era uma versão aprimorada do mentor de Stahl, Johann Becher, que teria publicado um livro de título *Physica Subterranea*, onde defendia que havia a libertação de um elemento específico quando um material ardia.<sup>8</sup>

Note-se que esta teoria do flogisto já era uma evolução de uma outra teoria antiga, a teoria dos elementos de

---

<sup>7</sup> Publicação original: Brock, W. (1992) *The Fontana History of Chemistry*, Glasgow: HarperCollinsPublishers,

<sup>8</sup> Publicação original: Taylor, S. (2010) *Alchemists, Founders of Modern Chemistry*, Montana: Kessinger Publishing.

Empédocles, que defendia existirem 5 elementos no mundo: o fogo, a terra, o ar, a água e o éter.

Novamente, esta teoria teria uma base (ingenuamente) empírica. Stahl teria feito uma série de observações nos processos realizados em metalurgia: quando um material entrava em combustão, sofria uma corrosão. Quanto maior a corrosão, maior era a quantidade de flogisto libertado por esse material.

Esta teoria foi aceite pelos pares de Stahl devido ao seu superior poder explicativo em relação à teoria de Empédocles. A teoria do flogisto permitia explicar por que é que um material orgânico perdia massa quando entrava em combustão: tal dava-se devido à perda do elemento flogisto para o ar.

A incapacidade de haver combustão sem ar era também explicada por esta teoria: tal dava-se pela impossibilidade do flogisto ser captado pelo ar. Finalmente, explicava-se o fim do processo de combustão com o esgotar da presença do flogisto no material.<sup>9</sup>

Novamente, foi preciso esperar algumas décadas para que a teoria do flogisto – e a própria relevância do conceito de “flogisto” para descrever a realidade – fosse colocada em causa.

---

<sup>9</sup> Publicação original: Crump, T. (2002) *A Brief History of Science, as Seen Through the Development of Scientific Instruments*, London: Constable & Robinson Ltd.

Um dos problemas desta teoria estava relacionado com o facto de, apesar da perda de massa ser uma realidade em matérias orgânicas, o mesmo não acontecia com os metais que, supostamente, nos processos de calcinação (i.e., oxigenação), também deveriam perder massa.

No entanto, não se verificava essa previsão particular. Pelo contrário, Antoine Lavoisier teria demonstrado que os metais, no processo de combustão, poderiam ganhar peso, o que contrariava a tese central da teoria do flogisto.<sup>10</sup>

Apesar desta evidência contrária, foi necessário esperar até à descoberta (acidental!) do oxigénio por Joseph Priestley em 1774, que teria percebido que era este elemento químico responsável pela combustão, em vez do flogisto.<sup>11</sup>

Depois de diversos progressos aos longo das décadas seguintes, a teoria química dos elementos foi-se solidificando e acabou por substituir totalmente a teoria de Stahl, que perdia qualquer relevância na descrição do mundo.

Novamente, estamos perante um conceito, “flogisto”, que parecia ter total relevância. Todavia, a investigação

---

<sup>10</sup> Publicação original: Bell, M. (2005) *Lavoisier in the Year One*, New York: Atlas Books.

<sup>11</sup> Publicação original: Jackson, J. (2005) *A World on Fire: A Heretic, An Aristocrat and The Race to Discover Oxygen*, New York: Viking.

viria a demonstrar que era um pseudoconceito, um engano linguístico que, ao deixar de ser considerado, perdeu toda a sua pertinência.

Em continuidade com a história da química – desta vez aplicada à biologia –, podemos encontrar uma teoria desenvolvida no século XVIII que defendia que o calor era um fluído invisível e sem odor que todos os corpos orgânicos possuíam na sua constituição.

A sua quantidade estava correlacionada directamente com a temperatura desse corpo: maior quantidade de fluído era equivalente a maior temperatura. Curiosamente, essa teoria teria sido avançada por aquele que teria contribuído para refutar o anterior conceito de “flogisto”: Lavoisier havia substituído esse conceito por outro, o de fluído calórico.

Como a teoria do flogisto era inconsistente com os resultados experimentais de Lavoisier, este propôs uma alternativa conceptual: a de considerar este fluído calórico como a substância do calor.<sup>12</sup>

Segundo esta teoria, o fluído calórico teria uma existência finita na sua quantidade, transferindo-se de corpos mais quentes para corpos mais frios. Além disso, não poderia ser criado ou destruído, pelo que a tese central desta teoria era a sua conservação

---

<sup>12</sup> Publicação original: Fox, R. (1971) *The Caloric Theory of Gases*, Clarendon Press: Oxford.

constante, o que explicava as interações de temperatura entre os diferentes corpos.

Ora, novamente, a incapacidade da teoria calórica de explicar variados fenômenos, como a evaporação e a sublimação, levou ao aparecimento de uma teoria mais promissora, a teoria cinética, através do trabalho de Benjamin Thompson, também conhecido como Conde Rumford.<sup>13</sup>

Rumford observou que o calor não era uma substância material com forma fluída, mas, antes, deveria ser considerado como uma energia em movimento. Tendo realizado experiências que demonstravam que o calor poderia ser gerado por atrito, desafiava a ideia de que o calor era um fluído indestrutível.

Uma das suas experiências mais interessantes envolveu o atrito gerado pela perfuração de canhões, tendo notado que a quantidade de calor produzida era muito maior do que a esperada pela teoria calórica.

Essas (e outras) observações foram fundamentais para a compreensão moderna do calor como uma forma de energia em constante transformação, contribuindo para o desenvolvimento da teoria da conservação de energia e refutando, na sua totalidade, o conceito de fluído calórico.

---

<sup>13</sup> Publicação original: Brown, G. (2001) *Count Rumford: The Extraordinary Life of a Scientific Genius – Scientist, Soldier, Statesman, Spy*, Gloucestershire: Sutton Publishing.

De forma interessante, este erro conceptual não está só presente na história da química: podemos observar o mesmo acontecimento na história da psiquiatria. A teoria da possessão demoníaca fez parte, durante séculos, da explicação apropriada para diversas doenças do foro psiquiátrico.<sup>14</sup>

Esta teoria teve grande prevalência na Idade Média, tendo tido várias versões em diferentes períodos e regiões do mundo. A possessão demoníaca funcionava como explicação para determinados comportamentos considerados anormais para os padrões daquele tempo.

O exorcismo era tido, muitas vezes, como o único tratamento disponível. Nos registos que nos chegaram, encontramos uma relação directa entre tipos particulares de demónios e sintomatologias associadas a doenças específicas, como transtornos psicóticos, neuróticos ou epilepsia.<sup>15</sup>

Curiosamente, alguns filósofos medievais que tiveram grande impacto noutras áreas da filosofia e da teologia

---

<sup>14</sup> Publicação original: Kemp, S. e Williams, K. (1987) "Demonic possession and mental disorder in medieval and early modern Europe", *Psychological Medicine*, 17 (1):21-9.

<sup>15</sup> Publicação original: Espí Forcén, C. e Espí Forcén, F. (2014) "Demonic possessions and mental illness: discussion of selected cases in late medieval hagiographical literature", *Early Scientific Medicine*, 19 (3): 258-79.

deram uma grande contribuição para a notoriedade desta teoria.

Por exemplo, Agostinho de Hipona defendia que os demónios teriam um corpo material, enquanto que Tomás de Aquino defendia que os demónios seriam entidades não-corpóreas, inteligíveis e separadas do mundo físico.

Singularmente, a posição de Aquino era substantiada pela ideia de que vários fenómenos empíricos descritos na literatura não poderiam existir caso os demónios tivessem efectivamente uma existência corpórea.<sup>16</sup>

Claro, esta teoria começou a perder toda a sua relevância com o desenvolvimento da psiquiatria como uma disciplina baseada em evidências e teorias enquadradas em conhecimentos médicos que se foram desenvolvendo no século XX e XXI, embora seja necessário indicar que – como verá em alguns dos temas abordados durante o livro – a natureza das doenças de foro psiquiátrico revela ainda grandes desafios e questões para os cientistas e filósofos.

O ponto central é que, com todos estes casos históricos agora relatados – a existência do flogisto, do fluído calórico, dos demónios – a sociedade compreendeu que, graças ao desenvolvimento do conhecimento,

---

<sup>16</sup> O'Neill, S. (2016) "Augustine and Aquinas on Demonic Possession: Theoria and Praxis", *Proceedings of the American Catholic Philosophical Association*, 90: 133-147.

passámos a estar perante conceitos destituídos de sentido e relevância para compreender a realidade.

Este ponto serve para ilustrar uma consequência essencial quando entendemos que, afinal, os conceitos usados não são úteis para compreendermos o mundo: as próprias perguntas e questões que estruturam o pensamento dos filósofos e cientistas deixam também de fazer sentido e perdem, assim, a necessidade de obter uma resposta.

Para entendermos este ponto crucial, vejamos a teoria geocêntrica que defendia a tese que o planeta Terra era o centro do universo, onde o Sol e a Lua giravam à sua volta uma vez por dia, e as estrelas estariam fixas no que era chamado de “esfera celestial”, que rodava sob o eixo da Terra.

Neste sistema geocêntrico, as grandes perguntas que todos os cientistas e pensadores procuravam responder estavam associadas às seguintes questões: o que faz a esfera celestial rodar pela Terra? Como é que as estrelas estão fixas e coladas a essa esfera? Como é que a Lua e o Sol se mantêm a rodar em torno da Terra com uma precisão tão grande?

Todas estas questões eram as grandes questões que os pensadores daquele tempo tinham de resolver. Contudo, com a refutação deste modelo, protagonizada inicialmente por Copérnico e depois por Galileu e Kepler, conjuntamente com o surgimento da mecânica

clássica de Newton, todas estas questões deixaram de fazer qualquer sentido.

Para o modelo heliocêntrico, a postulação de uma esfera celestial era simplesmente ilógica e absurda: o que parecia necessitar de uma resposta ou uma teoria deixava de requerer explicação.

Será que, com o conceito de consciência, encontramos na mesma situação que todas estas teorias históricas de variadas áreas do conhecimento humano? Será que estamos a colocar questões que podem, afinal, não justificar qualquer explicação, dado que poderão ser inquietações sem sentido?<sup>17</sup>

Esta é uma provocação filosófica com que teremos de lidar durante esta odisseia pela mente. Não podemos negar que tudo o que fazemos e sentimos é, de alguma forma, parte da mente consciente.

Podemos então estar preocupados: como sabemos tão pouco sobre algo que faz parte da nossa natureza humana? Será que a ciência contemporânea não consegue explicar a mente de forma satisfatória? Existirá alguma teoria da consciência suficientemente robusta que possa explicar todos os seus fenómenos?

Nesta primeira parte do livro, iremos tentar responder a estas perguntas dialogando com quatro dos

---

<sup>17</sup> Secção baseada em: Gouveia, S. (2022) *Philosophy and Neuroscience: a Methodological Analysis*, New York: Palgrave.

pensadores mais relevantes sobre a natureza da consciência – David Chalmers, Susan Blackmore, Nicholas Humphrey e Sir Roger Penrose – com quem tive a honra de debater.

Para que possa ter uma maior compreensão dos diálogos que se seguirão na secção II de cada parte, farei, nos próximos capítulos, uma breve introdução dos temas abordados nesses diálogos.

Comecemos, então, por compreender por que é que a consciência possui esta aura misteriosa em relação à sua existência e à capacidade de conseguirmos fornecer explicações plausíveis sobre si mesma.



## II. A CONSCIÊNCIA MISTERIOSA

O primeiro diálogo conta com a participação do filósofo que colocou o tema da consciência no mapa dos grandes debates filosóficos e científicos. Falamos de David Chalmers, Professor de Filosofia e Ciência Neural e co-diretor do Centro para a Mente, Cérebro e Consciência da Universidade de Nova York e um dos principais proponentes da ideia de que há algo de especialmente complexo sobre a natureza da consciência.

Como verá no debate, Chalmers cunhou o famoso “problema difícil da consciência”, mas o problema não é totalmente novo. No fundo, é uma característica fundamental de quem aceita o problema mente-corpo como sendo uma questão real, isto é, quem acredita que a mente tem propriedades que não podem ser esgotadas pela fisicalidade do mundo.

Muitos fenómenos mentais como aprender, raciocinar, memorizar, etc., podem ser explicados em termos de desempenhar um determinado “papel funcional”: se

descobriremos a “função” de determinado sistema, saberemos tudo o resto sobre esse sistema.<sup>18</sup>

Contudo, para Chalmers, tal não acontece com a consciência:

“O que torna o problema difícil *difícil* e quase único é que ele vai além dos problemas sobre o desempenho das funções. Observe que mesmo quando explicamos o desempenho de todas as funções cognitivas e comportamentais nas proximidades da experiência – discriminação perceptiva, categorização, acesso interno, relato verbal – ainda pode permanecer uma questão sem resposta: por que o desempenho dessas funções é acompanhado de experiência?”<sup>19</sup>

Se aceitarmos que este problema é real, uma das aparentes consequências é estarmos a defender que uma explicação científica da consciência nos irá escapar verdadeiramente.

Nesse sentido, outro filósofo fundamental para compreender a natureza do “difícil” problema da consciência é o filósofo Thomas Nagel. Nagel vê o

---

<sup>18</sup> Uma excelente introdução a este debate pode ser encontrada em: Fernandes, J. (2020) *A Velha Xereta : na Fronteira entre Filosofia e Ciência*, São Paulo: Editora Intermeios.

<sup>19</sup> Publicação original: Chalmers, D. (1996) *The Conscious Mind*, New York: Oxford University Press.

problema como a “subjectividade” dos estados mentais conscientes.<sup>20</sup>

Este filósofo argumenta que os factos sobre os estados conscientes são inerentemente subjectivos: eles só podem ser totalmente compreendidos a partir de formas limitadas de pontos de vista, apenas acessíveis pelo sujeito que está a ter essa mesma experiência consciente.

Por exemplo, no cenário inicial de estar a beber uma caipirinha na praia de Copacabana, só o leitor poderá ter um acesso imediato e subjectivo à experiência do sabor dessa caipirinha e ninguém de fora poderá compreender, exactamente, o que está a sentir naquele momento.

No entanto, a explicação científica exige uma caracterização objectiva dos factos, que se afaste de qualquer ponto de vista particular. Assim, os factos sobre a consciência iludem a ciência e tornam “o problema mente-corpo realmente intratável”.<sup>21</sup>

Chalmers acaba por introduzir um argumento original e interessante chamado “O Argumento dos Zombies Filosóficos” para demonstrar essa característica da consciência.

---

<sup>20</sup> Publicação original: Nagel, T. (1974) “What Is It Like to Be a Bat?”, *The Philosophical Review*, 83 (4): 435-450.

<sup>21</sup> Publicação original: Nagel, T. (1974) “What Is It Like to Be a Bat?”, *The Philosophical Review*, 83 (4): 435.

Este argumento prende-se com a possibilidade lógica de um mundo idêntico ao nosso sem as suas qualidades secundárias (o que chamamos em filosofia de *qualia*), sem quaisquer estados de consciência – um mundo zombie. Neste mundo, haveria um zombie idêntico a cada um dos leitores, molécula por molécula. No entanto, esse zombie não teria quaisquer experiências conscientes, embora do ponto de vista de fora, se iria comportar como se fosse consciente.

Imagine que o leitor encontra, na rua, a sua celebridade favorita por quem tem uma paixoneta desde a primeira vez que a viu a actuar no seu filme favorito. Embora não saiba, essa pessoa é um zombie filosófico.

Com todo o entusiasmo, num momento de romantismo extremo, declara-se a essa celebridade que, por espanto, reage ao seu avanço, sendo recíproco/a no sentimento sentido, dizendo-lhe que está também muito apaixonada/o e comportando-se até desse modo.

Neste exemplo claramente ficcional – infelizmente, a probabilidade de uma celebridade famosa estar apaixonada por nós é muito pequena para ser tomada a sério –, a pessoa que está a interagir connosco não sente de facto nada: ela não tem qualquer experiência consciente, sendo que, do ponto de vista do leitor, ela apenas parece ter consciência através do seu comportamento.

Partindo deste exemplo hollywoodesco, David Chalmers argumenta que podemos afirmar, indirectamente, que os zombies possuem uma possibilidade lógica – embora possam não ser uma possibilidade natural – se compreendermos a organização funcional das pessoas.

Por exemplo, podemos conceber que um isomorfo de silicone do leitor – com chips no lugar dos seus neurónios – não venha a desenvolver a sua consciência e, portanto, os factos da minha organização funcional não implicam necessariamente os factos da minha consciência.

Neste ponto, é importante realçar que a noção de zombie filosófico parece ser conceptualmente coerente e logicamente possível e, se tal for o caso, então estados conscientes e subjectivos não são idênticos aos estados físicos do cérebro.<sup>22</sup>

Tal leva-nos ainda a outro filósofo, que defende que a existência da consciência no mundo físico causa uma “lacuna explicativa” (*explanatory gap*). Joseph Levine

---

<sup>22</sup> Historicamente, a experiência mental dos zombies filosóficos tem o propósito de refutar a teoria fisicalista da mente que defende que os estados conscientes podem ser reduzidos – no mesmo sentido que descrevemos na secção anterior – a estados físicos específicos, em relações de identidade estrita.

argumenta que existe uma “lacuna explicativa” especial entre a consciência e o físico.<sup>23</sup>

O desafio de fechar essa lacuna explicativa é o que origina, de certa forma, o problema difícil da consciência. Levine argumenta que uma boa explicação científica deve implicar dedutivamente o que ela explica, permitindo-nos inferir a presença do fenómeno-alvo a partir de uma demonstração de leis ou mecanismos e condições iniciais.

A implicação dedutiva é um tipo de relação lógica com a seguinte configuração formal: se as premissas de um argumento forem verdadeiras, então segue-se que a sua conclusão é, necessariamente, verdadeira.

Note o seguinte exemplo: uma vez que descobrimos que um relâmpago nada mais é do que uma descarga elétrica, saber todas as condições adequadas para uma descarga elétrica na atmosfera no momento  $t$  permite deduzir que o relâmpago deve ter ocorrido nesse momento  $t$ .

Contudo, tanto para Levine como para Chalmers, tal não parece ser o caso quando pensamos na consciência: não importa o quão detalhada seja a nossa especificação de mecanismos cerebrais ou das leis físicas. Por mais detalhada que esta seja, há sempre uma questão em aberto sobre se a consciência está

---

<sup>23</sup> Publicação original: Levine, J. (1983) “Materialism and qualia: The explanatory gap”, *Pacific Philosophical Quarterly*, 64: 354-361.

presente e acompanha determinado mecanismo psicológico subjacente ou não.<sup>24</sup>

Claro, muitos pensadores fisicalistas – que acreditam que tudo o que existe no mundo tem de ser explicado pelas leis da física e da química – rejeitam este argumento defendendo que um mundo em que soubéssemos tudo sobre os mecanismos teria necessariamente de conter uma explicação sobre a consciência, pois esta seria necessariamente gerada a partir de qualquer conjunto de circunstâncias físicas que desconhecemos no momento, mas que nada impede de podermos vir a conhecer.

Para o leitor compreender mais facilmente este “mistério da consciência”, podemos solicitar uma preciosa ajuda da imaginária Doutora Mary, a neurocientista mais brilhante que o mundo já conheceu – infelizmente não possuímos esse tipo de conhecimento na actualidade, tratando-se de outra experiência mental.

Mary é, então, uma neurocientista genial que conhece tudo o que há para saber sobre as propriedades físicas do cérebro, principalmente as propriedades ligadas à percepção visual da cor. Além disso, Mary é tão

---

<sup>24</sup> Para maior aprofundamento, sugiro duas obras essenciais em língua portuguesa: Miguens, S. et al. (eds.) *Filosofia da Mente: uma Antologia*, Porto: U. Porto Editorial; Miguens, S. (2009) *Será que a Minha Mente está Dentro da Minha Cabeça? Da ciência cognitiva à filosofia*, Porto: Campo das Letras.

brilhante que conhece também todos os factos físicos sobre a luz e as cores.

Assim, podemos dizer que Mary sabe tudo o que há para saber sobre a forma como os seres humanos experienciam as cores visualmente, tanto os processos cerebrais envolvidos, como os processos físicos.

Contudo, há um pequeno problema: Mary é prisioneira desde o seu nascimento numa casa em que só tem acesso ao seu quarto. Pior que isso, o quarto não tem qualquer iluminação: todo o conhecimento que Mary aprendeu do cérebro e da mente deu-se num ambiente monocromático, através de material e ferramentas a preto e branco.

O que é que este cenário procura mostrar sobre a consciência? Esta experiência mental procura problematizar a seguinte situação: imagine que Mary é finalmente libertada desse quarto monocromático e um namorado, com quem trocava cartas de amor, finalmente se encontra pessoalmente com ela.

O namorado, numa acção plena de romantismo, traz uma rosa vermelha a Mary: pela primeira vez, Mary é confrontada com um objecto de cor que ela nunca viu antes, embora tenha todo o conhecimento teórico do cérebro e das cores, inclusive das rosas vermelhas.

Ora, será que Mary já saberia o que esperar ao ver, pela primeira vez, a cor vermelha da flor? Ou será que ela iria aprender algo de novo nessa experiência visual,

algo que, afinal, lhe escapava, mesmo estando na posse de todo o conhecimento de como é que o cérebro e as cores funcionam?

Esta experiência mental, desenvolvida originalmente pelo filósofo Frank Jackson, procura mostrar ao leitor, duas teses particulares.<sup>25</sup> A primeira, que estas experiências subjectivas ligados aos *qualia* têm uma existência real, não sendo uma mera fabulação teórica de alguns filósofos e cientistas. A segunda, que a tese de que a mente consciente pode ser descrita apenas apelando a propriedades físicas e químicas está equivocada, dado que, no cenário de Mary, há verdades sobre as cores que parecem não serem físicas.

Apesar deste argumento ter alguma força intuitiva, alguns pensadores<sup>26</sup> acham que não podemos efectivamente concluir da experiência de Mary essas duas teses: no máximo, a conclusão que segue, embora contra-intuitiva, é que Mary já saberia o que esperar quando o seu namorado romântico lhe ofereceu a rosa vermelha.

Claro, a aceitação destas duas teses vai depender, no fundo, da posição filosófica em que o leitor se encontra em relação à consciência: se levar a experiência de

---

<sup>25</sup> Publicação original: Jackson, F. (1982) "Epiphenomenal Qualia", *Philosophical Quarterly*, 32 (127): 127-136.

<sup>26</sup> Por exemplo: Churchland, P. M. (1985) "Reduction, Qualia, and the Direct Introspection of Brain States", *The Journal of Philosophy*, 82 (1): 8-28.

Mary a sério, acredita que há algo de fundamental na consciência que não pode ser descrito pela ciência actual.

Contudo, se achar que Mary já sabe o que esperar ao ver a rosa vermelha oferecida pelo seu namorado, então acredita que não há nada de especialmente misterioso com a consciência.

Uma área relevante onde podemos entender esta problemática do mistério da consciência está ligada aos novos desenvolvimentos tecnológicos e à possibilidade de podermos estar a viver num mundo simulado, ao estilo do filme “The Matrix”.

A possibilidade de estarmos a viver numa simulação foi defendida pelo filósofo Nick Bostrom no seu popular artigo *Are you living in a computational simulation?*<sup>27</sup> O argumento da simulação indica três possibilidades sobre o futuro do mundo:

- (i) É muito provável que a espécie humana seja extinta antes de atingir um estágio “pós-humano”;
- (ii) É extremamente improvável que uma civilização pós-humana futura possa executar um número significativo de simulações da sua história evolutiva (ou variações da mesma);

---

<sup>27</sup> Publicação original: Bostrom, N. (2023) “Are you living in a computer simulation?”, *Philosophical Quarterly*, 53 (211): 243-255.

(iii) Estamos quase certamente a viver numa simulação computacional.

Bostrom argumenta que, entre estas três proposições, pelo menos uma deve ser verdadeira. Suponhamos que (i) seja incorreta: alguma civilização no universo atinge a maturidade tecnológica.

Agora, consideremos que (ii) também seja incorreta: nenhuma civilização usa os seus recursos para realizar simulações. Se assim for, podemos, supostamente, concluir o seguinte: uma vez que essas civilizações ancestrais têm a capacidade de executar um número enorme de simulações, se as primeiras duas possibilidades forem falsas, pode existir um número maior de entidades simuladas do que de entidades não simuladas neste universo.

Todos os seres humanos estariam – pelo menos na sua maioria – a viver dentro de simulações e não fora delas. O raciocínio pressuposto no argumento é que, se rejeitarmos as duas primeiras hipóteses, teremos de aceitar que a terceira segue necessariamente, dado que para cada mundo real, teríamos milhões de mundos simulados.

Portanto, a probabilidade de estarmos actualmente numa simulação parece ser genuína, se aceitarmos as premissas avançadas por Bostrom, algo que Chalmers parece aceitar, defendendo o seguinte argumento:

(iv) É mais provável que simulações humanas conscientes sejam possíveis;

(v) É mais provável que, se simulações humanas conscientes forem possíveis, muitas populações semelhantes a humanos as criarão;

(vi) Há uma boa hipótese (25% ou mais) de sermos simulações de computador.<sup>28</sup>

Mas será que a hipótese da simulação faz sentido? Se recordarmos o épico filme “The Matrix”, há várias cenas em que Neo, a personagem principal, se encontra numa situação de total incerteza sobre se está no mundo real ou no mundo simulado.

De facto, se estivermos numa simulação, será muito difícil, do ponto de vista da nossa subjectividade, perceber isso. No entanto, alguns pensadores defendem que este argumento probabilístico falha porque ambas as suas premissas principais são implausíveis.

Ora vejamos: pode muito bem acontecer que, algum dia, os computadores tenham consciência, mas é improvável que a consciência deles seja igual à nossa, dado que os mecanismos físicos dos computadores são muito diferentes dos mecanismos neuronais – de carne e osso – que produzem a consciência humana.

---

<sup>28</sup> Publicação original: Chalmers, D. (2022) *Reality+: Virtual Worlds and the Problems of Philosophy*, New York: W. W. Norton.

A suposição de que a consciência do computador será a mesma que a consciência humana pressupõe uma independência de substrato: os estados mentais podem operar numa ampla gama de sistemas físicos independentemente da sua constituição material.<sup>29</sup>

Todavia, podemos criticar este pressuposto e defender que a consciência depende efectivamente de um substrato neurobiológico como aquele que a evolução “desenvolveu” através dos nossos corpos e cérebros, como veremos em vários diálogos deste livro. E se esse for o caso, nenhum sistema artificial poderá reproduzir de facto a consciência, pelo menos a consciência subjectiva e humana.

Além disso, podemos também colocar em causa a aceitação de que a segunda premissa (ii) seja falsa: de facto, não é assim tão óbvio por que razão uma civilização humana futura consideraria gastar inúmeros recursos para simular milhares de universos em vez de aplicá-los em outras prioridades existenciais.

Ademais, essa mesma civilização poderá considerar que tais simulações sejam eticamente imorais: se simular o universo implicar simular os seres humanos, e se a tese da independência de substrato estiver correcta – algo que não sabemos ser o caso

---

<sup>29</sup> Secção baseada em: Gouveia, S. & Neiva, D. (2017) “The Problem of Consciousness on the Mind Uploading Hypothesis” In *Philosophy of Mind: Contemporary Perspectives*, Cambridge: Cambridge Scholars Publishing.

actualmente – então simular esses seres humanos será simular também a sua consciência e, por consequência, a sua capacidade de sentir dores e de sofrer.

Se podemos simular essa capacidade senciente, então poderemos estar a cometer o maior erro ético de toda a história da humanidade (e bem sabemos que a lista desses erros já é bastante longa!): poderíamos estar a simular um número infinito de sofrimento humano que deveríamos ter a obrigação moral de evitar a todo o custo.<sup>30</sup>

Se estamos ou não a viver numa simulação, deixo ao critério do leitor. Mas lembre-se de fazer a escolha mais sábia quando for confrontado com a decisão de escolher entre a pílula vermelha – e o mundo real – ou a pílula azul – e o mundo simulado.

Por falar em consumo de pílulas, vejamos, de seguida, com a companhia da psicóloga e investigadora Susan Blackmore, outras formas mais “alternativas” de pensar e investigar o fenómeno da consciência, através dos variados estados alterados de consciência e do que estes nos podem ensinar sobre a natureza da mente consciente.

---

<sup>30</sup> Secção baseada em: Gouveia, S. (2020) “Ética da Simulação” In *Homo Ignarus: Ética Racional para um Mundo Irracional*, Coimbra: Editora Minerva.

### III. A CONSCIÊNCIA ALTERADA

O segundo diálogo conta com a participação de Susan Blackmore, escritora, conferencista, e Professora-visitante na Universidade de Plymouth, no Reino Unido. Sendo doutorada em Psicologia, procura estudar a natureza da consciência através de metodologias mais fora da caixa.

Normalmente, quando falamos de “consciência”, pressupomos pensá-la de um ponto de vista da sua normalização, isto é, pensamos este fenómeno mental no seu estado “normalizado”: quando estamos acordados a experienciar as sensações do mundo, em vez de estarmos a sonhar.

No entanto, desde que se conhece relatos de actividade humana, sabemos que existe uma categoria de estados conscientes que são... diferentes dos estados normais a que estamos habituados. Estes estados podem ser induzidos por consumo de substâncias psicoactivas, pela meditação, ou até por sonhos lúcidos, como iremos ver já a seguir.

E porquê focarmos neste tipo de estados diferentes daqueles em que estamos a maior parte do tempo? Porque podem ser úteis para identificar e comparar diferentes informações quando a consciência se

encontra num estado alterado, o que chamamos de “Estado Alterado de Consciência” (EAC).

Como podemos tentar definir estes estados? Uma forma objectiva de definir EACs é especificar de que forma é que esse estado foi induzido. Por exemplo: uma alteração de consciência provocada pelo consumo de Ayahuasca conduz a um estado de consciência diferente de uma alteração induzida por hipnose ou por meditação.

Embora esta tentativa seja uma proposta interessante, ela falha porque um estado induzido no leitor por Ayahuasca, por exemplo, terá efeitos diferentes na sua consciência do que se eu, autor deste livro, consumir a mesma droga: não há uma relação causal e necessária entre uma substância e um específico EAC.

Uma outra forma de definir EACs é baseada em medições fisiológicas que ocorrem quando estamos sob a influência de substâncias, como a temperatura corporal, o ritmo cardíaco, o consumo cortical de oxigénio, entre outros. O desafio, no entanto, reside na variabilidade desses critérios fisiológicos, que não se mantêm uniformes em todas as pessoas, tanto em termos de experiência subjectiva quanto na forma como são induzidos.

Sabemos que mudanças pequenas na fisiologia estão associadas a grandes mudanças no estado subjectivo, e vice-versa, de modo que nenhum mapeamento

directo parece ser possível. Assim, é necessário ter algum cuidado em definir estes estados muitíssimo voláteis e com uma força experiencial subjectiva enorme a partir de elementos objectivos como a fisiologia humana.

Há algumas definições que podemos encontrar em manuais de psicologia, neurociência ou filosofia como as seguintes:

- “(...) uma alteração qualitativa no padrão geral de funcionamento mental, tal que o experimentador sente que a sua consciência é radicalmente diferente da forma como funciona normalmente”;<sup>31</sup>
- “(...) uma mudança temporária no padrão geral de experiência, tal que o indivíduo acredita que o seu funcionamento mental é distintamente diferente de certas normas gerais para o seu estado de vigília normal”;<sup>32</sup>
- “(...) existe um EAC sempre que há uma mudança de um padrão normal de funcionamento mental para um estado que

---

<sup>31</sup> Publicação original: Tart, C. (1972) “States of consciousness and state-specific sciences”, *Science*, 176: 1203-1210.

<sup>32</sup> Publicação original: Farthing, G. (1992), *The Psychology of Consciousness*, Englewood Cliffs, NJ: Prentice Hall.

parece diferente para a pessoa” (Nolen-Hoeksema et al., 2014: 640)<sup>33</sup>.

Embora estas definições sejam interessantes, todas elas falham num pormenor importante: tentam definir os EACs pela perspectiva da comparação negativa a estados de consciência “normais”. O problema é que não temos, também, uma definição útil de estados normalizados e, por isso, este tipo de definições não nos ajuda a especificar, concretamente, o que poderia ser considerada a natureza comum de um estado alterado de consciência.

Uma outra alternativa é olharmos para exemplos concretos que possam aclarar um pouco esta dificuldade conceptual. Começemos, assim, por aqueles estados que o leitor mais liberal talvez já tenha experienciado: os estados induzidos por drogas.

As drogas psicoativas são todas aquelas que têm efeitos no funcionamento mental ou na consciência. Todas elas funcionam alterando a acção de neurotransmissores ou neuromoduladores endógenos.

Por exemplo, podem aumentar o efeito de um neurotransmissor estimulando a sua libertação ou bloqueando a sua recaptação para que os seus efeitos durem mais, ou pode reduzir os efeitos por inibição ou

---

<sup>33</sup> Publicação original: Nolen-Hoeksema, et al. (2014) *Atkinson & Hilgard's Introduction to Psychology*, Andover: Cengage Learning.

bloqueando a sua recepção na membrana pós-sináptica. Uma razão pela qual os efeitos das drogas que alteram a mente podem ser tão abrangentes deve-se ao facto de um único neurotransmissor poder estar activo em muitas regiões diferentes do cérebro e até do corpo.

A anfetamina é o estimulante mais conhecido e estudado e possui três efeitos principais no cérebro:

- induzir a libertação de serotonina;
- induzir a libertação de dopamina;
- inibir a recaptação da serotonina.

A serotonina desempenha um papel importante na regulação do humor e do dormir, e a dopamina ajuda a mediar o comportamento motivado pela recompensa, bem como respostas interpretativas a si mesmo, ao outro e ao ambiente.

O consumo de anfetaminas inclui consequências como aumento de energia, aumento das sensações tácteis e outras, e o aumento – tanto da frequência como da intensidade – de sentimentos de amor e empatia pelo outro.<sup>34</sup> A experiência e os seus efeitos dependem, muitas vezes, do ambiente onde a droga é consumida.

---

<sup>34</sup> Publicação original: Holland, J. (ed.) (2001) *Ecstasy: The complete guide: A comprehensive look at the risks and benefits of MDMA*, Rochester, VT: Park Street Press.

A sua toma frequente causa tolerância e, por isso, adição a longo prazo.

No entanto, em contexto terapêutico e de baixa frequência, pode ser usada para tratar condições como transtorno de stress pós-traumático e ansiedade social com alguma eficácia, quando comparando com tratamentos mais convencionais.<sup>35</sup>

A mais comum droga psicadélica é a canábis, tendo cerca de 85 componentes canabinoides. Descrever os efeitos subjectivos da cannabis não é tarefa fácil, em parte porque estes diferem muito de pessoa para pessoa. No entanto, a investigação científica tem revelado alguns efeitos comuns, como efeitos emocionais, incluindo euforia e relaxamento em doses mais baixas, e medo e paranoia em doses mais altas.

Os efeitos sensoriais incluem uma maior percepção de profundidade, uma sensação de aperfeiçoamento dos nossos sentidos, aumento da capacidade de resposta e prazer sexual, uma sensação de desaceleração do tempo e de ampliação do espaço, e um maior foco no presente.

Algumas pessoas encontram um senso do sagrado ou divino no consumo desta droga. Sabemos também que,

---

<sup>35</sup> Publicação original: Danforth, A., et al. (2016) *MDMA-assisted Therapy: a new treatment model for social anxiety in autistic adults*, *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 64: 237-249.

em algumas pessoas, o consumo pode levar ao aumento da criatividade e, em outras, pode levar a um pensamento lento com a memória a curto prazo a ser afectada negativamente. Em doses elevadas, alguns indivíduos relatam fenómenos de sinestesia, isto é, a troca de elementos sensoriais, como por exemplo, ver uma cor ao ouvir um som.

A ayahuasca é outro tipo de psicadélico cada vez mais comum e frequente. Considerada uma droga curativa, leva inicialmente a episódios de vômitos sendo que, passados alguns minutos, surge uma variedade desconcertante de sensações corporais, transformações, visões e percepções<sup>36</sup>. A contemplação da morte é comum, assim como os *insights* místicos sobre assuntos pessoais e questões existenciais profundas.

Durante o consumo desta substância, o cérebro sofre alterações significativas, como a diminuição da conectividade entre o parahipocampo e o córtex retrosplenial, que se correlacionou com relatos de dissolução do “Eu”, o que explica por que é que quem passa por uma experiência de toma de ayahuasca relata, muitas vezes, ter sentido que o seu Eu se tenha

---

<sup>36</sup> Publicação original: Luna, L. e White, S. (2016) *Ayahuasca Reader: Encounters with the Amazon's Sacred Vine*, Santa Fe: Synergetic Press.

diluído ou fundindo com o ambiente onde se encontrava<sup>37</sup>.

Finalmente, um dos mais poderosos piscadélicos é a LSD (dietilamida do ácido lisérgico). Os efeitos dessa substância abarcam uma ampla gama de experiências que incluem tanto sensações positivas, como alegria e euforia, quanto sensações negativas, como terror e desespero e a desintegração do “Eu”.

Nos primeiros estudos de imagética cerebral controlados por placebo, os participantes receberam 75 microgramas de LSD por via intravenosa: foi detectado um aumento da conectividade funcional em todo o cérebro, enquanto que efeitos locais coincidiram com mudanças na experiência. Por exemplo, alucinações visuais foram correlacionadas positivamente com um aumento de fluxo sanguíneo cerebral e da conectividade funcional no córtex visual primário (V1)

Todas estas substâncias trazem algo de novo a quem as consome: uma experiência consciente diferente e anormal, o que pode ajudar a compreender a natureza da consciência. Porquê? Porque podemos comparar, por exemplo, o cérebro normal de um sujeito sem indução de droga com um cérebro induzido por droga

---

<sup>37</sup> Publicação original: Uthaug, M. et al. (2018) “Sub-acute and long-term effects of ayahuasca on affect and cognitive thinking style and their association with ego dissolution”, *Psychopharmacology*, 235: 2979-2989.

e ganhar determinados conhecimentos relevantes sobre a actividade neuronal, o dinamismo das redes neuronais, ou as relações de determinadas regiões do cérebro.

Vejamos agora uma outra forma de pensar os Estados Alterados de Consciência, desta vez através da Meditação. Será que a meditação pode ser considerada como um EAC? Algumas definições parecem implicar que sim:

- “(...) a meditação pode ser considerada como um processo lento, cumulativo e de longo prazo para produzir um estado alterado de consciência”<sup>38</sup>.

Alguns praticantes do Budismo podem ter experiências de *kenshō* (despertar), incluindo vislumbres da suposta natureza da mente. Sabemos que quem medita sente que a sua mente foi radicalmente alterada e, neste sentido, pode ser considerada um EAC.

No entanto, há pensadores que defendem que a meditação nada mais é do que uma forma particular de... dormir. O neuropsiquiatra Peter Fenwick<sup>39</sup> mostrou que os perfis de EEG na meditação não são

---

<sup>38</sup> Publicação original: Wallace, B. e Fisher, L. (1991) *Consciousness and Behavior*, Boston, MA: Allyn and Bacon.

<sup>39</sup> Publicação original: Fenwick, P. (1987) “Meditation and the EEG” In M. West (ed.) *The Psychology of Meditation*, Oxford: Clarendon Press.

exactamente semelhantes aos de sono ou sonolência, mas muitos praticantes entram de facto num estado de “microsono” durante a meditação. Noutro estudo, os praticantes dormiram cerca de um terço do tempo durante a meditação<sup>40</sup>.

Os efeitos benéficos da meditação, como a redução da ansiedade e da depressão e a melhoria da performance cognitiva, podem ser explicados em parte pela semelhança com os benefícios conhecidos dos microsonos. No entanto, e de forma interessante, muitos praticantes relatam conseguir distinguir entre meditação profunda e estados de sono, embora essa distinção seja difícil de explicar objectivamente.

E por falar em sono, uma outra forma de pensar os Estados Alterados de Consciência são os Sonhos Lúcidos. Um sonho lúcido é caracterizado pelo facto de se saber que se está a ter um sonho nesse preciso momento. Todos os dias quando dormimos passamos por um ciclo de três estados: vigília, REM (*rapid eye movement*) e sono não-REM, sendo que uma típica noite de sono consiste em quatro ou cinco ciclos entre o sono não-REM e REM, e muitas vezes alguns micro-despertares não conscientes.

---

<sup>40</sup> Publicação original: Austin, J. (1998) *Zen and the brain: Toward an Understanding of Meditation and Consciousness*, Cambridge, MA: MIT Press.

Os estados de “acordado” e de sono são caracterizados por indicadores comportamentais, como a rapidez em despertar, os movimentos dos olhos, a tensão muscular e a actividade cerebral, que pode ser medida por meio de eletroencefalograma (EEG), entre outros métodos.

Sabemos que no sono REM, o cérebro está bastante activo e o seu EEG assemelha-se ao cérebro em estado de vigília. No sono não-REM, a taxa geral de disparo dos neurónios é tão alta quanto nos estados de vigília, mas o padrão é bem diferente, com o EEG dominado por ondas longas e lentas em vez de complexas e rápidas.

Durante o sono, o cérebro isola-se de várias formas: há uma inibição da entrada sensorial no nível tálamo-cortical durante o sono não-REM, sendo que na fase REM, essa inibição é mais periférica. Estudos com EEG e fMRI mostram que, durante estímulos auditivos na fase REM, o córtex auditivo está ainda activo, e na fase intermitente de REM, que inclui movimentos oculares e espasmos musculares, o cérebro opera num circuito fechado e encontra-se funcionalmente isolado do exterior<sup>41</sup>.

No sono REM, o tronco encefálico bloqueia os comandos motores na medula espinhal, o que impede a tradução das actividades mentais em movimentos

---

<sup>41</sup> Publicação original: Wehrle, R. et al. (2007) “Functional microstates within human REM sleep: First evidence from fMRI of a thalamocortical network specific for phasic REM periods”, *European Journal of Neuroscience*, 25: 863-871.

físicos. Entretanto, se esse mecanismo falhar, tal poderá explicar por que é que aquele tio do leitor costuma levantar-se a meio da noite para comer a sobremesa do frigorífico sem se dar conta disso: o sonambulismo explica-se exactamente quando há uma falha neste mecanismo.

Nesta fase REM, a amígdala, o hipocampo e o cíngulo anterior exibem actividade elevada, assim como partes do sistema visual e áreas de associação visual. Já o córtex pré-frontal dorsolateral, associado a funções executivas como memória de trabalho, resolução de problemas e organização motora, exibe uma actividade mais reduzida em comparação com o estado de vigília.

A investigação sobre sonhos lúcidos fornece um contexto interessante para explorar os correlatos neuronais da consciência, já que permite correlacionar variáveis fisiológicas, neuroquímicas e comportamentais com descrições subjectivas dos sonhos e pode ajudar-nos a ter uma maior compreensão da dimensão da experiência consciente.

Finalmente, uma última forma de pensarmos a natureza da consciência é através de algo muito especial que pode acontecer à nossa mente em determinado momento, como aconteceu a Professora Susan Blackmore que, aos 19 anos, se viu a experienciar uma “Experiência Fora do Corpo” (EFC) (cf. Diálogo com Susan Blackmore).

Uma EFC é uma experiência na qual uma pessoa parece experienciar o mundo de um local fora do seu corpo físico, onde se consegue observar a si mesma de um ponto de vista de fora do seu corpo. As pessoas que têm EFC relatam mais experiências psíquicas e maior crença no paranormal do que aquelas que não têm, bem como melhor recordação de sonhos e maior frequência de sonhos lúcidos<sup>42</sup>.

O que as EFC nos podem dizer sobre a consciência? Algumas pessoas utilizam estas experiências incomuns como uma prova absoluta de que a consciência é uma efectivamente alma imaterial e independente do corpo ou do cérebro. Embora seja compreensível que, para alguém com muito pouco conhecimento científico, se possa acreditar nessa tese, há efectivamente outras explicações possíveis.

Ora vejamos: se a alma conseguisse realmente ver o mundo físico durante a EFC, então estar-se-ia a admitir que ela está a interagir com esse mundo e, portanto, deve ser uma entidade física detectável. No entanto, tal contraria a suposta natureza não-física da alma. Por outro lado, se a alma é não-física, então ela não pode interagir com o mundo físico para poder observar o corpo (físico) de um ponto de vista de fora.

---

<sup>42</sup> Publicação original: Blackmore, S. (2017) *Seeing Myself: The New Science of Out-of-body Experiences*, London: Robinson.

Uma alternativa a estas teorias dualistas é argumentar que, apesar do que parece que vemos e sentimos, nada "sai" realmente do corpo. Alguns argumentos de natureza psicanalítica sugeriram que a EFC poderia reflectir o medo da morte, a regressão do ego ou a revivência do trauma do nascimento. No entanto, essas teorias são difíceis de testar cientificamente e tiveram um impacto limitado no nosso conhecimento sobre este fenómeno.

Contudo, se olharmos para a neurociência mais actual, ficamos a saber que o lobo temporal tem uma grande influência nos EFCs porque, em doentes epilépticos, a estimulação do lobo temporal causa episódios deste tipo, assim como experiências psíquicas e místicas. O neurocientista Michael Persinger<sup>43</sup> propôs que as crenças e experiências místicas seriam "criações" da função do lobo temporal do nosso cérebro, tendo conseguido induzir vários EFCs e distorções corporais ao usar a Estimulação Magnética Transcraniana que, focando-se no lobo direito, produziu autênticas EFCs aos pacientes.

A área do cérebro específica envolvida na estimulação é a junção temporoparietal direita, o que faz sentido ter impacto nas EFCs, dado que essa área é responsável por processar as informações visuais, tácteis,

---

<sup>43</sup> Publicação original: Persinger, M. (1983) "Religious and mystical experiences as artifacts of temporal lobe function: A general hypothesis", *Perceptual and Motor Skills*, 57: 1255-1262.

proprioceptivas e vestibulares e são elas que, juntas, formam o esquema corporal de cada um de nós.

Desta forma, podemos oferecer uma explicação científica a um fenómeno que apenas parece místico, mas que tem, na sua base, uma natureza perfeitamente humana.

De seguida, iremos introduzir outras ideias da psicologia informada pela neurociência: desta vez, de uma psicologia de influência mais evolutiva para tentar explicar por que razão nós, seres humanos e alguns animais, temos experiências subjectivas conscientes.



## IV. A CONSCIÊNCIA EVOLUTIVA

O terceiro diálogo conta com contribuição do reputadíssimo psicólogo e pensador Nicholas Humphrey, Professor Emérito de Psicologia na London School of Economics. Neste debate irá confrontar-se com ideias maravilhosas sobre a natureza da consciência de uma perspectiva evolutiva.

A história de Nicholas Humphrey no estudo da consciência começou, como verá no diálogo, quando este tinha apenas 23 anos e era um estudante de doutoramento num laboratório de psicologia na Universidade de Cambridge.

Durante a sua investigação, o Professor investigou inicialmente um macaco que, estando anestesiado, possuía um eléctrodo inserido no seu cérebro, mais especificamente no cóliculo superior, uma área mais “antiga” na evolução neuroanatômica do cérebro responsável pelo processamento visual.

Este colículo superior antecede o córtex visual mais desenvolvido que permite que a percepção visual consciente aconteça nos seres humanos e mamíferos próximos. Ora, o curioso deste episódio é que, embora o macaco não estivesse acordado, as células nervosas do colículo superior estavam activas, sugerindo que,

talvez, a sua visão estivesse a acontecer sem uma sensação consciente associada.

Algum tempo depois, Humphrey conheceu a macaca Helen, cujo córtex visual tinha sido totalmente removido, tendo se mantido apenas o colículo superior. Ao longo de várias interacções e muitos meses, Helen, que deveria estar totalmente cega, teria desenvolvido novamente a capacidade de ver: era capaz de pegar na sua fruta favorita, era capaz de reconhecer o Professor, entre outros factos espantosos que não deveriam estar a acontecer com um mamífero que logicamente não poderia ter qualquer visão.

Ora, o autor de *Soul Dust* começou a suspeitar que Helen poderia estar a ter percepções visuais sem ter experiências conscientes associadas: ela poderia processar informações do seu ambiente sem que criasse uma imagem consciente dessas informações.

Toda esta investigação levou o seu orientador de doutoramento, o Professor Larry Weiskrantz, a observar num paciente humano cujo córtex visual estaria reduzido a metade devido a um acidente, que conseguia identificar com alta precisão objectos que estariam presentes no seu campo visual, embora o próprio paciente admitisse nada ter visto.

Teria sido descoberto o famoso fenómeno da “visão cega”: este fenómeno ajudava a perceber muito sobre o funcionamento da visão, mas também levantava

questões fundamentais sobre a natureza da consciência, sendo que todo este episódio desencadeou em Nicholas Humphrey um incentivo e uma curiosidade enorme em contribuir no esclarecimento da existência de experiências conscientes no mundo.

A principal contribuição de Humphrey é avançar com uma teoria da consciência que pode ser explicada através da teoria evolutiva darwinista. Nas palavras do autor

“(...) argumentarei que a verdade sobre a consciência – se e quando a virmos da perspectiva certa – é que ela é de facto o produto de uma parte altamente improvável da engenharia biológica: uma obra de arte maravilhosa da natureza que dá origem a todos os tipos de impressões misteriosas em nossas mentes, mas algo que tem uma explicação física relativamente directa.”<sup>44</sup>

E eis como começa esta história maravilhosa de como é que a consciência poderá ter evoluído através da selecção natural.

O primeiro passo é argumentar que a consciência tem um impacto no nosso comportamento. E porquê o comportamento? Uma vez que a consciência, como a conhecemos, é uma característica da vida na Terra,

---

<sup>44</sup> Publicação original: Humphrey, N. (2011) *Soul Dust: the Magic of Consciousness*, Princeton: Princeton University Press.

podemos presumir que — como qualquer outra característica especializada dos organismos vivos — ela evoluiu porque confere alguma vantagem selectiva.

De uma forma ou de outra, a consciência deve estar a auxiliar o organismo a sobreviver e a reproduzir-se. Ora, tal só pode acontecer se, de alguma forma, estiver a alterar a maneira como o organismo se relaciona com o mundo exterior, algo que normalmente implica comportamento.

Mas de que forma a consciência distingue um organismo em termos de comportamento? Ao analisar as "criaturas conscientes", como lhes chama Humphrey, percebemos que elas não possuem características físicas distintivas em relação às criaturas não conscientes: a consciência não confere maior saúde, força ou beleza, mas parece exercer os seus efeitos na sobrevivência do organismo por meio do que podemos chamar de "psicologia" da criatura.

Portanto, ser fenomenalmente consciente deve estar relacionado com a forma como um organismo pensa, deseja ou acredita e o facto de ser consciente leva-o a agir de maneira adaptativa ao mundo através de comportamentos específicos que conferem uma vantagem e que podem ser identificados.

No entanto, apesar desta "vantagem" adaptativa, não é ainda claro por que razão a consciência foi seleccionada como elemento relevante pela selecção natural. Este é

o ponto central da teoria evolutiva da consciência: se a selecção natural pode “ver” os efeitos – sejam eles quais forem – da mudança psicológica no comportamento, presumivelmente, outros observadores externos também podem perceber as vantagens adaptativas de ser consciente.

Estamos, assim, no caminho certo para criar uma história plausível sobre a evolução da consciência, uma vez que, além de se apresentar uma vantagem inerente a cada organismo, essa vantagem intrínseca influencia o comportamento externo e isso possibilita a sua detecção resultando numa vantagem adaptativa identificável.

Mas de que forma, evolutivamente, a consciência surgiu para ter chegado como uma característica fundamental do ser humano nos nossos dias? Humphrey acredita que o surgimento da consciência ocorreu rapidamente e foi um fenómeno de “tudo ou nada”, tendo aparecido mais tarde na história evolutiva, quando os nossos ancestrais homínídeos desenvolveram diversas habilidades sociais como imitação, decepção, a linguagem.

Assim, podemos argumentar que a consciência é uma propriedade emergente que evoluiu pela sua função social, pela capacidade de entender, prever e manipular o comportamento de outros indivíduos. Foram esses indivíduos que a evolução “favoreceu” em relação aos que não possuíam essas capacidades.

Como as espécies de grandes símios que vivem hoje, nós, humanos, sempre vivemos em grupos sociais complexos nos quais conhecer as intenções de outros indivíduos pode ser extremamente útil para determinar quem está acima de nós na hierarquia social, em quem podemos confiar, com quem podemos formar alianças, e assim por diante.

Assim, de acordo com Humphrey, os ancestrais que eram capazes de entender, prever e manipular o comportamento de outros indivíduos tinham uma vantagem adaptativa definitiva. Neste ponto, pode ser argumentado que os humanos poderiam muito bem ter adquirido essas habilidades simplesmente a observar o comportamento dos outros e as suas consequências de fora, um pouco como os behavioristas.

Mas o Professor Emérito da Universidade de Cambridge acha que haveria uma melhor forma de conseguir esse resultado: através da hipótese de que os indivíduos adquiriram a capacidade de olhar para si mesmos, de se colocar no lugar de outra pessoa e tentar "observar" a sua mente.

Humphrey compara essa capacidade a um "olho interno" direccionado não para o mundo exterior, como a maioria dos órgãos sensoriais, mas para o mundo interior do indivíduo. Esse "olho interno" não consegue observar o funcionamento neuronal do cérebro, mas, em vez disso, percebe uma versão

psicológica mais acessível dessa actividade, os estados conscientes subjectivos.

Poderá tal significar que a consciência pode ser considerada uma “invenção”, no sentido que não existia num determinado período do desenvolvimento humano, e que passou a existir depois?

Para Humphrey, a resposta é dupla: a consciência é uma faculdade cognitiva, desenvolvida pela selecção natural, projectada para nos ajudar a dar sentido a nós mesmos e ao que nos rodeia; ao mesmo tempo, é uma fantasia, criada pelo cérebro, projectada para mudar o valor que atribuímos à nossa existência.

Contudo, se tal é o caso, corremos o risco de estar a cometer o erro do “teatro cartesiano” como apontado por Daniel Dennett, esta ideia de que há um lugar no cérebro onde o “filme” do mundo exterior acontece.

Esta abordagem à consciência não implica necessariamente uma tese ilusionista da consciência – esta ideia de que a consciência é uma ilusão criada pelo cérebro – porque o que acontece dentro do “teatro” não é de facto uma replica do mundo exterior, mas do mundo “interior”, isto é, do mundo da subjectividade e dos *qualia*.

É certo que o autor de *Sentience* começa por negar o “realismo” dos *qualia* tal como são defendidos por filósofos como David Chalmers ou Thomas Nagel que, como vimos anteriormente, os consideram como

existindo de forma fundamental e independente, acabando por se aproximar de uma teoria ilusionista.

Esta abordagem defende que, apesar das experiências subjectivas conscientes parecerem ter essas propriedades não-físicas maravilhosas, tal só pode ser o caso porque o cérebro está a fazer-nos uma “partida”.

Tal é possível porque o cérebro é um motor computacional que lida com símbolos, e símbolos com base física podem perfeitamente representar estados de coisas que não representam, e até mesmo que poderiam não existir.

Como afirma Daniel Dennett, considerado um dos maiores representantes da teoria ilusionista: “A consciência é uma ilusão do cérebro, para o cérebro, pelo cérebro”.<sup>45</sup>

Assim, quando temos uma experiência subjectiva consciente, o nosso próprio cérebro está a “montar” um espectáculo de magia que nos faz acreditar que estamos a experienciar algo fora de nós. De certa forma, esta abordagem enfraquece o mistério da experiência humana, do facto de sermos únicos, mágicos e excepcionais no universo.<sup>46</sup>

---

<sup>45</sup> Publicação original: Dennett, D. (1991) *Consciousness Explained*, New York: Little Brown.

<sup>46</sup> Publicação original: Humphrey, N. (2020) “The Invention of Consciousness”, *Topoi*, 39: 13-21.

Mas para o autor do *Soul Doust*, tal pode ser também uma vantagem: assim que percebermos que não há nada de especialmente misterioso na existência da consciência, poderemos usar essa informação como trunfo para nos “engradecer” como seres humanos, através das artes, da ciência, e de superar uma existência meramente “terrena”.

Para concluirmos esta nota introdutória, vejamos as palavras do próprio pensador sobre a excepcionalidade da consciência no mundo:

“Mesmo que a ideia de que uma característica naturalmente evoluída com “intenção” esteja errada, imagino que o próprio Darwin poderia ter visto a consciência fenomenal como uma conquista “última” – a glória culminante do processo evolutivo que começou com o Big Bang. É uma invenção tão sublime que, se deixasse de existir, diminuiria de facto o propósito de toda a criação.”

Este ponto de vista evolutivo da consciência traz-nos de facto uma abordagem interessante que pode ser estudada empiricamente. No entanto, não esclarece, ao certo, de que forma é que o nosso cérebro produz especificamente esta propriedade emergente que é a subjectividade consciente.<sup>47</sup>

---

<sup>47</sup> Secção baseada em: Humphrey, N. (2023) *Sentience: the Invention of Consciousness*, Cambridge, MA: MIT Press.

Para fecharmos esta primeira parte do livro dedicada à consciência, irei introduzir as ideias do último convidado com quem debati este tema, o Professor Sir Roger Penrose.

## V. A CONSCIÊNCIA QUÂNTICA

O quarto e último diálogo é, porventura, aquele que o leitor poderá ter maior dificuldade em seguir. De modo a prepará-lo para uma das conversas mais estimulantes deste livro, tentarei introduzir alguns dos conceitos centrais desenvolvidos pelo Prémio Nobel da Física 2020, Sir Roger Penrose, Professor Emérito da Universidade de Oxford.

De que forma é que podemos pensar a consciência? Alguns cientistas e filósofos ponderam a consciência como uma propriedade emergente da computação entre os neurónios que se interconectam e alternam em sinapses mediadas quimicamente. O problema da consciência, aqui, é o problema de descobrir que computação acontece ao certo e se podemos replicar esse tipo de procedimento.

O cérebro, nesta perspectiva, é uma espécie de computador digital que processa informação do ambiente para a mente, e a consciência terá algum tipo de papel nesse processo.

Contudo, a noção de “computação” neste sentido clássico é bastante problemática: para o físico britânico, o cérebro está numa escala de processamento totalmente diferente de um computador digital actual.

Assim, a abordagem de Penrose pressupõe ampliar o foco de poder explicativo da física clássica/neurociência computacional tradicional (i.e., o neurónio) para um nível mais “pequeno”.

Através de alguns elementos da física quântica, propõe-se uma teoria da consciência com uma explicação quântica. Para alcançar este passo, Penrose irá criticar a ideia de que as propriedades quânticas deixam de ter relevância quando “subimos” na escala de complexidade da biologia, que defende que o neurónio, as interacções neuronais, o cérebro, etc., não são sujeitos apropriados para serem interpretados pela mecânica quântica (MQ), sendo que devem ser interpretados pela física clássica.

Contrariando esta tese, um grupo de cientistas<sup>48</sup> mostrou que a MQ tem um papel fundamental na fotossíntese, descobrindo que nas células de clorofila, a transferência de luz está perto de 100 por cento de eficiência, sendo que tal supera largamente qualquer tipo de tecnologia humana moderna.

Além disso, também sugere que deve haver algum tipo de “espaço seguro” da MQ para manter a coerência do sistema, de modo a conseguir alcançar esse tipo de eficiência.

---

<sup>48</sup> Publicação original: Engel, G. et al. (2007) “Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems”, *Nature*, 446 (2007): 782-786.

Porque precisamos de ir além da computação digital para alcançarmos uma teoria viável da consciência? Penrose argumenta que a física clássica possui três pressupostos que impedem uma explicação da consciência humana:

- Determinismo causal: podemos conhecer todas as condições iniciais dos sistemas físicos e, com isso, podemos prever o seu comportamento futuro a partir de uma cadeia causal;
- Localidade/independência: dois sistemas separados no espaço não podem interagir instantaneamente;
- Realismo objectivista: objectos na realidade externa existem com propriedades bem definidas e independentes de eventuais observadores.

No entanto, o Prémio Nobel da Física defende que a consciência é não-computacional (i.e. não-algorítmica) como consequência directa do teorema de Gödel.<sup>49</sup>

Penrose demonstrou como a qualidade mental da 'compreensão' (*understanding*) não pode ser encapsulada por nenhum sistema computacional e deve derivar de algum efeito 'não-computável'. O ingrediente não-computável necessário para a

---

<sup>49</sup> Publicação original: Gödel, K. (1931) "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I", *Monatshefte für Mathematik und Physik*, 37: 349–360.

consciência e compreensão humana, sugere Penrose, terá de estar numa área onde as nossas teorias físicas actuais estão fundamentalmente incompletas, embora de importante relevância para as escalas pertinentes ao funcionamento dos nossos cérebros.

A única possibilidade “séria” é a Incompletude da Teoria Quântica, uma incompletude que tanto Albert Einstein quanto Erwin Schrödinger já haviam reconhecido, referida como o Problema da Medição. Uma maneira de resolver este problema seria fornecer uma extensão da estrutura-padrão da MQ, introduzindo uma forma objectiva de redução de estado quântico – denominada de “redução objectiva” (RO).

Tal significa que é a consciência que reduz os estados quânticos? Nesta versão que chamamos da interpretação clássica da MQ, é o acto de observação do sujeito que define o estado quântico e viola o princípio da superposição: o gato de Schrödinger<sup>50</sup> estaria morto ou vivo até que alguém (i.e., um observador) observa o sistema, definindo a superposição numa posição particular.

Para entender a superposição da perspectiva clássica, imagine que atira uma moeda ao ar para escolher o que vai jantar: se calhar coroa, será seitan panado; se calhar

---

<sup>50</sup> Publicação original: Schrödinger, E. (1935) “Die gegenwärtigen situation in der quantenmechanik”, *Naturwissenschaften*, 23: 807-812, 823-8, 844-9.

cara, será raclette. Logo depois de atirar a moeda ao ar, durante os poucos segundos em que ela se encontra no ar, a moeda estará num estado de superposição: ela está simultaneamente em ambos lados, cara e coroa. O lado da moeda só fica definido quando o leitor pegar na moeda e observar o lado (que seja cara!). As partículas sub-atômicas são como esta moeda: encontram-se em estados de superposição até que uma observação as faça colapsar num determinado estado.

No entanto, a visão de Penrose é exactamente oposta a esta descrição: para ele, há um choque entre a mecânica quântica e a teoria da relatividade geral de Einstein<sup>51</sup> que pede que algo ‘novo’ aconteça.

Na visão do Prémio Nobel da Física, a redução do estado acontece não por observação, mas espontaneamente, e é neste processo que a consciência aparece. Mas será que tal faz sentido? Repare-se: o nível da MQ só funciona como teoria num nível muito “pequeno” (sub-atómico) e o nível do processamento de informação neuronal acontece numa escala bastante mais elevada.

Se a física quântica chega a níveis mais “elevados”, passa a encontrar-se num contexto “quente”, “húmido”, “ruidoso” devido ao contacto com o ambiente, levando

---

<sup>51</sup> Publicação original: Einstein, A. (1916) “Die Grundlage der allgemeinen Relativitätstheorie”, *Annalen der Physik*, 49 (7): 769-822.

necessariamente ao colapso dos estados quânticos e à sua decoerência quântica.

Essa decoerência poder ser explicada no seguinte exemplo: imagine que o gato do leitor é um gato quântico que está simultaneamente vivo e morto de acordo com o princípio da superposição. Se o seu gato interagir com partículas do ambiente (por exemplo, com fótons), a decoerência irá acontecer: dado que o ambiente "observa-o" indirectamente, essa observação faria com que o seu gato assumisse um estado definido, vivo ou morto, em vez de existir em ambos os estados ao mesmo tempo (esperemos, claro, que seja a primeira opção quântica).

Assim, seguindo esta linha de raciocínio, parece impossível que o nível da física quântica possa funcionar no nível da informação do cérebro – entre neurónios – e possa produzir a consciência. Será que tal é um obstáculo à teoria da consciência de Penrose?

Para o físico de Oxford, esta crítica não funciona por razões, curiosamente, evolutivas: a biologia evoluiu durante milhões de anos para resolver o problema da decoerência quântica, nomeadamente através do desenvolvimento de uma estrutura sub-atômica do neurónio, os chamados "microtúbulos" que, devido à sua estrutura peculiar, permitem garantir a coerência quântica – como acontece com o processo natural da fotossíntese que referimos anteriormente.

A referência aos microtúbulos não é original de Penrose, mas de um colega seu, o anestesiológico americano Stuart Hameroff, que propôs a Penrose que a coerência quântica no cérebro poderá acontecer nesses microtúbulos, que são estruturas proteicas e que possuem uma estrutura tubular dentro de células eucarióticas (parte do citoesqueleto), que são responsáveis por várias funções: determinar a forma da célula, coordenar o seu movimento, coordenar a divisão celular, entre outras.

Hameroff sugere que os microtúbulos são o dispositivo quântico que Penrose estava à procura na sua teoria, dado que estes ajudam a controlar a força das conexões sinápticas, e a sua forma tubular pode protegê-los do ruído circundante do neurónio maior. Imagine que os neurónios são pequenas fábricas: os microtúbulos seriam os trilhos que ajudam a guiar e organizar o movimento dentro dessas fábricas.

Contudo, é necessário mais do que apenas uma variedade contínua de momentos aleatórios de coerência quântica para ter algum impacto na consciência: são esses momentos de percepção consciente que, orquestrados pelos microtúbulos nos nossos cérebros, têm a capacidade de armazenar e processar informações e memória.

Para Hameroff e Penrose, os microtúbulos podem preservar suficientemente a coerência quântica até atingirem o nível neuronal: para a consciência, é

necessário que muitos microtúbulos em vários neurónios diferentes actuem de uma forma ‘orquestrada’.

Qual é, então, a diferença entre estados quânticos normais, e estados quânticos que levam à consciência? Para esta dupla de cientistas, é a coerência global: daí defender-se a redução objectiva (RO), quando os estados quânticos colapsam numa opção ou noutra.

A parte “orquestrada” é avançada para garantir que é necessário um trabalho de conjunto entre diversos microtúbulos para influenciar o nível neuronal.

Para resumir as ideias até então, temos a seguinte configuração teórica-conceptual:

- pressupõe-se um modelo segundo o qual a consciência surge através de efeitos quânticos que ocorrem dentro de estruturas sub-celulares internas aos neurónios conhecidos como microtúbulos;
- este modelo postula os chamados “colapsos objectivos” que envolvem o sistema quântico a mover-se de uma superposição de vários estados possíveis para um único estado definido, mas sem a intervenção de um observador ou medição como na maioria dos modelos da mecânica quântica;
- de acordo com Penrose e Hameroff, o ambiente interno aos microtúbulos é

especialmente adequado para tais colapsos objetivos, e os autocolapsos resultantes produzem um fluxo coerente que regula a actividade neuronal e possibilita processos mentais não-algorítmicos.

Nesta parte, é importante introduzir a chamada “Interpretação de Penrose” da mecânica quântica, dado que esta é fundamental para compreender a sua teoria da consciência.

A interpretação de Penrose é uma especulação sobre a relação entre a MQ e a relatividade geral de Einstein que propõe que um estado quântico permanece em superposição até que a diferença de curvatura do espaço-tempo atinja um nível significativo para o colapsar (o chamado “autocolapso”).

Esta perspectiva é uma alternativa à “Interpretação de Copenhaga”, que postula que a superposição falha quando uma observação é feita (mas que não é de natureza objectiva) e é também uma alternativa à “Interpretação dos ‘Muitos Mundos’”, que afirma que os resultados alternativos de uma superposição são igualmente “reais”, enquanto que a sua decoerência mútua impede interacções observáveis subsequentes.

A interpretação de Penrose nega estas duas interpretações que dependem de observadores (ambas são teorias subjectivas), defendendo um tipo de teoria objectiva do colapso. Para esta teoria, a função

de onda é uma onda física: assume-se o colapso da função de onda como um processo físico, sendo que os observadores não têm nenhum papel especial ou causal.

Penrose teoriza que a função de onda não pode ser sustentada numa superposição além de uma certa diferença de energia entre os estados quânticos, sendo que esse *threshold* estará, então, relacionado com as influências gravitacionais das partículas.

A proposta da “Redução Objectiva” teria o seu início determinado por uma condição referida como critério de “one-graviton”. A proposta de Diósi-Penrose fornece um limiar físico objectivo, indicando um tempo de vida plausível para estados quânticos superpostos. Esta proposta sugere que cada evento RO, que é um processo puramente físico, é em si mesmo um tipo primitivo de “observação”, um momento de “experiência protoconsciente”.<sup>52</sup>

Para tal, é necessário que a superposição evite a decoerência ambiental imediata e se mantenha até atingir um determinado limite temporal. Tal é alcançado, então, defendendo-se que uma superposição quântica é:

---

<sup>52</sup> Publicação original: Hameroff, S. & Penrose, R. (2014) “Consciousness in the universe: A review of the ‘Orch OR’ theory” *Physics of Life Review*, 11 (1): 39-78.

- ‘orquestrada’, ou seja, adequadamente organizada, imbuída de informações cognitivas e capaz de integração e computação;
- isolada de um ambiente aleatório não-orquestrado por tempo suficiente para a superposição evoluir para o colapso e, nesse momento, criar um momento de consciência.

O problema desta abordagem é que a equação de Schrödinger é considerada uma equação que descreve o formalismo quântico de um sistema a temperatura zero: seria absurdo considerar que um cérebro consciente esteja no ambiente térmico aproximado de zero (pelo contrário, está longe disso!).

Sabemos também que os computadores quânticos actuais exigem temperaturas muito próximas de zero graus na escala de Kelvin para serem funcionais, pelo que se segue que considerar actividades quânticas relevantes no cérebro a temperaturas longe de zero pode ser altamente problemático e um obstáculo a toda a teoria avançada por Penrose.

Ora, o interessante é que, novamente, parece que a biologia e natureza já arranjam várias formas de desenvolver mecanismos térmicos específicos para promover a coerência quântica e evitar a decoerência.

Como já apontámos, há evidências que mostram que as plantas usam rotineiramente o transporte de

electrões (com coerência quântica) à temperatura ambiente na fotossíntese.<sup>53</sup> Mas... e no cérebro humano? Há alguma evidência?

Em 2009,<sup>54</sup> Anirban Bandyopadhyay e colegas do Instituto Nacional de Ciências dos Materiais no Japão usaram nanotecnologia para abordar as propriedades eletrônicas e ópticas de microtúbulos individuais, tendo descoberto que os efeitos quânticos podem ocorrer efectivamente nos microtúbulos a temperaturas biologicamente relevantes (i.e., longe do zero) e que a existência de estados quânticos coerentes em microtúbulos a temperaturas do cérebro são uma possibilidade real.<sup>55</sup>

Além disso, há uma outra evidência curiosa vinda da biologia: não de um organismo complexo e altamente desenvolvido, mas de um organismo unicelular muito simples de nome “paramecium” que não possui um único neurónio, mas, no entanto, consegue deslocar-se, procriar e alimentar-se. De que forma? Utilizando estruturas específicas chamadas... microtúbulos.<sup>56</sup>

---

<sup>53</sup> Publicação original: Hildner, R., Brinks, D. et al. (2013) “Quantum coherent energy transfer over varying pathways in single light harvesting complexes”, *Science*, 340 (639): 1448-1451.

<sup>54</sup> citar

<sup>55</sup> Publicação original: S. Sahu, S. Ghosh, K. Hirata, D. Fujita, A. Bandyopadhyay (2013) “Multi-level memory-switching properties of a single brain microtubule”, *Applied Physics Letters*, 102: 123701.

<sup>56</sup> Publicação original: Nakagaki, T., Yamada, H. e Toht, Á. (2000) “Maze – solving by an amoeboid organism”, *Nature*, 407: 470.

Embora esta teoria da consciência seja de facto complexa, ela tenta resolver dois dos maiores problemas científicos e filosóficos do século XXI: explicar de que forma é que a consciência existe no ser humano tentando unificar duas teorias que parecem não ser conciliáveis numa “Teoria do Tudo” – a mecânica quântica e a teoria da relatividade de Einstein.

Com esta proposta, Penrose aponta algumas luzes sobre como resolver estes dois problemas de uma só vez. Será que o leitor ficou convencido?

Depois destas notas introdutórias, espero que o leitor possa seguir com mais facilidade os diálogos com estes quatro pensadores incríveis com quem tive a honra de pensar a natureza da consciência.

Não sei se ficará mais esclarecido que confuso, mas uma confusão clareada é sempre melhor do que uma certeza obscura.



# DIÁLOGOS I

## Consciência

# **DIÁLOGOS I**

## **Consciência**

## VI. Diálogo com David Chalmers



David Chalmers é Professor de Filosofia e Ciência Neural e co-diretor do Centro para Mente, Cérebro e Consciência da Universidade de Nova York. É também Distinguished Professor of Philosophy na Australian National University.

Doutorou-se em Filosofia e Ciência Cognitiva na Universidade de Indiana no grupo de investigação de Inteligência Artificial de Douglas Hofstadter.

Chalmers é conhecido pela formulação do "problema difícil" da consciência e do seu trabalho sobre "a mente estendida", a ideia de que a tecnologia que usamos pode literalmente tornar-se parte de nossas mentes.

O seu trabalho em linguagem, metafísica, computação e inteligência artificial também atraiu muito interesse.

É cofundador e ex-presidente da Association for the Scientific Study of Consciousness e é codiretor da PhilPapers Foundation.

É autor de *Reality+: Virtual Worlds and the Problems of Philosophy* (2022), *The Conscious Mind* (1996), *The Character of Consciousness* (2010) e *Constructing the World* (2014).

Mais informações: <https://consc.net/>

**Pergunta:** No seu livro *A Mente Consciente* (1996) avançou com o chamado “problema difícil da consciência”, em contraposição aos “problemas fáceis da consciência”, estes ligados aos correlatos neuronais da consciência e com a sua implicação no comportamento. Passados estes anos todos, ainda acha que o problema da consciência é tão “difícil” como era nessa altura?

**David Chalmers:** Diria que estamos a progredir na compreensão do problema através de várias vertentes, pelo que podemos afirmar que há, definitivamente, um progresso no estudo da natureza da consciência. Dito isso, não acho que alguém tenha resolvido o problema “difícil” ainda, e acho que continua a ser fundamentalmente um problema muito difícil da ciência conseguir resolver.

A distinção básica que avancei é entre os problemas “fáceis” da consciência, que são os de explicar várias funções comportamentais e cognitivas, para as quais temos um paradigma para explicá-las, e o problema “difícil”, que é o problema de explicar experiência subjectiva, para a qual não temos o mesmo paradigma, pois parece que os métodos-padrão actuais da ciência cognitiva deixam em aberto a questão de por que toda a nossa biologia deveria dar origem a experiências subjectivas conscientes em primeira instância.

Ademais, considero ainda que esta distinção básica é relevante na medida em que se alguém oferecer uma

explicação de ciência cognitiva comum, tal não resolverá o problema difícil. Então, precisamos de algo novo.

Dito isto, a ciência da consciência tem-se desenvolvido muito bem sem ter resolvido o difícil problema através de, por exemplo, procurar os correlatos neuronais da consciência, ou talvez até conseguir avançar teorias da consciência como a Teoria da Informação Integrada (*Integrated Information Theory*), que não tenta reduzir a consciência, mas tenta conectá-la às propriedades físicas e ao cérebro.

Em última análise, pode ser que o melhor que possamos conseguir alcançar em relação ao problema “difícil” seja algo como um esboço dos princípios fundamentais que conectam a consciência aos processos físicos, e talvez teorias como a TII estejam a tentar alcançar esses princípios. Não creio que, no entanto, e para que fique claro, nenhuma teoria até então tenha alcançado o tipo de evidência e consenso necessários para ser realmente aceite como teoria da consciência.

Enquanto isso, variados cientistas e filósofos têm explorado muitas ideias diferentes que são muito melhor compreendidas hoje do que há 30 anos, seja o pampsiquismo, a abordagem da mecânica quântica ou a teoria do ilusionismo. Acho que um progresso importante foi feito em cada uma destas teorias, mas,

fundamentalmente, acho que o problema “difícil” é tão difícil quanto sempre foi e continuará a ser.

**Pergunta:** Como conceberia uma ciência rigorosa de modo a ser considerada uma abordagem sólida à natureza da consciência? Estaria tal relacionado com possuir uma forte validade matemática e formal? Ou acha que também seria necessário aliar dados empíricos e evidências à prova matemática? Ou seja: que tipo de critérios considera relevantes para conseguirmos desenvolver uma ciência séria da consciência?

**David Chalmers:** Vejo a ciência da consciência como uma integração de dados de terceira pessoa, isto é, descrições objectivas de sistemas cognitivos, com dados de primeira pessoa, isto é, o tipo de dados que podemos obter pela experiência subjectiva.

Podemos obter dados objectivos através de métodos-padrão da ciência contemporânea, como uma observação de medição, especialmente do comportamento e do cérebro. E podemos obter dados de primeira pessoa ao considerar a experiência subjectiva. Assim, o que queremos de uma ciência da consciência é que possa alcançar, em última análise, princípios que possam conectar o reino da primeira pessoa ao reino da terceira pessoa, formulados de maneira rigorosa.

Tal irá exigir que uma ciência da consciência seja realmente rigorosa em várias etapas. Em primeiro lugar, serão necessários métodos para colectar dados da primeira pessoa que sejam tão rigorosos quanto os nossos métodos para colectar dados em terceira pessoa e, embora vários investigadores tenham reflectido bastante sobre este ponto, os nossos métodos para alcançar dados da primeira pessoa são ainda muito primitivos.

Precisaremos, também, de métodos para formular a estrutura da consciência em termos rigorosos. O Steven mencionou a matemática: eis um lugar onde considero que a matemática possa ter relevância. Podemos, por exemplo, trabalhar para encontrar descrições matemáticas da estrutura da consciência. E pode ser que, pelo menos parcialmente, princípios matemáticos conectem a estrutura dos processos físicos (isto é, dados de terceira pessoa) à estrutura da experiência subjectiva (isto é, dados de primeira pessoa).

Acho que podemos encontrar uma aproximação de tal abordagem na Teoria da Informação Integrada: acho que é uma metodologia muito promissora para uma ciência da consciência. No entanto, devo ser claro em afirmar que uma descrição matemática da consciência não esgotará toda a natureza da consciência.

A experiência mental da “Mary, a brilhante neurocientista” procura mostrar um pouco este ponto:

Mary poderia conhecer a estrutura matemática da consciência de antemão, mas a experiência de, pela primeira vez, ser confrontada com a cor vermelha é uma experiência consciente nova, que não se podia reduzir a todo o conhecimento que ela já possuía sobre ver cores.

Assim, acho que a estrutura matemática da consciência é, na verdade, pelo menos uma caracterização parcial muito boa da consciência. E isso pode fornecer o que algumas pessoas chamam de “fenomenologia objectiva”, que pode desempenhar um papel crucial no desenvolvimento da ciência da consciência.

**Pergunta:** Em relação ao seu novo livro *Reality+: Virtual Worlds and the Problems of Philosophy* (2022), apresenta a “hipótese da simulação” que, muito brevemente, defende que os cenários apresentados pela ficção científica como o filme “The Matrix” podem ter uma maior plausibilidade do que parecem ter à partida, ou seja, que talvez possamos estar, neste momento, a viver num mundo simulado. Essa simulação poderá ter sido criada por uma civilização ancestral que foi capaz de desenvolver esse tipo de tecnologia com capacidade para simular mundos inteiros. Focando nesta ideia de simulação, argumenta que seres simulados podem ser tão conscientes quanto nós e gostaria de saber por que razão acha que tal é o caso?

**David Chalmers:** Podemos voltar parcialmente à velha questão de saber se necessitamos da biologia (de base carbónica) para produzir consciência, ou se é mais uma questão de processamento de informação, da computação propriamente dita e da organização funcional do sistema. Sempre estive do lado dos filósofos e cientistas que afirmam que o que realmente importa para a consciência não é a biologia específica, mas algo mais como a organização funcional e/ou o processamento da informação.

A forma como eu levantei este debate no meu livro *The Conscious Mind* nos anos 90, foi imaginar esta experiência de pensamento de substituir gradualmente os neurónios por chips de silício e argumentar que, se fizermos isso gradualmente, isto é, se substituir neurónio a neurónio até obtermos um cérebro totalmente constituído de chips, e se tal preservar o normal funcionamento da organização desse sistema, então tal cérebro artificial seria consciente.

Não considero, no entanto, que este tipo de métodos nos levaria a criar simulações, mas levaria pelo menos a isomorfos de silício em que estaríamos conscientes como somos agora só com biologia. Chegando a este nível de desenvolvimento tecnológico, acho que não será necessário muito mais para criar seres simulados que são igualmente conscientes como nós.

Porquê? Bom, porque o que defendo é que uma simulação do meu cérebro será basicamente análoga a

um isomorfo de silício do meu cérebro: ele terá muitas partes interactivas que estão a processar exactamente as mesmas informações. Assim, se o que importa de facto é a estrutura ou o processamento da informação, podemos já estar, em princípio, perante uma simulação suficientemente boa do mundo real.

No entanto, é verdade que continuaria a existir o problema difícil da consciência, então não podemos afirmar que realmente compreendemos como é que uma simulação poderia ser consciente. No entanto, não acho que isso seja um argumento demolidor, porque nós também não compreendemos de que forma é que um cérebro pode ser consciente. O que argumento, no fundo, é que o problema da consciência é igualmente difícil no caso da simulação como no caso do cérebro humano e que, em princípio, se um pode ser consciente, não vejo por que razão o outro não possa também ser consciente.

**Pergunta:** Interessante. Diria, porém, que mesmo o processamento da informação é realmente dependente da “parte” material do sistema, ou seja, que há algo de realmente único e especial no tipo de biologia orgânica que o processo evolutivo desenvolveu e que serve de base à experiência consciente nos seres humanos. Mas o Professor acha que tal não é o caso, certo? Que o material que constitui esse substrato não

é relevante, desde que mantenha o seu poder funcional?

**David Chalmers:** Acho que o substrato pode fazer uma diferença muito significativa na forma como a informação é processada, e tenho a certeza de que a estrutura dos neurónios, por exemplo, produz uma grande diferença na forma como a informação é processada no cérebro. Dito isto, não vejo por que isso não possa, em princípio, ser simulado também. Quaisquer que sejam as propriedades idiossincráticas dos neurónios, parece-me, até onde posso afirmar, que não há nada que seja incomputável ou não-simulável.

Agora, se o Sir Roger Penrose estiver certo, então pode haver alguns processos mecânicos quânticos especiais que ocorrem nos neurónios que não podem ser simulados num computador clássico. E se tal for o caso, então precisaremos de fazer mais trabalho.

Mas mesmo que Sir Roger esteja certo, ainda me pergunto se talvez possa haver algum tipo novo e especial de simulação quântica, dado que temos computadores quânticos que basicamente exploram certas propriedades físicas da mecânica quântica.

Se o Prémio Nobel da Física estiver certo, talvez as coisas vão muito além disso: a gravidade quântica envolverá novos tipos de processos que não podem ser simulados, mesmo num computador quântico comum. Se tal for o caso podemos, mesmo assim, construir

novos computadores de gravidade quântica que possam explorar o tipo especial de computação de Penrose e, em princípio, não vejo por que não poderíamos construir uma simulação de um cérebro num desses novos computadores especiais de gravidade quântica.

Esse tipo de simulação iria além da simulação num computador clássico, mas acho que ainda significaria uma abordagem interessante para a simulação.

**Pergunta:** Sabemos que o problema da consciência faz parte da filosofia há centenas e até milhares de anos, mas a contribuição dos filósofos não tem sido muito profícua. De que forma é que os filósofos podem trabalhar para encontrar uma solução para o problema difícil da consciência? Qual pode ser o papel real da filosofia para resolver o problema difícil da consciência?

**David Chalmers:** O que a filosofia pode fazer para resolver o problema difícil da consciência? Essa é uma boa pergunta: acho que os filósofos exploraram diferentes abordagens para abordar o problema difícil. Nos últimos 20 anos, vimos muitos trabalhos interessantes sobre o pampsiquismo com propostas a tentar abordar o que torna o pampsiquismo realmente difícil de ser defendido, o que é conhecido como o “problema da combinação”, por filósofos como Philip Goff, Hedda Mørch e Galen Strawson. Outra teoria

interessante desenvolvida é a teoria do ilusionismo sobre a consciência, desenvolvida por cientistas cognitivos, mas que foi realmente impulsionada por muitos filósofos, incluindo Dan Dennet, Keith Frankish e outros. Portanto, os filósofos têm realmente tentado procurar soluções possíveis para, pelo menos, aspectos particulares do problema difícil da consciência. Sendo sincero, não me importo muito se é um filósofo ou um cientista que conseguirá resolver o problema, mas suspeito fortemente que tal terá de envolver algum tipo de interacção entre filosofia e ciência.

**Pergunta:** Seguindo Sir Roger Penrose, acha que precisamos de um novo “tipo” de física para nos aproximar do problema difícil da consciência? Que tipo de teorias poderiam ser bem-sucedidas para lidar com a natureza da experiência subjectiva?

**David Chalmers:** Que tipo de teoria poderia explicar a consciência? Gosto da ideia da matemática da consciência, onde podemos mapear a estrutura da consciência e podemos, num segundo nível, mapeá-la nos processos físicos do cérebro. A minha opinião é que uma teoria da consciência pode exigir leis psicofísicas que conectem os processos físicos à consciência e, finalmente, leis psicofísicas fundamentais.

Poderá ser, por exemplo, o pampsiquismo, mas não precisa ser o pampsiquismo: um dualista também pode

acreditar em leis psicofísicas fundamentais. Giulio Tononi pode ser visto como tendo uma proposta para uma teoria da consciência, tal como Sir Roger Penrose. No fundo, acho a questão de qual destas teorias é a teoria correcta vai depender muito do desenvolvimento da ciência. Alguém pode objectar, neste ponto, que uma teoria puramente matemática da consciência ainda estará sujeita ao problema da “Mary”.

Assim, argumentaria que devemos distinguir entre o carácter qualitativo da consciência do seu carácter estrutural, o que nos poderia permitir alcançar leis psicofísicas matemáticas objectivas que podem explicar o carácter estrutural da consciência. Claro, ainda terá de haver algum trabalho em relação ao carácter qualitativo da consciência, como a experiência de vermelhidão do vermelho.

**Pergunta:** Muitos neurocientistas e filósofos têm uma posição de que a consciência não é nada mais que um erro de categoria, um conceito que já não faz sentido existir no nosso vocabulário científico e filosófico. Por que devemos preocupar-nos com o problema da consciência em primeiro lugar? Que razões podem ser oferecidas para acharmos este problema filosófico realmente relevante?

**David Chalmers:** Não consigo convencer ninguém a importar-se com nada em específico: até certo ponto,

tal depende apenas de cada pessoa. Há muitos assuntos fascinantes no mundo e nem todos têm que ficar interessados pelo difícil problema da consciência. Por que este é interessante e importante?

Se estivermos interessados em prever o comportamento de outras pessoas, então, para esse propósito, talvez não seja necessário esclarecer nada de fundamental sobre a consciência. Tal vai depender de muitas questões complicadas, como se o dualismo interacionista é verdadeiro, por exemplo. ´

No entanto, acho que nos importamos muito mais do que com o comportamento dos outros: preocupamo-nos com as outras pessoas (ponto final!) muito mais do que nos importamos com o comportamento delas.

Acho que, para muita gente, é precisamente pelo facto das outras pessoas serem conscientes que nos importamos com elas. Daí que para questões éticas e políticas como “Como devemos tratar os animais?”, considero que é absolutamente crucial saber se os animais são conscientes e que tipo de estados conscientes eles possuem: por exemplo, “eles podem sentir dor como nós?”.

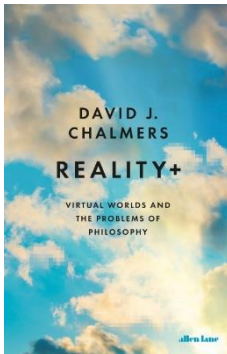
Neste ponto, alguém pode afirmar “Posso saber tudo sobre o animal apenas observando o seu comportamento”. Ótimo, mas tal não me diz o que é mais crucial: saber como devo tratar moralmente os

animais, que é assumir algo sobre a sua experiência subjectiva consciente.

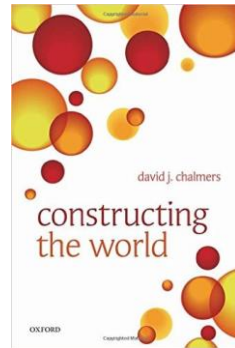
Se considerarmos que a experiência subjectiva está realmente na base do valor e do significado das nossas vidas, então temos pelo menos uma razão para nos preocuparmos com a natureza fundamental da consciência.

Podemos também encontrar razões somente intelectuais: a consciência é incrivelmente interessante, é uma anomalia na nossa imagem do universo, não temos uma abordagem teórica para a compreender, e sem uma teoria da consciência, nunca poderemos ter uma teoria completa do universo. Acho que estas são razões mais que suficientes para nos importarmos com a consciência.

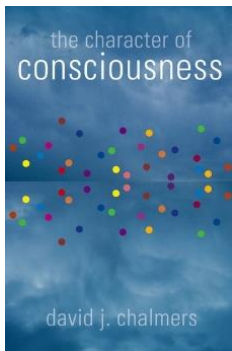
## Livros de David Chalmers



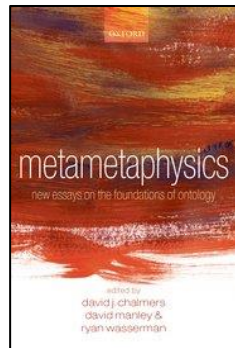
2022



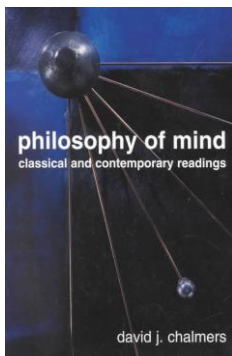
2012



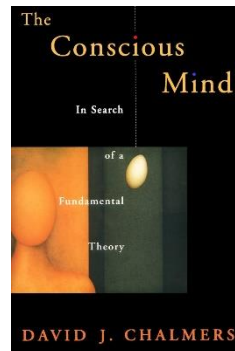
2010



2009



2002



1996



## VII. Diálogo com Susan Blackmore



Susan Blackmore é escritora, conferencista, radialista e Professora-visitante na Universidade de Plymouth, Reino Unido.

Possui formação em Psicologia e Fisiologia pela Universidade de Oxford (1973), Mestrado e Doutorado pela Universidade de Surrey (1980).

Os seus interesses de investigação incluem memes, teoria evolutiva, consciência e meditação.

A Professora Blackmore escreve para várias revistas e jornais, bloga para a *Psychology Today* e é colaboradora frequente e apresentadora de rádio e televisão.

É autora de mais de quinze livros, sessenta artigos académicos e cerca de oitenta contribuições em livros.

Alguns dos seus livros são *Dying to Live: Near Death Experiences* (1993), *The Meme Machine* (1999), *Conversations on Consciousness* (2005), *Zen and the Art of Consciousness* (2011), *Seeing Myself: What Out-of-body Experiences Tell Us About Life* e *Consciousness: An Introduction* (2018).

Mais informações: <https://www.susanblackmore.uk/>

**Pergunta:** A Professora Blackmore foi das primeiras cientistas a dar uma relevância filosófica maior aos estados alterados de consciência. Começamos pela sua experiência “maluca” que lhe aconteceu quando era mais jovem: a Professora teve uma espécie de experiência “fora do corpo” e isso influenciou muito o seu interesse pela consciência em geral. Pode contar-nos um pouco dessa história pessoal e do impacto que teve nas suas investigações?

**Susan Blackmore:** Todo esse episódio durou cerca de duas horas e meia. Estava sentada no quarto de um amigo, no meu primeiro período em Oxford – tinha 19 anos, estamos a falar de há 50 anos atrás. Lembro-me que estava muito cansada, sem dormir há alguns dias, e que tinha fumado um pouco de cânabís, não o suficiente para explicar todo o fenómeno, embora tenha possivelmente contribuído para o mesmo.

Em determinado momento, comecei a descer por um túnel. Era como um túnel de árvores, com folhas ao redor, em direcção a uma luz. Tudo isto aconteceu antes mesmo de o termo “experiência quase de morte” ter sido inventado.

Nesse período, não tinha qualquer conhecimento sobre este tema. O túnel que estava a experienciar transformou-se numa experiência fora do corpo: eu parecia estar fora do meu corpo, podia olhar para baixo e ver o meu próprio corpo ali estendido no sofá. Durante este período, ainda estava comunicativa; um

amigo estava a tentar conversar comigo e perguntava-me “o que estás a ver agora?” e nesse preciso momento comecei a viajar – o que parecia ser viajar pelo mundo – e ver todo o tipo de coisas.

Para resumir sucintamente a história, tentei voltar ao meu corpo, mas isso foi muito difícil: senti-me a ficar cada vez mais pequena, e encolhi, e encolhi, e encolhi e fiquei muito pequena mesmo. E então fiquei muito assustada, e de repente comecei a ficar cada vez maior e a experiência expandiu para uma experiência mística clássica, sobre a qual, novamente, eu não sabia nada.

Com todas estas fases de experiências diferentes, senti que o meu Eu era agora um “tudo”. Não havia mais nenhum Eu particular, uma Susan, mas havia algo que parecia ser tudo. Muitas (outras!) experiências aconteceram ao longo deste episódio, mas fiquei eventualmente exausta e o meu amigo perguntou-me: “Há mais alguma coisa?” e pensei “Não, porque sou Tudo, como poderia haver outra coisa?”. Os pensamentos ainda estavam a acontecer, mas não através de palavras. E então tive uma espécie de percepção de que há sempre algo mais.

Precisei de dois dias para voltar a sentir que estava dentro do meu corpo. Esta experiência fez-me acreditar em todo o tipo de coisas psíquicas e místicas e em coisas do outro mundo. Tal foi, claro, ilógico, mas compreensível na época para uma jovem de 19 anos que não sabia nada sobre a natureza destes

fenômenos. Este episódio aconteceu no início da década de 1970: não havia neurociência para nos dar algumas respostas. Deste episódio, fiquei a acreditar em telepatia, clarividência, psicocinese, fantasmas, poltergeists, tudo! Mundos do além, vida após a morte, almas, espíritos, etc.

Por causa desta experiência decidi não fazer um doutoramento reputado que me foi oferecido numa universidade de prestígio, e decidi avançar com o meu próprio doutoramento sobre o “paranormal”. Levei cerca de cinco anos de investigação sobre todos estes temas paranormais para descobrir que é quase certo que eles não existem de todo e são uma criação da nossa mente.

Esta é a breve história de como me transformei de uma crente em todo o tipo de coisas estranhas para decidir que as verdadeiras questões não são: existe um espírito, ou uma alma, ou consciência além do cérebro? Mas a verdadeira questão está aqui e agora: que experiência fenomenal é esta que tenho? Como é que um cérebro cria esta experiência? Como é que um cérebro, um mundo e um corpo contribuem para isso? Qual é a natureza da consciência? Estas são as verdadeiras questões que são muito mais emocionantes do que se existe vida após a morte, ou espíritos, ou qualquer outra coisa.

**Pergunta:** Seguindo essa sua experiência, sabemos que muitos tipos de drogas podem induzir os chamados estados alterados de consciência. Acha que podemos usar, por exemplo, psicadélicos como DMT ou LSD para estudar a natureza da mente consciente? Acha que essas metodologias são úteis do ponto de vista de compreendermos o fenómeno da subjectividade, dadas as várias dificuldades metodológicas que existem actualmente na investigação científica da consciência?

**Susan Blackmore:** Dependerá um pouco do que se pretende estudar em relação à natureza da mente. Se ficarmos convencidos que estes métodos nos vão dar respostas rápidas, então não considero que sejam de facto úteis. Mas de uma perspectiva mais geral, sim, podem ter alguma utilidade.

Pode depender também da relação pessoal com essas substâncias, e se tal pode permitir o ganho de alguma compreensão através das suas próprias vivências ou se a investigação que está a realizar nos oferece alguma explicação. Mas diria que a mente muda de forma tão radical durante um episódio de influência de substâncias que não conseguimos apreender algo de substancial no imediato.

Pode dar-nos alguma luz sobre um estado normal de consciência antes e depois, por exemplo. Mas como William James afirmou em 1890, nós sabemos que, além do véu desta experiência particular que estamos

a ter, existem inúmeras outras maneiras de estar consciente que estão muito próximas desta, mas precisamos de algo como uma droga para nos levar até lá, ou uma experiência mística espontânea.

Acho que uma das investigações mais interessantes que li sobre este tema é o artigo muito recente de Timmermann e colegas<sup>57</sup> sobre a DMT que é o ingrediente psicoativo da ayahuasca que perturba as principais redes funcionais do cérebro.

No cérebro encontramos várias redes principais, mas uma delas em particular, a rede de modo padrão (*default mode network*, DMN) é, efectivamente, aquela que sublinha o sentido do Eu. A DMN é a rede de longo alcance que reúne o esquema corporal com as memórias sobre si mesmo, as suas opiniões, etc.

A DMN conecta-se à junção temporoparietal direita e é nesta zona do cérebro que o esquema corporal se conecta aos sistemas de controlo no córtex frontal e à memória nos lobos temporais, e assim por diante. Esta rede é interrompida quando se toma DMT (dimetiltryptamina). Ao mesmo tempo, noutras partes do cérebro, por exemplo, no córtex visual, há um aumento da actividade local que pode explicar as

---

<sup>57</sup> Publicação original: Timmermann, et al. (2019) "Neural correlates of the DMT experience assessed with multivariate EEG", *Scientific Report*, 9: 16324.

alucinações visuais que ocorrem com o consumo de DMT.

Quando o sistema que produz o sentido do Eu é perturbado, deixamos de sentir que existe um Eu. Na minha perspectiva, quando tal acontece, podemos observar a natureza ilusória do poderoso Eu consciente, o Eu que consideramos ter consciência ou livre arbítrio.

Para mim, todos estes são ilusões na minha mente, e o facto de a DMT perturbar o Eu, significa que um indivíduo pode – durante algumas horas, se for ayahuasca, ou 15 minutos se fumar DMT – ter esta experiência sem o sentido normalizado do Eu.

Todo este tema parece-me relevante para conseguirmos ver o início de um aspecto das ilusões da consciência. Acredito que a maioria das teorias da consciência permanecem presas a estas ilusões e não têm qualquer progresso por confundi-las com algo de real.

Talvez os psicadélicos em geral possam ajudar neste sentido e acredito que toda a investigação que estamos a realizar actualmente está a começar a revelar muito sobre a forma como a mente funciona. E haverá mais para descobrir: agora que tudo começou e a lei não pode mais travar esta linha de investigação (houve no passado uma lei negativa contra o uso de drogas na

investigação científica), acho que iremos aprender muito.

**Pergunta:** E em relação aos sonhos lúcidos? Sabemos que os sonhos lúcidos estão muito mais próximos de estar acordado do que de estar em estado de sono profundo. A Professor Blackmore acha que podemos mapear alguns estados de consciência a partir dos sonhos lúcidos, por exemplo, através de investigação empírica?

**Susan Blackmore:** Quando o Steven afirma que o sonho lúcido é estar num estado semelhante ao estar acordado, não tenho certeza que tal seja mesmo o caso. Estes estados ocorrem, principalmente, na fase REM do sono, e isso significa que precisamos de estar a dormir adequadamente. É nesta fase REM que podemos ter sonhos lúcidos.

Estes estados acontecem em períodos mais activos do REM, onde há mais actividade a acontecer do que nos períodos de menor actividade do REM. Mas o que é realmente interessante é que – e, mais uma vez, diz respeito à rede de modo padrão (DMN) – as conexões entre a junção temporo-parietal e os lobos frontais tornam-se mais fortes nos sonhos lúcidos, o que sugere que esta é uma base psicológica para a sensação de “ah, eu estou aqui, agora, e posso controlar o sonho!”.

Mais uma vez, adoro descobrir estas coisas, porque elas simplesmente destroem o tipo de teorias alternativas sobre o que acontece num sonho lúcido, como o despertar das nossas almas e todo o tipo de coisas absurdas que as pessoas imaginam. Infelizmente, a psicologia convencional não está ainda realmente focada na relevância científica dos sonhos lúcidos e não encontramos actualmente investigadores suficientes a fazer investigação sobre este tema. Considero que estes estados são, sem sombra de dúvida, uma fonte de aprendizagem muito relevante sobre os nossos cérebros e as nossas mentes.

**Pergunta:** A Professora Blackmore é praticante de Meditação Zen há cerca de 30 anos, certo? De que forma é que acha que a meditação pode fornecer uma estrutura útil para estudar ou compreender a mente consciente? Acha que há alguma relação entre estes dois conceitos? E se sim, qual será essa relação?

**Susan Blackmore:** Sim, mas pergunto-me o que quer dizer com “compreensão”, no sentido de que a meditação não serve para compreender nada, mas para esclarecer a mente, ou pelo menos essa é, para mim, a melhor razão para alguém querer meditar. De certa forma, a meditação me ajudou a entender coisas sobre a mente, mas também ajudou-me abandonar as ilusões, a abandonar a crença de que existe um Eu aqui,

e um mundo lá fora, a abandonar o desejo de me tornar algo mais importante.

O que sabemos é que a mente de quem medita muda de forma bastante radical no sentido de sermos mais abertos a tudo no mundo, mais receptivos, menos apegados, e é para isso que a meditação serve para mim, embora essa mudança seja muito lenta. Se tal pode ajudar-nos a compreender a mente? Bem, sim, claro, porque os praticantes de meditação de longo prazo possuem uma enorme eficácia em entrar em diferentes estados de consciência, e podemos criar investigações interessantes a partir disso.

Nos últimos anos tenho praticado a Meditação Jhanas, na qual se supõe haver oito estados discretos de consciência que podem ser alcançados através de concentração profunda. Ouvi falar sobre isso há décadas atrás e pensei em como seria incrível que a existência de estados alterados específicos poderiam ser alcançados simplesmente seguindo determinadas instruções. Achei que nunca seria capaz de alcançar esses estados pela meditação. Mas graças a um grande professor de Jhanas que conheci em Inglaterra, tenho tido o privilégio de participar nos seus retiros.

Tenho descoberto que parece ser de facto possível, simplesmente seguindo algumas instruções (bastante intensas), alcançar alguns desses estados: começando com uma concentração profunda e depois fazendo várias outras coisas, podemos entrar em estados

alterados claramente demarcáveis que podemos controlar nós mesmos. Até agora, só consegui entrar nos três primeiros estados, e um pouco do quarto talvez, e continuo a praticar com o objectivo de conseguir alcançar outros estados.

Pela prática que tenho feito, estou convencida de que esses povos antigos que praticavam estas meditações realmente descobriram esses diferentes estados de consciência e descobriram também como chegar até eles. Isso informa-nos sobre o que a mente é capaz: e quando consegue fazer isso ou mesmo apenas ler e entender o que as pessoas estão a relatar sobre essas experiências, então sabemos que a mente é capaz de estar num tipo de relação completamente diferente dela mesma, do mundo e da ideia que ela tem de si mesma. Portanto, sim, acho que a resposta é que a meditação nos pode aclarar algumas ideias sobre a natureza da mente.

**Pergunta:** Tenho impressão de que esses diferentes estados alterados da consciência podem ser reconhecidos através de imagética neuronal, certo? Será que, com prática suficiente, poderemos ver oito tipos de actividade cerebral durante as oito fases diferentes dessa meditação específica?

**Susan Blackmore:** Do que tenho conhecimento, apenas foram realizadas duas experiências

neurocientíficas sobre esse tema, ambas com o praticante de meditação Lee Brasington.<sup>58</sup> Estas experiências cerebrais com fMRI ou alguma técnica imagética similar em neurociência são muito dispendiosas.

O problema é que o sujeito tem de ser capaz de indicar aos cientistas, neste estado de concentração muito profundo, quando está a mudar para o estado meditativo seguinte e de associar essa mudança à sua actividade cerebral. Certamente que os neurocientistas devem ter identificado algumas mudanças no momento em que o sujeito indica que está a passar do estado dois para o terceiro, por exemplo. Mas não me parece que tenhamos dados suficientes e não creio que Leigh tenha conseguido passar pelos oito estados, porque os dois últimos estados são estados associados ao que se poderia chamar de “inconsciente”.

São estados que estão tão “afastados” que seria muito difícil comunicá-los naquele ambiente experimental – mas tal seria incrível, com certeza. É muito, muito difícil meditar adequadamente dentro de um scanner fMRI, com todo aquele barulho e tudo mais. Então, acho que essa previsão teórica é talvez demasiado esperançosa.

---

<sup>58</sup> Publicação original: Hagerty M. et al. (2013) “Case study of ecstatic meditation: fMRI and EEG evidence of self-stimulating a reward system”, *Neural Plasticity*, 653572.

**Pergunta:** Sabemos que a consciência é realmente difícil de investigar de forma rigorosa. No entanto, encontramos indivíduos como Deepak Chopra – com quem a Professora Blackmore já teve vários debates – que apoiam algumas ideias espirituais bem duvidosas sobre a mente. O que acha dessas ideias em geral?

**Susan Blackmore:** Se alguém estiver interessado em Deepak Chopra, dê uma vista de olhos no meu Facebook e assista a um vídeo que coloquei sobre o meu debate com ele, e verá um autoproclamado guru espiritual a comportar-se de forma bastante duvidosa.<sup>59</sup>

Não posso responder à pergunta sobre *todos* esses gurus, mas posso responder sobre Deepak, que defende que a consciência é primária e a matéria não existe. Tal é simplesmente absurdo: não tem qualquer significado real e a maior parte das coisas que ele afirma não significam nada.

Os conselhos práticos que ele dá sobre habilidades de como treinar a mente, e algumas das coisas que ele diz sobre meditação e as suas consequências são precisas: nesse aspecto, ele realmente tem um conhecimento interessante. Mas quando se trata da relevância das suas teorizações sobre a mente, estas são muito insatisfatórias de um ponto de vista filosófico informado pela melhor ciência.

---

<sup>59</sup> Vídeo original: [www.youtube.com/watch?v=ZFGkqhNhgM](http://www.youtube.com/watch?v=ZFGkqhNhgM).

Consideremos o problema filosófico mente-corpo: se somos materialistas, defendo que tudo o que existe no mundo é constituído por matéria, e aí não conseguimos explicar a experiência subjectiva, seguindo o que chamamos de “problema difícil da consciência” que não está resolvido. Talvez seja, na verdade, a pergunta errada, o problema errado.

Mas se afirmarmos que somos idealistas, temos de defender que tudo o que existe é consciência. Mas aí deixamos de poder explicar a matéria. Então, claramente, nenhuma dessas duas posições funciona. O dualismo não funciona porque estamos perante dois tipos de coisas completamente diferentes. Assim, há algo de profundamente errado, e por isso defendo um suposto ilusionismo em relação à consciência, mas Deepak é apenas um idealista, embora não admita que o seja.

Ele apenas diz que tudo é consciência e que a matéria não existe. Ótimo! Esta tese não faz quaisquer previsões, não há base teórica para defendê-la, não há conclusões que se possam tirar desta afirmação e no entanto as pessoas adoram ouvi-lo falar! As pessoas adoram porque pensam: “de alguma forma, a minha consciência é tão maravilhosa e importante, e eu posso fazer tudo isto com a minha consciência! Ele está realmente consciente, então ele deve ser muito espiritual!”.

Claro, estou a fazer uma caricatura aqui, mas acho realmente que as ideias dele, em termos do que ele pensa, são, do ponto de vista científico, simplesmente vazias. E o interessante é que ele recorre a alguma investigação científica, mas distorce e faz interpretações absurdas para as adequar a essa tese idealista que defende.

**Pergunta:** A Professora Blackmore afirmou que é uma ilusionista, mas disse que rejeitaria tanto o materialismo quanto o idealismo. Podemos dizer que o ilusionismo talvez seja uma espécie de combinação com o materialismo, no sentido de que a dor é um material que cria uma ilusão?

**Susan Blackmore:** É uma pergunta interessante essa. Tudo o que posso dizer é que a maioria dos ilusionistas são materialistas. Isso é uma certeza. Não me considero materialista no sentido tradicional, mas por outro lado, não tenho nada para substituí-lo. A minha ideia de ilusionismo é bastante ampla: todas estas formas existentes de pensar sobre a consciência estão erradas, e ainda não descobrimos uma saída para este dilema, e é isso que chamamos do “problema difícil”.

Por outro lado, quem tenta perceber como a consciência surge do cérebro parece cair numa espécie de estrutura dualista: embora todos estes materialistas, com excepção de David Chalmers, que

inventou o termo “o problema difícil”, continuam a ser materialistas sem realmente resolverem o problema. Eles estão a afirmar, como eu, que “se pensarmos que a consciência é algo que temos, se pensarmos que tem poder e faz alguma coisa, se pensarmos que a consciência evoluiu para um propósito, então estamos iludidos”.

Penso que Daniel Dennett e Keith Frankish, por exemplo, diriam “sim, existe um cérebro material a partir do qual as ilusões são construídas”, e estão a tentar resolver o problema de “como é que essas ilusões surgem na mente?”. Assim, eles substituem o problema difícil pelo problema da ilusão, ou o metaproblema, como Chalmers o chama. Esta alteração do problema central é um progresso relevante do meu ponto de vista.

Não sou filósofa, mas, se me perguntarem, diria que sou uma monista neutra. Sou monista porque acho que o dualismo não funciona. Mas não acredito que os nossos conceitos actuais de “material”, de “matéria” ou de “mente” sejam úteis.

Não sei de facto em que consiste o universo: talvez alguma física fundamental possa ajudar, embora entremos em grandes dilemas por não sermos capazes de combinar a mecânica quântica com as teorias einsteinianas. Podemos também considerar algumas das teorias mais recentes sobre a informação que afirmam que tudo o que existe é informação. E

podemos pensar também em termodinâmica e entropia.

Enfim, acho que é nestas possíveis abordagens que alguém poderá chegar a uma solução: da minha parte, estou a tentar compreender todos estes novos conceitos, o que é um pouco difícil dada a multiplicidade de abordagens e conhecimentos envolvidos. Mas creio que é por isso que o tema da consciência é o desafio mais estimulante da ciência actual.

**Pergunta:** A Professora abordou a sua pessoa ilusionista em relação aos estados conscientes. No entanto, diria que poderá haver também uma relação entre a ilusão da consciência e a ilusão do Eu? Qual é a relação entre elas, se esta existir?

**Susan Blackmore:** O melhor que posso dizer é que são ilusões extremamente próximas. O início da ilusão da consciência é a separação do “Eu” do “Outro”. O problema das outras mentes, se preferir; a separação de mim das outras pessoas. A sensação de que esta é a minha consciência e a minha é diferente da das outras pessoas, porque não posso saber como é para os outros ver o azul e se é igual ao meu azul e todos esses tipos de coisas.

É tudo uma ilusão? Não: diria que as coisas que me interessam em termos de possíveis ilusões são principalmente as armadilhas.

Se olharmos para o trabalho do Daniel Dennett, principalmente no seu livro *Consciousness Explained*, ele fala sobre todas as armadilhas em que as pessoas caem, e também faz trabalhos posteriores sobre as “bombas de intuição”,<sup>60</sup> e assim por diante.

A teoria dos múltiplos rascunhos (*multiple draft theory*) avançada por Dennett leva-nos a concluir algo que, para a grande parte dos neurocientistas, é totalmente bizarro: que não há distinção entre processos neuronais que são conscientes daqueles que não são conscientes. Na verdade, não há diferença. Não há nenhum facto sobre isso, é algo sem sentido afirmar que esses processos cerebrais que ocorrem aqui ou ali são os da consciência.

Toda a procura pelos correlatos neuronais da consciência na minha mente está completamente errada. Decorre destas ilusões, porque diz que a própria consciência *emerge* de algum processo particular, de alguma área particular do cérebro. Mas, de acordo com Dennett, simplesmente não é esse o caso.

---

<sup>60</sup> Publicação original: Dennett, D. (2013) *Intuition Pumps and Other Tools For Thinking*, New York: Penguin.

Se eu perguntar ao leitor: “está consciente agora neste preciso momento?!” – ao ler esta pergunta, pode ter tido uma experiência extraordinária e pensado: “claro que estou! Mas, espere, um momento atrás eu estava apenas a ler esta frase, mas agora algo mudou. Tornei-me mais consciente!”.

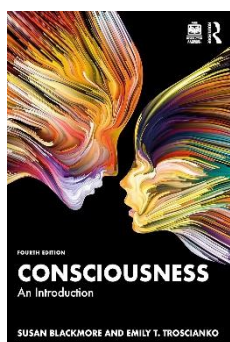
Tal leva-me a pensar que – voltando a relacionar com a teoria dos rascunhos múltiplos – se o leitor não sabe se estava consciente antes de pensar sobre isso, então ninguém sabe. Não podemos colocar nenhum medidor de consciência no cérebro e declarar onde está a consciência, quando uma pessoa está consciente ou não num determinado momento.

Então, se não sabe, Dennett diria que não há nenhum facto relevante sobre isso. Tudo o que fazemos é, retrospectivamente, depois do facto, dizer que estávamos conscientes disto e não daquilo: é sempre uma atribuição posterior ao facto em si.

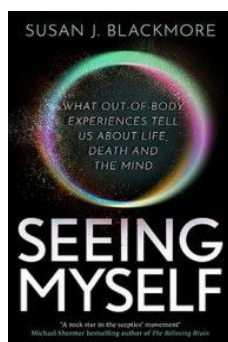
Voltando à sua pergunta, será que o Eu e a consciência são baseados na mesma ilusão? De certa forma, são, porque sou *Eu* quem está consciente, e isso envolve a separação entre mim e essa coisa chamada “a minha consciência” e as coisas das quais estou consciente de. Então, suponho que minha resposta seja “sim” e “não”: sim, estão intimamente relacionados, mas não, podemos separá-los em vários níveis diferentes de ilusão.

Para concluir, penso que o campo dos estudos da consciência está atolado de ilusões e não faremos nenhum progresso real até compreendermos como essas ilusões surgem e aprendermos a “ver” através delas.

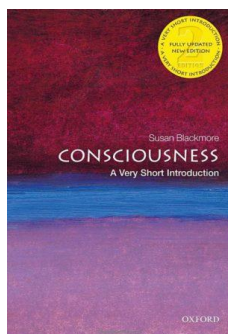
## Livros de Susan Blackmore



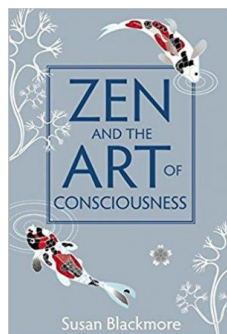
2024



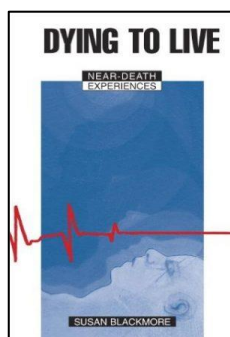
2020



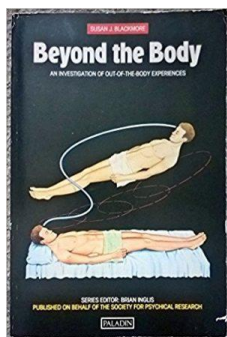
2017



2011



1993



1992



## VIII. Diálogo com Nicholas Humphrey



Nicholas Humphrey é Professor Emérito de Psicologia na London School of Economics, Professor-visitante de Filosofia no New College of the Humanities e membro sénior do Darwin College, em Cambridge.

É psicólogo teórico, conhecido internacionalmente pelo seu trabalho sobre a evolução da inteligência e da consciência humanas. Os seus interesses são variados: estudou gorilas da montanha com Dian Fossey em Ruanda; foi o primeiro a demonstrar a existência de "visão cega" após dano cerebral em macacos; propôs a célebre teoria da "função social do intelecto".

É um autor prolífico tendo publicado vários livros: *Consciousness Regained* (1983), *The Inner Eye* (2003), *Seeing Red: A Study in Consciousness* (2009), *A History of the Mind* (2012), *The Mind Made Flesh* (2012), e *Soul Dust: the Magic of Consciousness* (2012).

Publicou recentemente o seu novo livro *Sentience: The Invention of Consciousness* (2023),

Mais informações: <https://www.humphrey.org.uk/>

**Pergunta:** No verão de 1966 o Professor Humphrey conheceu a Helen, uma macaca cega cujo córtex visual tinha sido removido para propósitos de investigação científica. O que é que o Professor aprendeu sobre a consciência com a vivência com a Helen?<sup>61</sup>

**Nicholas Humphrey:** Foi uma experiência muito estranha quando encontrei esta macaca pela primeira vez. Era estudante de doutoramento na Universidade de Cambridge e no laboratório havia uma macaca que não tinha sido operada por mim mesmo, mas por outro colega, com o objectivo de descobrir qual era a função do córtex visual. Então, uma colega minha removeu todo o córtex visual do cérebro desta macaca e, sem surpresas, ela parecia ter ficado completamente cega.

Conheci Helen um ano depois da operação e ela ficou ali sentada, a olhar vagamente para o nada: ela não estava nem um pouco interessada em usar os seus olhos. Mas fiquei intrigado com isso, porque Helen teve o córtex removido na parte de trás do cérebro, mas deixou-se intacto o sistema visual mais antigo, que é o sistema visual usado por peixes, sapos e outros vertebrados não-mamíferos, o sistema visual do colículo superior do mesencéfalo, que lhes permite ver perfeitamente bem: um sapo pode usar os olhos e

---

<sup>61</sup> Gravação em vídeo de Helen em acção: [www.youtube.com/watch?v=rDIsxwQHwt8](http://www.youtube.com/watch?v=rDIsxwQHwt8).

apanhar as moscas, e assim por diante, sem nenhum problema.

Então, como esse sistema ainda estava intacto neste macaco, fiquei a pensar se, de facto, ele poderia ser treinado para voltar a ser usado, mesmo que um macaco com um cérebro normal não fizesse uso desse sistema visual antigo. Será que Helen poderia recuperar a sua visão usando este sistema mesencéfalo?

Quando me deparava com estas questões, o meu supervisor viajou durante uma semana para uma conferência em Zurique na Suíça e aproveitei a oportunidade para sentar-me com esta macaca e brincar com ela, para tentar fazer com que ela interagisse comigo da maneira que pudesse. Para minha surpresa, descobri que, em poucas horas, ela estava a usar os seus olhos.

Ela obviamente estava a prestar atenção ao que eu estava a fazer: por exemplo, passei um pedaço de fruta à frente dela para que estendesse a mão e pegasse nessa fruta e ela fez isso mesmo. No final de uma semana, já conseguia estender a mão para tocar uma pequena luz que era colocada na frente dos seus olhos, ou tocava num objecto que era colocado na ponta de uma vara e que era passado à frente dos seus olhos.

Com estes resultados incríveis (dado que não era suposto a Helen observar nada do que eu lhe estava a

colocar à frente dos olhos), fiquei muito entusiasmado e enviei um telegrama ao meu supervisor em Zurique e ele não gostou do que lhe tinha escrito. Enviei um telegrama a dizer “Ensinei a Helen a ver. Não vai acreditar”. Na altura, eu tinha 20 anos e o meu supervisor era um grande professor sénior e não ficou muito satisfeito com isso. Ele voltou para Cambridge e um dia ou dois dias depois, convenci-o a ir ver a Helen, e ele teve que concordar que algo surpreendente havia acontecido naquela semana: ela era claramente capaz de usar os olhos novamente.

Então, o resultado deste episódio todo foi que o meu supervisor permitiu-me continuar a trabalhar com ela e acabei por trabalhar cerca de 7 anos com a Helen. No final desses 7 anos, ela estava, aparentemente, com o seu sistema visual normalizado. Ela corria pelo quarto, pegava em objectos do chão sem esbarrar em obstáculos, podendo até estender a mão e capturar moscas que passassem por ela.

Muitas pessoas concluíram que que ela tinha uma visão normal, mas fiquei intrigado com isso. Quando a macaca Helen estava comigo, relaxada e confiante, ela conseguia ver, mas se ela estivesse ameaçada ou ansiosa, a sua visão desaparecia, ela começava a tropeçar, como se estivesse numa escuridão cega novamente. Então, parecia que só conseguia ver aquilo em que não precisava de pensar muito. Se tivesse de

pensar, a sua confiança desaparecia, assim como a sua capacidade de ver.

Achei que era uma forma de visão extraordinária: era uma visão na qual o próprio macaco não acredita de facto que está a ver. Então, escrevi um artigo onde chamei essa capacidade de “ver e o nada”, ecoando o livro de Jean-Paul Sartre [Ser e o Nada], já que achei que estava claramente a faltar alguma coisa. E logo descobrimos o que faltava: inspirado por este episódio, o meu supervisor testou pacientes humanos de uma forma diferente e estabeleceu o famoso fenómeno da visão cega em humanos.

Ele descobriu que um ser humano com grandes danos no córtex visual que acreditasse ser cego poderia, de facto, usar os olhos para ver naquela área do campo cego: o paciente conseguiria adivinhar o que estava a acontecer à frente do seu campo de visão, adivinhando a forma e a posição de um objecto particular, mas ao mesmo tempo afirmava que “não tem nada aí”, “não entendo”, “isso não faz sentido”, “a minha visão não tem nada a ver comigo”.

Então, a partir disso, passei a colocar a grande questão: se um individuo pode ver, e um sapo pode ver, e um peixe pode ver, sem usar o córtex visual e sem ter sensações visuais, então para que serve o sistema secundário e qual é o seu sentido? Qual é a utilidade das sensações visuais? E essa é realmente a questão na qual tenho trabalhado desde o resto da minha vida.

Nos últimos 50 anos tenho tentado descobrir o papel funcional das sensações físicas.

E, claro, isso levou-me a colocar outra grande questão: por que é que as sensações visuais e todas as outras sensações têm as qualidades fenomenais tão estranhas que têm? Como é ver o vermelho, ou sentir o gosto do açúcar, ou ouvir um grito ou um toque de sinos? Não estamos apenas a obter informações sobre o objecto: estamos a obter informações numa dimensão diferente, na dimensão fenomenal, de como ele se relaciona connosco, de como é para nós receber essa informação.

**Pergunta:** O Professor Humphrey afirma que a consciência é um produto da evolução, porque confere benefícios específicos de sobrevivência à humanidade...

**Nicholas Humphrey:** Não digo que seja um produto da evolução *porque* confere vantagens de sobrevivência. O que pergunto é: o que mais pode ser, senão um produto da evolução? Pelo menos nesse caso, deve confirmar as vantagens de sobrevivência. Então, temos que descobrir quais são essas vantagens, exactamente...

**Pergunta:** Mas também sustenta que a consciência é uma forma de ilusão que o nosso cérebro cria para nós mesmos. Nesse ponto de vista, a consciência é percebida como uma construção mental, uma representação interna da realidade que experimentamos.

**Nicholas Humphrey:** Costumava afirmar isso, sim. Mas tal sugere que a consciência é uma ilusão. E li vários artigos e até outros livros que descrevem a consciência como uma ilusão, e que agora se tornou uma visão popular entre certos filósofos, como Keith Frankish e Daniel Dennett, que são considerados ilusionistas da consciência, mas rapidamente afastei-me desse rótulo.

Não creio que seja a maneira correta de descrever a consciência. Penso que chamar algo de ilusão sugere que é um erro, que estamos errados ao atribuir as qualidades que fazemos à experiência. Não acho que seja assim: acho que quando vemos vermelho ou cheiramos uma rosa, esta é realmente como parece ser. É realmente assim que sentimos aquela experiência subjectiva. É assim que ela é, e descrever isso como uma ilusão é subestimar o papel que desempenha na nossa psicologia e na nossa fenomenologia.

Para mim, a consciência fenomenal é uma descrição verídica de como é para um ser humano ter essas experiências. E a questão central é “por que

representamos isso dessa forma?”. Sensações são representações, poderiam ter sido apenas representações dos factos, simples factos físicos, mas não são. São representações de como nos sentimos ao ter esses estímulos através dos nossos corpos.

**Pergunta:** É interessante seguir esse pensamento porque, no presente, considera-se um “surrealista” em relação aos *qualia* ou à consciência. Pode explicar-nos o que pretende defender com essa posição?

**Nicholas Humphrey:** Estava a tentar encontrar outro termo para descrever a minha posição. Houve um artigo escrito e publicado sobre ilusionismo pelo filósofo Keith Frankish. Na minha resposta a esse artigo, defendi que “Não, acho que não é uma ilusão; não é irreal; na verdade, é super real”. A vermelhidão fenomenal é mais vermelha que o vermelho; a dor fenomenal é mais dolorosa que a dor. E tirei essa frase de Pablo Picasso, que foi um dos primeiros surrealistas, embora não aceitasse o termo. Picasso tinha produzido uma escultura famosa de uma cabra,<sup>62</sup> tendo afirmado “a minha cabra é mais cabra do que qualquer cabra real”.

---

<sup>62</sup> Escultura de título “She-Goat” (1950) em exibição no Museu de Arte Moderna de Nova York.

Noutras palavras, Picasso tentava expressar na sua arte a essência do objecto que criava. E dizer que foi além, e que foi mais profundo que a própria realidade da realidade física. E acho que talvez essa seja a maneira mais interessante de falar sobre as sensações fenomenais e a consciência. Num certo sentido, parecem ir mais fundo do que os factos superficiais da realidade.

**Pergunta:** O Professor Humphrey também defende que há uma relação estreita entre os *qualia* e o que podemos chamar de “Eu fenomenal”. Por que acha que existe uma relação próxima entre estes dois conceitos?

**Nicholas Humphrey:** Penso que os *qualia*, a experiência fenomenal, são a base do Eu. E, de facto, esse é o caminho a seguir na nossa economia psicológica, e é por isso que ela evoluiu. O que as sensações fazem é dar uma realidade substancial ao nosso sentido de quem somos e da nossa existência no mundo.

Notoriamente, David Hume, o filósofo, afirmou que, quando tenta examinar a sua própria mente e descobrir o que significa ser ele mesmo, não encontra nada além de sensações. E ele ficou desapontado com isso, uma vez que as sensações são evanescentes, não parecem ter nenhuma realidade contínua e não existe

uma base substancial para o Eu, baseada apenas nas sensações.

Acho que tal visão está completamente equivocada: as sensações dão-nos de facto a base mais sólida que poderíamos ter para a nossa existência no mundo. São evidências sempre presentes de como vivemos as nossas vidas e de como somos importantes, de como é ser nós mesmos e, mais importante, da nossa própria individualidade. As experiências que temos são diferentes de tudo no mundo e das outras pessoas.

É óbvio que podemos assumir que outros humanos têm experiências como as nossas. Mas a evidência não está aí: tudo o que sabemos é que a minha sensação de vermelho se assemelha à sensação de vermelho que tive antes. Parece ter algo em comum com as minhas outras sensações e outras modalidades. Tudo é feito no meu estilo, mas essa é a prova da minha realidade contínua.

Descrevo a continuidade das sensações como sendo semelhante à continuidade das obras de arte de um determinado pintor. E assim como podemos identificar estilo de Cézanne, de Vermeer, ou de Picasso nas suas obras, todas as minhas sensações são minhas, e isso dá-me uma realidade no âmbito das coisas que acho muito significativa. Quando crescemos, descobrimos que é isso que constitui o nosso centro psicológico, e passamos então a assumir que outros humanos têm

um centro do Eu equivalente, no qual o jogo de sensações é semelhante às nossas.

Para cada um deles, esse jogo será individual, privado e importante. Uma vez que adotamos essa visão das outras pessoas, os parâmetros da vida social começam a mudar. Passamos a viver no que chamei de “A Sociedade dos Eus”, dos Eus fenomenais. E essa é certamente a base da cultura humana e da civilização humana. A grande questão é se isso vai além dos humanos e se outros animais também podem ser considerados como possuindo um Eu fenomenal.

**Pergunta:** Seguindo essa ideia, o Professor Humphrey defende também que a consciência tem um papel específico relacionado com fornecer uma espécie de significado à vida e à existência humana em geral. Poderia elaborar um pouco mais sobre esse aspecto específico da consciência, que é muitas vezes ignorado pelos pensadores que se dedicam a estudar este fenómeno?

**Nicholas Humphrey:** Essa é uma grande questão, dar sentido à vida, sim. Bem, por onde começamos a encontrar significado? Primeiro, dá-nos uma noção da nossa própria importância, de que não somos fenómenos puramente físicos, de que existimos num outro plano ou de que temos uma dimensão espiritual nas nossas vidas. De certa forma, existimos fora do

tempo e do espaço, num reino que não pode ser descrito pelas qualidades materiais da matéria física.

Esta é uma descoberta muito importante sobre nós mesmos e é, claro, uma base não apenas para um simples sentimento de “sim, eu sou importante”, mas pode aperfeiçoar-se, uma vez elaborada pela cultura e pela linguagem: pode dar-nos a sensação de que temos uma alma imortal.

Neste ponto, o leitor pode pensar que é uma ideia estranha para um cientista levar a sério, mas eu levo-a muito a sério. A crença na alma tem sido uma das forças motrizes da história humana e, de facto, é responsável pela maioria das coisas significativas que os humanos alguma vez alcançaram. Porque uma vez que acreditamos nas almas e na sua importância, e nas almas das outras pessoas e na sua importância, isso dá novas ambições para o que queremos deixar para trás e o que queremos alcançar nas nossas próprias vidas.

Agora, é claro que isso é sobre seres humanos. Não creio que os cães tenham o sentido de serem almas, por exemplo. Mas eles, no entanto, sentem-se como sendo algum tipo de indivíduo significativo e, para eles, temos que contar uma história um pouco diferente. Para os humanos, a consciência passou a ter resultados bastante inesperados e maravilhosos. Descrevi isso no meu mais recente livro: acho que é a jóia da coroa da evolução biológica e devemos levar essa ideia seriamente.

**Pergunta:** O Professor argumenta nesse seu novo livro *Sentience: the Invention of Consciousness* que a sentiência está restrita a mamíferos e pássaros por causa de uma característica fisiológica muito específica: considera que eles são sencientes porque têm sangue quente. Tal é, de facto, uma característica curiosa e inovadora que não foi pensada antes em relação a uma explicação plausível para a consciência na sua vertente senciente. O que o fez chegar a essa ideia?

**Nicholas Humphrey:** Esta não é uma tese científica firme, mas penso que o que temos de aceitar é que a sentiência, a consciência fenomenal, é um desenvolvimento relativamente tardio na evolução e não remonta aos organismos primitivos. De muitas maneiras, muitas pessoas pensam, incluindo o meu grande amigo e colega Daniel Dennett, que os animais são sencientes do começo ao fim: é um espectro entre organismos com mais sentiência e outros com menos sentiência.

Não creio que tal esteja certo: penso que a sentiência é como um limiar, surge num determinado estágio da evolução, porque envolve um tipo particular de mecanismo cerebral e seus respectivos ciclos de *feedback*. E só existe quando é necessário. Não servirá de nada para um animal que não pensa em si mesmo, não se relaciona com outros animais, que não tem que ser, como já disse, “psicólogos”.

Quando comecei a minha investigação há muitos anos atrás, comecei por questionar “bem, ok, então quando é que a senciência surgiu?” e percebi que há uma transição na evolução dos vertebrados que os filósofos, e, na verdade, os biólogos, não levaram muito a sério a esse respeito: quando os animais se tornaram de sangue quente.

Foi há cerca de 200 milhões de anos que os dinossauros passaram a ter sangue quente e, então, as aves e mamíferos próximos destes também tinham sangue quente. E o que isso significava era que toda a relação com o ambiente tinha mudado: tornaram-se seres autónomos, independentes do ambiente imediato e físico, podiam deslocar-se para onde quisessem, estavam activos dia e noite, e assim por diante. E acredito que isso lhes deu um novo e forte senso de individualidade, de autonomia: nesse momento, passaram a ter uma utilidade para a ideia de um Eu.

Mas a outra coisa que aconteceu – e isto foi algo de extraordinário – foi que, porque os nossos cérebros tinham aquecido, a velocidade das suas células nervosas também tinha aquecido. Quando se aumenta a temperatura de doze ou quinze graus para trinta e sete ou trinta e oito graus em humanos e mamíferos, e quarenta graus em pássaros, triplicamos a velocidade eléctrica das células nervosas. Então, tal significou que os nossos cérebros e os cérebros dos nossos ancestrais

passaram a funcionar muito mais rápido do que o de qualquer outro animal.

Acredito que tal produziu uma reorganização crucial no cérebro: permitiu o desenvolvimento de certas formas de *feedback* que simplesmente não teriam ocorrido de outra forma. Então, acredito que existe essa união, tanto para um estilo de vida diferente quanto para uma necessidade de uma forma de pensar sobre si mesmo, porque agora sou uma criatura independente do meio ambiente, que acompanhou essa mudança cerebral e que permitiu um novo tipo de imagem psicológica do que é o cérebro.

**Pergunta:** O Professor está familiarizado com o trabalho da filósofa Susan Schneider, que propôs uma espécie de teste de consciência para a inteligência artificial, para que possamos decidir se uma máquina artificial está ou não consciente. Acha que estas sugestões avançadas pela filósofa podem fornecer detalhes específicos para criar um tal teste da consciência artificial? Afirmando de outra forma, sabemos que a natureza inventou a consciência humana, mas acha que seremos capazes de criar ou inventar a consciência artificial?

**Nicholas Humphrey:** Bem, existem aí duas questões diferentes. A Susan Schneider sugere alguns testes que poderíamos usar. São testes muito bons, bastante

próximos do tipo de testes que eu sugeriria: excepto que ela não leva tão a sério o lado social da consciência. Mas para mim necessitamos dessa capacidade, pois permite entrar na mente de outras criaturas, como nós. Acho que ela deveria ter adicionado esse factor à sua lista de critérios para estar mais perto de um teste de diagnóstico, um teste de senciência numa máquina.

Sobre as máquinas, não há dúvidas neste momento. As pessoas têm especulado se essas novas linguagens como o ChatGPT poderiam ser sencientes. David Chalmers acredita – inacreditavelmente para mim – que há 10% de chance de que o ChatGPT já seja senciência. Ele é um filósofo: não acho que ele deveria fazer declarações fortes como essa. Não se pode ser 10% senciência, e não há razão alguma para pensar que qualquer coisa neste modelo de linguagem seja senciência, uma vez que não tem a necessidade, e também não tem a mecânica para isso, e não mostra nenhum dos critérios de diagnósticos da senciência.

Mas tal não significa que não possamos desenvolver uma máquina que cumpra estes critérios. No entanto, isso só aconteceria se projectássemos, deliberadamente, a máquina para ter essa capacidade: isso não vai acontecer por acaso, mas porque criámos uma máquina que processa dados cada vez mais rápido, ou que é cada vez mais inteligente.

Inteligência não é a mesma coisa que senciência e é por isso que não creio que as máquinas consigam alcançá-

la até que a possamos introduzir deliberadamente nas máquinas. E, neste momento, é claro que não sabemos como fazer isso. O que defendi no meu livro sugere o tipo de coisa que precisaríamos de construir numa máquina para que esta se tornasse senciente.

Mas então temos que perguntar: por que haveríamos de fazer isso? Uma razão pode ser: porque queremos, queremos que as máquinas tenham as mesmas capacidades de leitura de mentes e sensações e ganhar uma importância psicológica – e assim por diante – como nós seres humanos temos. Assim, poderíamos tirar a lição da natureza e aplicá-la às máquinas e quiçá desenvolver máquinas sencientes. E talvez um dia solicitemos às nossas máquinas sencientes que façam o nosso trabalho muito além da Terra, no espaço extraterrestre, por exemplo.

Os humanos nunca serão capazes de viver numa galáxia distante. Mas as máquinas poderiam servir esse propósito: poderíamos projectá-las para que chegassem a outros lugares do universo. Mas se quiserem estabelecer esse estilo de vida, se quiserem estar suficientemente interessados para começar a desenvolver uma ciência do seu novo ambiente e considerar que tipo de cultura querem desenvolver, precisarão de ter confiança na sua própria importância. Então, talvez precisemos primeiro de nos perguntar se nós, humanos, queremos que essas máquinas tenham consciência fenomenal ou não.

**Pergunta:** O Professor Humphrey argumenta também que alguns animais particulares são sencientes e outros não o são de todo. Sabemos que a “senciência” geralmente é uma espécie de propriedade moral que qualquer organismo precisa de possuir para ter estatuto moral e ser considerado um agente de direitos. Seguindo isso, que tipo de implicações éticas específicas acha que a sua investigação tem, em relação, por exemplo, à ética animal e aos direitos dos animais em geral?

**Nicholas Humphrey:** Certamente significa que, se estiver certo, devemos ser cautelosos antes de atribuímos sentiência a outras criaturas. Não devemos apenas dar-lhes o benefício da dúvida e presumir que a sentiência estará presente porque têm sistemas nervosos complexos ou porque mostram elevados níveis de inteligência, o que tende a ser o que as pessoas fazem neste momento. A maior parte dos escritos sobre sentiência e animais estão focados no quão inteligentes eles são, e não o quão conscientes eles são, no sentido de terem uma consciência fenomenal. Discuti um pouco esse aspecto no livro: não enfatizo muito quando e por que devemos levar em consideração outros animais como tendo estatuto moral apenas porque têm consciência como a nossa.

A minha opinião geral é que, como cientista e não necessariamente como cidadão, devemos assumir que a maioria dos animais não são sencientes: não creio

que as lagostas sejam sencientes, não creio que os polvos sejam sencientes, certamente não penso que as minhocas o sejam, e isso significa que podemos repensar os tipos de leis que foram agora aprovadas em todo o mundo sobre a senciência animal. Não sei qual é a situação em Portugal, mas na Grã-Bretanha foi aprovada uma lei que afirma que as lagostas são sencientes. Por consequência, é ilegal ferver uma lagosta viva. Pode haver muitos motivos para não querer ferver uma lagosta viva, mas ser senciente não é um deles, na minha opinião.

**Pergunta:** Quando considera que a senciência tenha “aparecido” nos seres humanos? Acha que começa a desenvolver-se durante a gestação, talvez até mesmo no útero materno, ou prefere estabelecer o início na primeira infância? Essa questão pode ter implicações profundas para a nossa compreensão do desenvolvimento humano e dos direitos éticos relacionados à vida e ao bem-estar.

**Nicholas Humphrey:** Essa é uma pergunta muito importante e interessante. Penso que não é na gestação ou no útero à partida, e uma das razões para afirmar isto é que os bebés humanos, quando nascem, ainda não têm o cérebro mielinizado: as bainhas de mielina não cobrem as células porque ainda não se desenvolveram, e certas áreas são disfuncionais por causa disso. Por exemplo, o córtex visual não funciona

num recém-nascido humano. E, portanto, acredito que as crianças humanas, se conseguem ver – e conseguem ver, sabemos isso pelo seu comportamento – devem ter algo semelhante à visão cega. Mas nesse ponto, elas não têm experiência fenomenal.

Novamente, faço parte de uma minoria quando afirmo isso. A grande parte dos meus colegas não acreditam que é possível ver sem ter sensações. E mesmo conhecendo estes casos de visão cega, não os levam a sério na sua prática clínica quotidiana.

Se virmos um bebé que é claramente capaz de responder ao sorriso da mãe, presumimos que o bebé está a ver o rosto da mãe da mesma maneira que nós veríamos. Não creio que isso seja verdade: penso que é por volta dos três ou quatro meses, quando o cérebro se torna mielinizado, que os circuitos seriam suficientes para sustentar uma visão com experiência fenomenal.

**Pergunta:** Para concluir, António Damásio, o neurocientista português que já publicou diversas obras sobre estes temas, fala sobre diferentes tipos de Eu, como o Eu autobiográfico, que pode ser visto como a consciência em certo sentido. Mas ele também fala do proto-Eu e do Eu nuclear, como bases do Eu autobiográfico. Onde se coloca neste esquema conceptual sobre a pessoalidade?

**Nicholas Humphrey:** Acho que eu e o António Damásio concordamos em muitas coisas, conversamos muito sobre isso em contextos diferentes. No entanto, quero ser muito mais específico do que ele ao dizer que a base do Eu são as sensações e é sobre isso que todo o resto é construído. O sujeito dos nossos estados mentais é o Eu, construído sob sensações. E se continuarmos a elaborá-lo, é claro que existimos como Eus sociais, como Eus autobiográficos, e assim por diante.

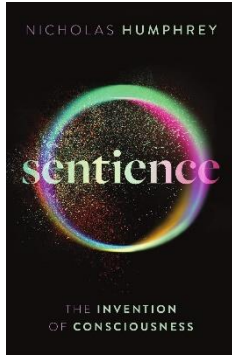
Embora eu ache que ele afirma sem grandes consequências que “os Eus existem em todos esses níveis diferentes, e cada um deles é importante, em aspectos diferentes”: acho que ele evita a questão de compreender por que é que alguém deveria ser extremamente importante para si mesmo em relação aos outros.

A minha crença é que se o indivíduo perder o Eu fenomenal, ou o Eu central, nada sobrará. Não pode manter um Eu autobiográfico ou um Eu social se perder a fenomenologia. E o interessante é que, em quase todos os casos de dissociação e de ausência, onde os Eus das pessoas se desintegram, elas permanecem realmente presentes como tendo Eus fenomenais, centrados na dor, nas cores e nas luzes, e assim por diante.

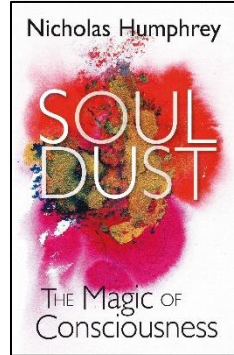
Mas não é isso que parece acontecer: se assim fosse, a pessoa não existiria. Na verdade, há casos em que os

pacientes afirmam que eles não existem. Na Síndrome de Cotard, o paciente insiste que morreu. E quando o médico fala “bom, mas não morreu mesmo, porque estou a conversar consigo agora, aqui na minha clínica”, o paciente responde; “Não, eu morri, não estou mais aí”: acho que o que eles estão a tentar expressar pode ser que o Eu fenomenal não existe mais para eles.

## Livros de Nicholas Humphrey



2023



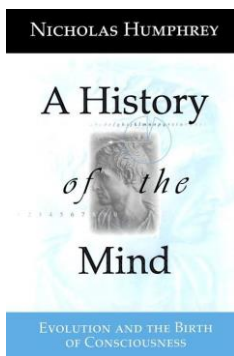
2011



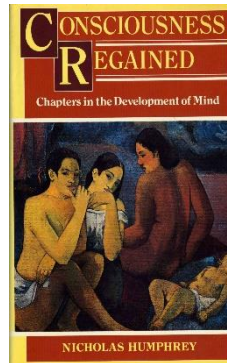
2006



2002



1998



1983



## IX. Diálogo com Sir Roger Penrose



Sir Roger Penrose é Professor Emérito de Matemática na Universidade de Oxford e Membro Emérito do Wadham College, Oxford, e membro honorário do St. John's College, Cambridge e University College London.

Foi nomeado Cavaleiro pelos seus serviços à Ciência em 1994 pela Rainha Isabel II.

Em 1969, com Stephen Hawking, Penrose provou que toda a matéria dentro de um buraco negro colapsa numa singularidade.

Recebeu vários prémios e distinções, incluindo o Wolf Prize in Physics de 1988, e o Prémio Nobel de Física de 2020.

É autor de vários livros focados na natureza da consciência, como *The Emperor's New Mind* (1989), *Shadows of the Mind* (1994) e *The Large, the Small and Human Mind* (1997), *The Road to Reality: A Complete Guide to the Laws of the Universe* (2007) ou *Cycles of Time: An Extraordinary New View of the Universe* (2013).

Mais informações:

<https://www.maths.ox.ac.uk/people/roger.penrose>

**Pergunta:** O Professor Penrose é um dos cientistas mais relevantes do século XXI, tendo ganho o Prémio Nobel da Física em 2020 mostrando que Albert Einstein estava, de facto, equivocado ao negar a existência dos buracos negros, sendo que o seu papel foi demonstrar que, matematicamente, essa existência – dos buracos negros – é uma consequência da teoria da relatividade de Einstein. Além deste seu interesse, tem também investigado a natureza da consciência. Quando é que se interessou por este tema da consciência?

**Roger Penrose:** Posso rastrear o meu interesse pela consciência quando era um estudante de doutoramento: uma das principais influências sobre a forma como penso a consciência deu-se precisamente por causa de uma série de palestras a que assisti em Cambridge quando estava a estudar (isto foi há muito tempo, como deve imaginar).

Estava a estudar Matemática Pura na época – Geometria Algébrica – e fiz três cursos que nada tinham a ver com este tema. Um desses cursos era focado na relatividade geral, oferecido por Hermann Bondi, que me influenciou bastante. Depois, fiz um curso com o grande físico Paul Dirac, sobre mecânica quântica, que também foi muito importante para mim. E o terceiro foi um curso dado por um homem chamado S.W.P Steen, que era um lógico especializado em lógica matemática e Máquinas de Turing, e a partir daí eu soube o que a

ideia de “computabilidade” significava no sentido técnico ou matemático.

Mas este professor de Lógica também descreveu, nesse curso, os dois principais Teoremas de Gödel e fiquei absolutamente surpreso com as suas ideias sobre os teoremas, pois quando ouvi falar deles pela primeira vez, foi-me passada a ideia de que haviam coisas na matemática que não podíamos provar matematicamente. De facto, não gostei dessa tese logo à partida.

O que fiz, então, foi inscrever-me neste tal curso e percebi que, afinal, não era bem esse o caso. O que acontece efectivamente é o seguinte: se usamos um certo método de prova que poderia, em princípio, ser colocado num computador, este tentaria provar ou refutar um determinado resultado.

O que podemos fazer a partir disso é produzir uma proposição – uma proposição matemática – sobre números, o tipo de coisa que o teorema pretende abordar e pode, de grosso modo, codificar a afirmação “não sou demonstrável por estas regras”. Agora, será que esta proposição é verdadeira ou é falsa? Vamos supor que é falsa: então tal indicaria que ela é demonstrável por essas regras.

Assim, o que se segue é que devemos acreditar que qualquer coisa que possa ser provada pelas regras é verdadeira: penso ser esse o ponto central do Teorema.

Nós escolhemos as regras, de modo que qualquer coisa demonstrável por essas regras deve ser verdadeira: se for falsa, então é demonstrável pelas regras e, portanto, é verdadeiro. E se for verdadeiro, então é verdadeiro e não pode ser provado pelas regras, então essa é a conclusão. Podemos ver que determinada afirmação é definitivamente verdadeira e que não pode ser provada pelas regras. Mas como é que podemos saber se ela é realmente “verdadeira”?

Sabemos se é verdadeira em virtude da fé nas regras: na crença de que as regras realmente provam coisas. Se elas indicam “sim, é verdade”, então acredita nessas regras, porque conseguimos compreendê-las – entramos de facto nas regras e realmente entendemos o que as regras significam, sendo que não as seguimos apenas por seguir.

Essa compreensão foi fundamental para mim. A compreensão do que isso significa é mais forte do que seguir as regras, sendo que essa compreensão é algo que é consciente, que requer a nossa consciência.

Seguindo estas ideias, argumentei que o pensamento consciente transcende o simples “seguir regras” e a conclusão a que cheguei é que não somos computadores: alcancei esta visão de que o pensamento consciente, seja ele qual for, não é algorítmico porque há algo mais além de seguir regras.

O meu livro *The Emperor's New Mind* (1989) foi uma tentativa de transmitir este ponto e também de explicar diferentes tópicos sobre matemática e física, que gostava muito, tentando torná-lo não tão difícil para que as pessoas pudessem ter algum interesse nele e pudessem aprender com o livro.

Achei que, quando terminasse o livro, aprenderia o suficiente sobre neurofisiologia humana para entender o que possivelmente poderia ser não computável na acção do cérebro, mas não consegui alcançar essa proeza. Cheguei à conclusão de que, embora entendesse quais eram os procedimentos, acreditava, naquela época, que era preciso levar em conta o colapso da função de onda na mecânica quântica e essa é uma parte muito relevante de todo o meu argumento: o que é que, na física, pode transcender a computação?

Segundo a teoria dominante da altura, teríamos de encontrar algo na física para explicar a consciência – e os fisicalistas pensam que o que acontece nos nossos cérebros faz parte do mundo físico – mas que parte do mundo físico é que seria relevante para a consciência? O meu argumento é que não é na física “normal” que usamos no nosso dia-a-dia, mas deve, talvez, ser algo que vai para além dessa física. Quando escrevi o livro, eu não tinha ideia do que haveria no cérebro que poderia fazer algo para além da física “normal”, então acabei por finalizar o livro sem nenhuma resposta convincente sobre a consciência.

**Pergunta:** Podemos então afirmar que a sua tese sobre a consciência é uma consequência directa do Teorema de Gödel, é isso? Visto que a sua afirmação é que a compreensão matemática é uma característica da consciência e não é computacional, isto é, não é computável por princípio ou definição.

**Roger Penrose:** Sim, é exactamente isso.

**Pergunta:** Seguindo essa tese, por que acha que tantas pessoas ligadas à Inteligência Artificial acreditam que a mente é uma espécie de computador digital e que podemos fazer engenharia reversa do cérebro humano? Por que acha que tantos pensadores – bastante inteligentes – defendem essas crenças filosóficas sobre a mente e o cérebro?

**Roger Penrose:** Acho que basicamente esses pensadores não seguiram ou não compreenderam o meu argumento. Existem várias críticas que podem ser apontadas ao meu argumento, e a mais forte é a seguinte: se não sabemos que tipo de algoritmos o nosso cérebro usa, então não podemos construir os resultados de Gödel e, portanto, o argumento não funciona.

O contra-argumento que apresento contra esta crítica no meu outro livro *Shadows of the Mind* é que, se supusermos que existe um algoritmo nas nossas cabeças – algo que não sabemos se é verdade – como

é que tal algoritmo surgiu em primeira instância? Temos que supor que veio, como tudo o resto, da selecção natural: tem de ser algo que nos oferece uma vantagem seletiva.

A imagem metafórica que usei no livro é imaginar os nossos ancestrais a fazer coisas úteis, como construir abrigos, domesticar animais, plantar... e em primeiro plano, vemos alguém a ter a ideia de como construir uma armadilha para mamutes: todas essas pessoas estão a fazer algo que tem vantagens selectivas para elas. Mas noutro plano da imagem, encontramos este pobre matemático que está a trabalhar em algum teorema e está prestes a ser devorado por um tigre dente-de-sabre. A moral desta metáfora é que fazer este tipo de matemática sofisticada não traz nenhuma vantagem selectiva, mesmo que tal possa produzir vários resultados matemáticos.

Um desses resultados matemáticos sofisticados que mais gosto é chamado de Teorema de Goodstein. É um resultado maravilhoso que pode ser explicado a pessoas que não possuem um conhecimento muito profundo de matemática – o leitor só precisa saber o que é elevar um número a uma potência em notação matemática.

O teorema de Goodstein afirma que, se seguir determinados procedimentos “A” e “B” repetidamente, sendo que cada um desses procedimentos é muito simples, tal não dura para sempre, e encontra um fim.

O problema é que isso requer tantas etapas que é completamente ridículo tentar demonstrar, de facto, o teorema, o que significa que não pode ter nenhuma vantagem selectiva. No entanto, como sabemos que esse teorema é verdadeiro?

Bem, sabemos que é verdadeiro por causa de alguns resultados maravilhosos graças ao matemático Georg Cantor e o mais incrível é que não precisamos de ser o Cantor para conseguir compreender esses resultados.

Pelo contrário, basta entender o raciocínio de Cantor e tal não exige muito de cada um de nós: requer uma pequena mudança do ponto de vista normal, mas o raciocínio real não é tão difícil de ser compreendido. Mas como sabemos disso? Como poderíamos ter evoluído se fôssemos apenas algoritmos? Não há como encontrar um algoritmo com tal sofisticação.

Isto é: a qualidade geral da compreensão, seja ela qual for – e não pretendo saber o que é ao certo – deve ser uma característica da consciência. Claro, não estou a afirmar que tal esgota a consciência, pois a consciência envolve muitos outros aspectos além da compreensão matemática, como a percepção da cor azul, por exemplo, ou a sensação de dor. Há muitos aspectos da consciência que não têm nada a ver directamente com a matemática.

Mas o meu argumento é que, se conseguimos compreender esses resultados em matemática, eles

devem estar além da computação: não podemos dizer que o Teorema de Goodstein está além da computação, porque uma vez que conhece o *input* de Cantor, poderíamos colocar uma versão particular dela num computador. Mas tal não é crível do ponto de vista da selecção natural: não há absolutamente nenhuma vantagem selectiva nisso – está muito fora do uso normal dos nossos cérebros.

O que estou a tentar defender, então, com o meu argumento, é que a qualidade geral da compreensão das coisas não é algorítmica: não estou a falar, aqui, sobre como é experienciar uma cor azul: a minha tese central é que, o que quer que seja a consciência, é algo que vai além de um procedimento computacional.

**Pergunta:** Uma das visões mais comuns na neurociência actual sobre a consciência é essencialmente a tese de que esta qualidade subjectiva é uma espécie de propriedade emergente do cérebro, uma consequência da actividade neuronal e da troca de informações entre os neurónios. Contudo, o Professor Penrose não concorda com esta visão, pois não?

**Roger Penrose:** O que penso é que não creio que exista alguma “outra coisa” misteriosa que crie a consciência nas nossas cabeças. Acho que tem a ver, fundamentalmente, com a neurofisiologia que está a acontecer dentro da cabeça: só não sabemos ainda

exactamente como é que a experiência subjectiva é criada no cérebro.

**Pergunta:** Mas em relação à natureza da consciência, o Professor acredita que a neurociência pode resolver esse problema por si só, ou acha que precisamos daquilo a que o Professor Penrose chama de “uma nova física”, de modo a conseguirmos enquadrar este fenómeno muito particular na imagem científica do mundo?

**Roger Penrose:** Sim, acho que precisamos de uma “nova” física, embora o meu argumento tenha várias partes, por exemplo: “quais são as outras partes da física que conhecemos?”

Podemos focar-nos numa das principais teorias que temos, a física newtoniana, e colocá-la num computador, como alguns investigadores fazem. E a teoria geral da relatividade de Einstein? Os cientistas descobriram como é que os buracos negros espiralam uns nos outros e que tipo de sinal sai dessa interacção – claramente, este é um procedimento computacional. Precisamos descobrir quais sistemas podem ser colocados num computador: eu diria que não é a relatividade geral.

E a mecânica quântica? Bom, na mecânica quântica existe a Equação de Schrödinger que pode ser colocada num computador também. Voltemos a um dos cursos

que frequentei quando era aluno de doutoramento de Paul Dirac. Na sua primeira aula, falou sobre o princípio da superposição: a ideia de que um átomo pode estar “aqui” ou “ali”. Na mecânica quântica encontramos estados onde o átomo está “aqui” e “lá” ao mesmo tempo: este é o princípio fundamental da mecânica quântica.

Seguindo isto, o Professor Dirac pegou num pedaço de giz e quebrou-o em dois pedaços, tentando descrever como os dois pedaços de giz podiam estar “aqui” e “ali” ao mesmo tempo, e o principal sobre este episódio é que eu estava presente na aula, mas a minha mente estava a vaguear: eu estava a olhar pela janela, a pensar em algo completamente diferente e, então, quando a minha mente voltou para a palestra, ele havia mudado para outro tema. Lembro-me dele estar a falar sobre uma energia envolvida na peça de giz, mas eu não conseguia entender o que tal tinha a ver com a superposição.

Então, não consegui ouvir a resposta, algo que se revelou, no final de contas, positivo: penso que, provavelmente, ele estava a tentar acalmar-nos de alguma forma, para não nos preocuparmos com este problema, porque as energias envolvidas são muito elevadas e não são relevantes neste nível de descrição. Mas fiquei com a sensação de que precisávamos de algo novo, precisávamos de uma “nova física”. E acho que é precisamente neste problema que precisamos de

uma nova física, já que a Equação de Schrödinger não descreve de facto o mundo, como todos os físicos sabem, mas fingem esquecer-se. Eles sabem disso, mas não o afirmam em voz alta.

Schrödinger estava muito consciente disso, pois ele descreve as pessoas após saberem que colocou esse pobre gato na caixa, colocando-o num estado de estar "vivo" e "morto" ao mesmo tempo – num estado de superposição. No entanto, as pessoas não entenderam o que Schrödinger estava a tentar dizer com essa experiência de pensamento, na minha opinião.

Os físicos costumam dizer algo como “bem, se tivesse uma experiência sofisticada o suficiente, poderia fazer um gato estar vivo e morto ao mesmo tempo, mas ainda não chegamos lá”: mas não é isto que Schrödinger estava a tentar afirmar. O que ele estava a tentar dizer era que tal cenário é simplesmente ridículo, dado que não podemos ter um gato vivo e morto ao mesmo tempo, e tal indica que há algo de seriamente errado com a sua própria equação (de Schrödinger). Na verdade, ele estava a tentar argumentar contra a sua própria equação: ele estava a afirmar que a Equação de Schrödinger não explica como o mundo funciona na sua totalidade.

**Pergunta:** E o Professor acha que há um hiato ou uma lacuna na mecânica quântica, que há algo que está em

falta para que a mecânica quântica seja uma teoria completa?

**Roger Penrose:** Sim, acho que há uma lacuna enorme. Mas é uma espécie de lacuna independente, dado que este é um argumento da física. O que estou a dizer é que há uma lacuna na mecânica quântica e talvez seja aí que resida a tal não-computabilidade que descrevi anteriormente.

Quando escrevi o *The Emperor's New Mind*, esse era o tom que estava a tentar transmitir. O ponto fraco do livro era que eu não tinha ideia de como os neurónios podiam fazer isso. Achei que, no final do livro, teria aprendido o suficiente sobre neurofisiologia para ver onde poderia haver uma lacuna desse tipo, e não consegui isso. Então meio que desisti do livro, mas tinha que terminá-lo de alguma forma e não fiquei muito feliz com tudo o que acrescentei na altura.

Foi aí que o Stuart Hammeroff, que tinha lido o meu livro, decidiu escrever-me uma carta onde especificava certas estruturas anatómicas que eu não conhecia, que são os microtúbulos e ele tinha uma teoria que foi ganhando apoio de outros ângulos de investigação, como por exemplo, de que forma os anestésicos podem criar diferentes estados de consciência. O ângulo de Hammeroff era o seguinte: há algo que consegue desligar de uma forma reversível a consciência – podemos ligá-la e desligá-la e ligá-la de novo. Como? Por exemplo, podemos ser induzidos em

coma e ficar sem consciência, e podemos acordar depois com a consciência a regressar.

Sendo um anestesiológico, o que Hammeroff está a tentar fazer não é somente deixar os pacientes inconscientes e depois acordá-los novamente, mas ele está a tentar entender o que realmente está a acontecer com essas substâncias que “desligam” a consciência. Tal é muito interessante porque a conclusão não é um processo químico: o que une essas diferentes substâncias está relacionado com um processo físico, e não com um processo químico.

E a pergunta que surge aqui é: o que está de facto a acontecer? O interessante é que, embora eu e o Hammeroff nos tenhamos encontrado pessoalmente na década de 1990, nessa altura, esta abordagem era considerada um ponto de vista muito controverso. No entanto, está a tornar-se numa das visões que os cientistas e filósofos levam a sério, fora os típicos pontos de vista que são normalmente considerados sobre a consciência, como o dualismo ou o reducionismo.

Acho que isso, para mim, é bastante reconfortante: saber que as pessoas estão a levar a sério esta nossa teoria sobre a consciência, havendo já experiências muito interessantes sobre a ação geral dos anestésicos na consciência e que estruturas específicas do cérebro são afectadas por essas substâncias. Parece que estas estruturas – os microtúbulos – podem

também ser um ingrediente importante para toda esta história da consciência: as coisas estão a evoluir para uma direção em que poderemos de facto confirmar se estamos perante uma teoria da consciência correcta ou não.

**Pergunta:** Vamos aprofundar-nos um pouco mais no seu modelo teórico da consciência. Basicamente, o Professor é um crítico da chamada “interpretação de Copenhaga” da mecânica quântica, pois argumenta que o que acontece é contrário do que essa interpretação descreve: temos o colapso da função de onda, mas não precisamos de nenhum sujeito consciente para que tal aconteça, não precisamos de um observador. Seguindo isso, como é que pode surgir a consciência deste processo de colapso da função de onda através da influência das forças gravitacionais, que é o que o Professor defende?

**Roger Penrose:** Sim, como o Steven disse, é o contrário. Há vários ingredientes que avanço: um que não mencionei é que o colapso da função de onda, a meu ver, tem a ver com a combinação da gravidade com a mecânica quântica. Este ponto particular não foi ainda confirmado ou refutado experimentalmente, mas existem algumas experiências destinadas a resolver esse problema. Há também razões teóricas para defender este ponto de vista.

Especificamente, tenho um argumento que apresentei há algumas décadas que procura mostrar que existe uma incompatibilidade entre os princípios fundamentais da mecânica quântica, que envolve o princípio da superposição que mencionei, e o princípio da relatividade geral, que é o Princípio de Galileu, que basicamente afirma que podemos cancelar a gravidade em queda livre.

Galileu imaginou que, se uma pedra grande e uma pequena caíssem da torre inclinada de Pisa sem atmosfera, elas cairiam juntas à mesma velocidade, e se houvesse um insecto sentado numa das pedras a observar a outra durante a queda, esse insecto pensaria que não existia gravidade. Esta experiência mental mostra que podemos cancelar a gravidade em queda livre e esse é o princípio da equivalência, que é a pedra fundamental da relatividade geral de Einstein, mas essa pedra fundamental é inconsistente com as pedras fundamentais da mecânica quântica, nomeadamente com o princípio da superposição.

O meu argumento é que há um conflito aí, e para resolvermos esse conflito, será necessária uma nova teoria. Embora não saiba qual é essa tal nova a teoria, o que posso dizer é uma estimativa de quando essa teoria entra em função, e em que tipo de escala poderíamos começar a observar o colapso da função de onda sendo um processo físico que podemos realmente medir.

Mas como o Steven disse, é o contrário do que muitos pensavam nos primórdios da mecânica quântica, ou seja, que o colapso da função de onda, que é quando fazemos uma medição, depende do observador, de um ser consciente. Pessoas como John von Neumann e Eugene Wigner defenderam este tipo de visão. Na verdade, conversei pessoalmente com Wigner sobre isso numa ocasião e descobri que ele não era tão dogmático quanto algumas pessoas pensam: ele certamente estava aberto a outras alternativas.

Wigner parecia pensar que este era um ponto de vista que deveria ser levado a sério, mas apenas até certo ponto, porque podemos notar que há realmente uma inconsistência e que não podemos confiar em seres conscientes para colapsar a função de onda. Acho que tem que ser um processo físico, mas como disse, acho que é o cérebro que, fazendo uso desse processo físico, produz a consciência.

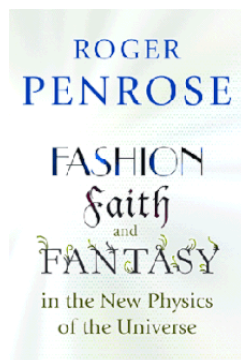
**Pergunta:** Para concluir e olhando para o fenómeno da espiritualidade, existem tantos livros baseados em temas como sincronicidade quântica, chakras, chi, entre muitos outros. Se o Professor pudesse, muito brevemente, dizer-nos o que pensa sobre estas abordagens e esses livros que são sucessos de venda.... Acha que existe alguma coisa à qual possa dar validade?

**Roger Penrose:** Não sou adepto de nenhuma dessas coisas: confesso que não li muito nesse sentido, mas a minha intuição é achar que não acrescentam muito. Podemos aprender uma ou outra coisa sobre a consciência observando essas ideias, mas concentrei-me apenas num aspecto muito limitado da consciência, que é essa qualidade de compreensão, especificamente, quando se trata de ideias matemáticas. Há muito mais para compreender a experiência consciente do que essa pequena área em que me foquei.

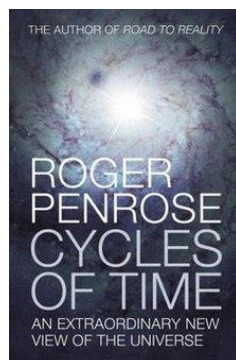
Tenho-me afastado dessas ideias principalmente porque não vejo como podemos falar delas de uma forma coerente e precisa. Acho que esse é o problema, sou matemático: gosto de poder juntar as coisas numa teoria onde posso ver as consequências a surgir, que são determinadas com precisão de maneiras claras por essa mesma teoria.

Essas coisas que mencionou, não as consigo ver nesta forma, então não perco o meu tempo a estudá-las. Não vejo nenhum benefício, pelo menos para mim, embora possa haver benefício para outras pessoas.

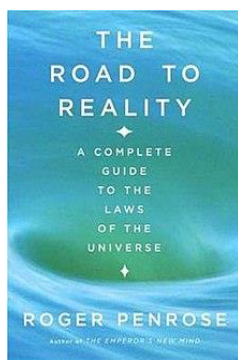
## Livros de Sir Roger Penrose



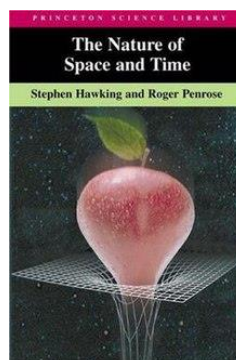
2016



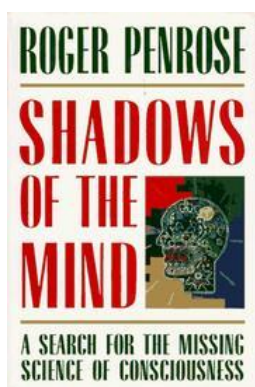
2010



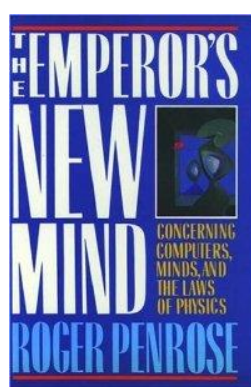
2004



1996



1996



1989



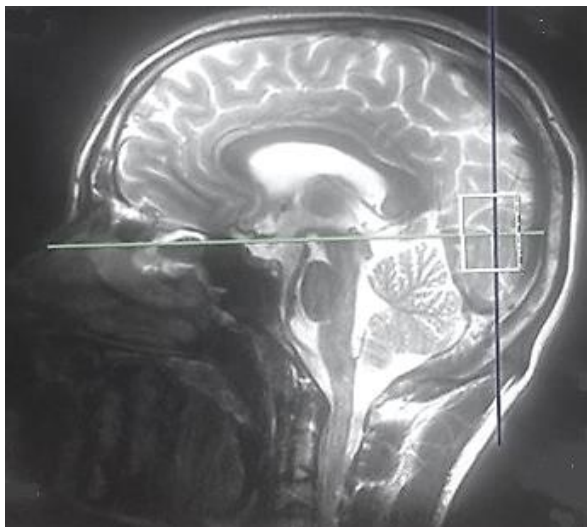
# **PARTE II**

## **Cérebro**

# PARTE II

## Cérebro

## I. INTRODUÇÃO: CÉREBRO



1. Imagem fMRI de um cérebro.

A natureza filosófica e científica do cérebro – como o cérebro da imagem acima – tem sido fonte de debate por centenas de anos em diversas disciplinas. Conhecemos actualmente muito mais sobre o cérebro do que os nossos avós conheciam: por exemplo, sabemos que, em média, pesa 1,5kg; sabemos que é feito de vários tipos diferentes de neurónios que partilham informações electroquímicas entre eles; e sabemos que esses neurónios podem criar diferentes redes neuronais que são relevantes para diversas

funções mentais, como memória de longo prazo, as emoções ou a criatividade.

O cérebro humano possui cerca de 86 biliões de neurónios conectados através de triliões de sinapses, estando dividido em várias regiões distintas. O núcleo mais primitivo, semelhante às antigas estruturas da evolução, é responsável pelos instintos básicos de sobrevivência. Logo de seguida, encontramos o sistema límbico, responsável pelas emoções e pela formação da memória. Por fim, encontramos a superfície do cérebro, o chamado córtex cerebral que se encontra dividido em vários lobos, responsáveis pelos processos cognitivos complexos.

Apesar de, presentemente, conhecermos bastante mais sobre o cérebro – embora também saibamos ainda muito pouco, como veremos nesta segunda parte do livro –, este órgão teve uma história controversa desde que a humanidade começou a pensar filosoficamente sobre si mesma como um ser constituído de carne e osso. Particularmente, dois grandes pensadores teceram algumas considerações sobre a ligação do cérebro à mente humana.

Num debate fervoroso, Aristóteles de Estagira defendia, no livro III do seu *Das Partes dos Animais*,<sup>63</sup> que o coração era a sede e a fonte das sensações, da consciência, do intelecto e tudo o que era relevante no

---

<sup>63</sup> Versão online: [penelope.uchicago.edu/aristotle/parts3.html](http://penelope.uchicago.edu/aristotle/parts3.html)

ser humano, acreditando que o cérebro teria uma única função: arrefecer a temperatura do sangue. Contra esta perspectiva, Hipócrates de Cós defendia no seu livro *Da Doença Sagrada*<sup>64</sup> que a perspectiva aristotélica estava equivocada: o cérebro deveria ser considerado como a fonte do prazer, alegria, da tristeza ou da dor, em vez do coração.

Apesar da Grécia Antiga ser um ponto de referência histórica aliciente, especula-se que as primeiras considerações conhecidas sobre o cérebro humano não comecem com os gregos antigos, como é comumente assumido, mas antes.

No final da Idade da Pedra, um período que começou há cerca de 50.000 anos na Europa e progrediu para o período do Neolítico – uma era que vai de 3.000 a 2.000 A.C. –, a humanidade desenvolveu práticas agrícolas inteligentes e domesticou determinados animais.

Nesse período, surgiram também a invenção de ferramentas de pedra polida, a criação de cerâmica e a formação de pequenos povoados, o que permitiu que algo absolutamente fascinante pudesse acontecer.

Do mesmo período, foram encontrados registos fósseis das primeiras perfurações do crânio humano realizadas por procedimentos cirúrgicos com ferramentas especializadas de pedra, o que indica que o homem neolítico usava vários tipos de instrumentos

---

<sup>64</sup> Versão online: [www.classics.mit.edu/Hippocrates/sacred.html](http://www.classics.mit.edu/Hippocrates/sacred.html)

afiados para realizar uma operação conhecida como “trepanação” (do grego *trypanon* que significa “furar”), que é a primeira prática cirúrgica conhecida a ser realizada pela humanidade.<sup>65</sup>

O primeiro crânio reconhecido como perfurado foi descoberto acidentalmente em 1864 por causa de... fezes de pássaro. Qual seria a ligação entre crânios e fezes? Em 1861, a Guerra Civil americana explode e passa a ser necessário garantir fertilização para a produção de agricultura, tornando-se uma das prioridades de Abraham Lincoln.

Curiosamente, os melhores fertilizantes do mundo, na época, vinham da América do Sul, com a extracção de uma substância de nome “guano” que provem das fezes de variados animais, sendo rica em fósforo e nitrogénio. Assim, Ephraim G. Squier, um arqueólogo e jornalista, é mandado por Lincoln para o Peru, para garantir a produção e envio do fertilizante para os Estados Unidos.

Depois de concluir a sua tarefa, Squier decidiu explorar o país, tendo chegado ao cemitério inca no Vale de Yucay, onde foi encontrado um crânio perfurado que possuía na sua constituição um buraco rectangular de 15x17mm, algo que o deixou espantadíssimo, dado que

---

<sup>65</sup> Baseado em: Weber, J. e Wahl, J. (2006) “Neurosurgical aspects of trepanations from Neolithic times”, *International Journal of Osteoarchaeology*, 16: 536–545.

não é comum encontrarmos ângulos rectos na natureza. Com esta descoberta, Squier decidiu voltar para casa, e apresentou o artefacto pela primeira vez à Academia de Medicina de Nova York em 1865.

Infelizmente, não foi “levado” a sério, dado que os cientistas dessa organização não acreditavam ser possível um conhecimento cirúrgico tão avançado de povos indígenas, numa clara posição racista.

Apesar disso, Squier não desistiu de analisar o crânio, tendo decidido enviá-lo para Paris, onde residia aquele que era considerado, na altura, a maior autoridade científica sobre o estudo do cérebro, o famoso fundador da primeira Sociedade Antropológica de França, o médico e professor Paul Broca.

Broca ficou totalmente surpreso ao analisar o crânio ao detalhe e perceber que a perfuração contida no mesmo teria sido claramente produzida por uma ferramenta de corte especializada e envolvia algum tipo de protocirurgia, algo que era considerado totalmente inviável até então.

Este foi o primeiro de muitos crânios perfurados que foram descobertos nas décadas seguintes, sendo que grande parte destes revelou um detalhe ainda mais espantoso: evidências de crescimento ósseo ao redor das perfurações. Ora, o que tal significa? Significa isso mesmo: que os indivíduos sujeitos as estas cirurgias

primitivas poderão ter sobrevivido vários meses ou até anos, o que é extraordinário.

Uma outra fonte histórica pode ser encontrada na região do Nilo, no Antigo Egito. As Grandes Pirâmides e os diversos procedimentos funerários eram indicadores que a vida após a morte era algo de grande relevância para os antigos egípcios.

Ora, para se garantir uma passagem bem-sucedida ao “outro mundo”, os egípcios acreditavam que, para garantir que a alma chegasse a bom porto, o corpo teria também de ser preservado. É esta crença que explica todo o processo de mumificação que ficou famoso em diversas obras literárias e de cinema.

De forma interessante, a parte do corpo mais importante era o coração, dado que se acreditava representar o Eu da pessoa. Os intestinos, pulmões, fígado e estômago também eram considerados importantes – sendo embalsamados e armazenados em potes canônicos ao lado da múmia.

No lado oposto da relevância encontrava-se o... cérebro que, imagine-se, era simplesmente deitado fora e extraído pelas narinas através de um gancho de ferro. Até onde sabemos, os antigos egípcios atribuíam pouca importância ao cérebro na sua relação com a mente ou a consciência.

Através de vários escritos egípcios – particularmente aqueles que adornam as paredes dos túmulos –

sabemos, no entanto, que o coração não era apenas tido como o repositório das acções terrenas da alma, mas possuía também capacidades cognitivas e conativas (emoções e sentimentos).

Esta primazia do coração foi mantida, curiosamente, até aos tempos bíblicos. De facto, quando o Antigo Testamento foi traduzido para grego, foi aceite que o intelecto e as emoções do Homem residiam no coração e não no cérebro, sendo que não existe nem uma única referencia ao cérebro na Bíblia.<sup>66</sup>

No entanto, apesar deste descrédito pelo cérebro, podemos possivelmente encontrar nos antigos egípcios a mais antiga referência escrita ao cérebro: um papiro de teor médico conhecido como o “papiro de Edwin Smith” que descreve 48 tipos de lesões na cabeça e pescoço, juntamente com conselhos sobre tratamento e intervenção cirúrgica, e que contém referências a textos escritos até 3.000 A.C.

Os autores do papiro aparentam ter uma certa compreensão da função do cérebro, dado que descreveram determinadas lesões na cabeça e as correlacionam com vários sintomas, como paralisia ou perda da fala, embora descrevam também erros grosseiros, como o facto de uma lesão no hemisfério

---

<sup>66</sup> Baseado em: York, G.K. e Steinberg, D.A. (2010) “Neurology in Ancient Egypt” In Finger, S., Boller, F. e Tyler, L. (eds) *Handbook of Clinical Neurology*, vol. 95, Elsevier: Amsterdam.

direito causar lesões no lado direito do corpo quando sabemos que, de facto, o lado esquerdo deveria ser afectado.

Uma outra referência histórica fascinante sobre o cérebro pode ser encontrada nas aventuras de Aquiles e Ulisses, ou seja, nos textos homéricos da *Ilíada* e da *Odisseia*, que inspiram este livro. Para os gregos, o Eu seria composto de diversas forças, sendo uma delas a psique, a força vital que permitiria manter a pessoa viva. Homero acabaria por identificar o mental em várias outras “formas” de alma que residiam no peito.

A mais importante dessas formas seria o *thymos*, mencionado mais de 450 vezes na *Ilíada*, que ficaria localizado no diafragma, sendo a fonte de emoções que levam alguém a agir. Já as capacidades intelectuais fariam parte do *noos*, que estaria localizado no peito. Na mesma linha dos egípcios, os textos homéricos não atribuem uma grande importância ao cérebro.

Há, no entanto, uma influência linguística muito importante de Homero, que introduziu no vocabulário três conceitos fundamentais: (1) *enkephalos*, o próprio cérebro; (2) *muelos*, referindo-se à coluna vertebral (de onde tiramos a palavra 'medula'); (3) *sinew*, que veio a originar o conceito de “neurónio”.<sup>67</sup>

---

<sup>67</sup> Baseado em: Singer, C. (1957) *A Short History of Anatomy and Physiology from the Greeks to Harvey*, Dover: New York.

Apesar desta desconsideração do cérebro, podemos encontrar na Antiga Grécia alguns pensadores que, contrariamente a esta crença partilhada pelos egípcios, defenderam que este órgão particular deveria ter maior relevância que aquela que estaria a ser dada até então.

Um desses pensadores era Alcmaeon de Crotona, que considerava o cérebro como o órgão das sensações, embora todos os seus textos tenham sido perdidos e não nos tenha chegado nenhuma referência directa aos nossos dias.

Apesar disso, chegaram-nos várias referências indirectas que apontam para o brilhantismo protoanatômico deste filósofo. Por exemplo, Teofrasto de Eressos afirmava que Alcmaeon teria sido o primeiro a estudar a anatomia de vários animais através de técnicas de dissecação dos mesmos, tendo descoberto que existiam dois canais que ligariam fisicamente a parte posterior do olho ao cérebro, o que conhecemos actualmente como os nervos ópticos, tendo levado a defender que os sentidos estariam todos conectados ao cérebro.

Outra referência indirecta pode ser encontrada em Aécio, contemporâneo de Alcmaeon, que defendia que o segundo era conhecido por defender que a inteligência, tal como as sensações, estariam relacionadas com o cérebro humano. Devido a estas e outras descobertas, o impacto das investigações de

Alcmaeon são comparáveis às descobertas de Copérnico e Darwin.<sup>68</sup>

Como vimos anteriormente, a outra grande referência grega sobre a relevância do cérebro é o fundador e pai da Medicina, Hipócrates. No *Corpus Hippocraticum*, são múltiplas as referências ao cérebro, sendo que o texto mais relevante é o *Da Doença Sagrada*, que se foca em grande parte numa tentativa de compreender a natureza da epilepsia.

Hipócrates é o primeiro a tentar demonstrar que esta doença não poderia ser explicada pela teoria médica em vigor até então, a chamada “Teoria da Possessão Demoníaca”, onde se defendia que a epilepsia estaria ligada a uma forma de possessão de um demónio particular como punição divina por algum pecado cometido naquela pessoa.

O fundador da medicina tentou contrariar esta teoria, argumentando que era uma noção defendida por charlatões e curandeiros que não queriam realmente saber de um tratamento para esta doença, defendendo que a epilepsia era uma doença do cérebro causada por um excesso de fleuma que bloquearia o fluxo de ar nos vasos sanguíneos. Ora, tal bloqueio só poderia ser corrigido através de uma convulsão epilética cuja finalidade seria desbloquear esse bloqueio.

---

<sup>68</sup> Baseado em: Mithen, S. (1996) *The Prehistory of the Mind*, Phoenix Books: Guernsey.

Além desta explicação naturalista (que não apela a demónios obscuros!), Hipócrates acabaria por argumentar que o cérebro é responsável por toda a nossa actividade mental. Vale a pena atentar às palavras do filósofo e médico:

“Deve ser do conhecimento geral que a fonte do nosso prazer, alegria, riso e diversão, bem como da nossa tristeza, dor, ansiedade e lágrimas, não é outra senão o cérebro. É este órgão que nos permite pensar, ver e ouvir, e distinguir o feio do belo, o mau e o bom, o agradável e o desagradável... é também o cérebro que é a sede da loucura e do delírio, dos medos e sustos que nos assaltam... é aí que reside a causa da insónia e do sonambulismo, dos pensamentos que não virão, dos deveres esquecidos e das excentricidades...”<sup>69</sup>

Este relato notável feito 400 anos AC poderia ter sido retirado de qualquer livro de neurociência deste século, que aponta o envolvimento do cérebro na percepção sensorial, julgamento e emoção, bem como a sua associação com distúrbios mentais. No livro referido, Hipócrates também demonstra os seus dotes anatómicos incríveis, onde fornece várias descrições anatómicas do cérebro humano, defendendo que este é semelhante a muitos animais, e que está dividido por

---

<sup>69</sup> Publicação original: Chadwick, J. e Mann, N. (eds.) (1983) *Hippocratic Writings*, Penguin: London.

uma estrutura que o divide em duas metades, uma clara referência ao corpo caloso e aos dois hemisférios.

Platão também tentou compreender o papel do cérebro no ser humano, ligando-o ao conceito de alma. Na filosofia platônica, “alma” distingue-se do sentido dado normalmente na religião, sendo que esta possuía uma estrutura tripartida: o *epithymetikon*, o *thymos* e o *logistikon*. A primeira estava associada ao fígado e ao intestino, onde estariam as necessidades vegetativas básicas do indivíduo, enquanto o *thymos* estava localizado no coração que instigava emoções como raiva, medo, orgulho e coragem.

Platão acreditava que estas duas partes distintas da alma faziam parte também dos outros animais, e que se extinguíam no momento da morte. No entanto, o *logistikon* distinguia-se destas duas formas: esta era tida como uma força espiritual única para os humanos, fornecendo ao indivíduo o pensamento e a inteligência, sendo imortal e capaz de reencarnar.

Ora, a parte mais interessante desta distinção tripartida é que Platão acreditava que esta última parte residia no cérebro. Por exemplo, no livro *Fédon*, Platão afirma que: “o cérebro é o poder originador das percepções da audição, da visão e do olfato, e a memória e a opinião podem vir dele”. Já no livro *Timeu*, Platão escreve: “a

cabeça... é a parte mais divina e domina o resto do corpo.”<sup>70</sup>

Contra esta visão platônica podemos encontrar, novamente, Aristóteles. Como já referimos muito brevemente, o pensamento aristotélico não dava muita importância ao cérebro. De forma interessante, essa era uma conclusão que usava dados empíricos e observações, mas que eram, todavia, interpretadas de forma equivocada pelo Estagirita.

Aristóteles era fascinado pelo funcionamento do corpo humano, tendo sido um dos primeiros seres humanos a dissecar uma multiplicidade de animais em vários dos seus estágios de desenvolvimento, o que incluía peixes, répteis, mamíferos e até elefantes.

Fez várias descrições anatómicas, observando duas membranas que cobriam o cérebro – o que sabemos agora serem as meninges (dura-máter e pia-máter) – e observou também uma estrutura localizada na “parte de trás” do cérebro a que chamou de *parencephalis*, que tinha um aspecto muito diferente, quer na textura como no aspecto, que sabemos agora ser uma referência ao cerebelo.

Apesar destas descrições precisas, Aristóteles não atribuiu nenhum papel relevante ao cérebro dado não

---

<sup>70</sup> Baseado em: Crivellato, E. e Ribatti, D. (2007) “Soul, mind, brain: Greek philosophy and the birth of neuroscience”, *Brain Research Bulletin*, 71: 327–336.

ter ficado encantado com a sua estrutura uniforme e fria, tendo relegado maior importância ao coração, tendo em conta que este era quente e irregular. Note-se que uma das crenças fundamentais daquela época estava ligada à tese de que o calor era essencial para a vida, dado que os corpos vivos eram quentes, e os cadáveres frios.

Curiosamente, esta tese fundamental tinha origens “empíricas”: Aristóteles teria examinado os vários estágios embrionários do ovo da galinha e teria concluído que o primeiro órgão a emergir seria o coração. O cérebro era então considerado uma espécie de ferramenta biológica de arrefecimento devido a outra observação empírica: o facto de o calor subir.

Tal levou o autor da *Metafísica* a deduzir que os múltiplos vasos sanguíneos que cobriam o cérebro serviriam para arrefecer o sangue do coração, o que o levou também a explicar o tamanho do cérebro humano: como os seres humanos eram mais quentes que outros mamíferos e animais em geral, necessitavam de um maior “aparelho” para diminuir a temperatura sanguínea.<sup>71</sup>

Poucas décadas depois destas experiências rudimentares de Aristóteles, surgia um outro filósofo na cidade de Alexandria, de seu nome Herófilo, com

---

<sup>71</sup> Baseado em: Gross, C.G. (1995) “Aristotle on the Brain”, *The Neuroscientist*, 1 (4): 245–250.

contribuições admiráveis para a anatomia cerebral. Reconheceu, por exemplo, que haveriam ligações muito específicas entre diversas partes do corpo e o cérebro, através da medula espinal, o que chamamos actualmente de nervos cranianos, tendo descrito sete pares, e tendo igualmente formulado as suas origens: o facial, auditivo, óptico, hipoglosso, trigêmeo e o oculomotor.

Assim, Herófilo contribuía para a tese de que era o cérebro – e não o coração! – responsável pelas funções mentais em geral. Pelas suas contribuições verdadeiramente notáveis, ficou conhecido como o primeiro grande anatomista, tendo fundado a Escola de Medicina de Alexandria, onde praticava grande parte das investigações em cadáveres.

Depois de uma passagem pelo Egipto Antigo e pela Grécia Antiga, chegamos agora à próxima paragem nesta viagem histórica: o tempo do Coliseu de Roma e do Império Romano, onde encontramos Galeno de Pérgamo, considerado o fundador da fisiologia que, ao contrário de Herófilo, estava proibido de dissecar cadáveres humanos, tendo usado diversos animais para concluir as suas descobertas.

Galeno teve a proeza de diferenciar uma constituição dupla do sistema nervoso: por um lado, era constituído por vias nervosas que entravam e saíam do cérebro através da base do crânio; por outro, existiam também um conjunto de vias nervosas conectadas através da

medula espinal. Admiravelmente, na sua obra *Dos Procedimentos Anatômicos*, Galeno descreve 10 dos 12 pares de nervos cranianos que cada ser humano possui, sendo que poderá ter confundido alguns deles entre si.

Estas descobertas – e outras descobertas, durante todo o período medieval – abriram caminho para a Modernidade que, com Descartes, veio “dividir” o mundo em duas grandes substâncias: a parte material, que representava a maioria das entidades existentes, e a parte imaterial, que seria composta pela alma ou mente humana. Chamamos a esta forma de pensar o mundo como “dualismo de substâncias”, porque divide o mesmo numa dualidade específica.

Assim, Descartes iria conceber o ser humano como dual neste preciso sentido: o corpo e o cérebro seriam constituídos de algo material e físico, e a alma ou a mente seria imaterial e não-física. O que se segue deste esquema dualista? Esta forma de pensar a natureza do ser humano dualista cria um problema: o problema de explicar a interacção entre duas substâncias de natureza diferente. Se a mente é imaterial, como poderia ela afectar algo material como o corpo ou o cérebro?

Descartes enfrentou esse problema e, seguindo as pisadas dos primeiros anatomistas, produziu um estudo muito rudimentar do cérebro humano, tendo observado que, anatomicamente, a maioria das

estruturas cerebrais vinham em pares (por exemplo, hemisfério direito e esquerdo). No entanto, uma estrutura específica parecia ser unitária: a glândula pineal. Através desta descoberta, Descartes passaria a argumentar que a interacção se daria nesse órgão particular: é a glândula pineal que garante a interacção causal entre o mental (imaterial) e o físico (material).

O autor da obra *Princípios de Filosofia* era, além de filósofo e matemático, tutor da princesa Elisabete da Boêmia, a quem Descartes acabaria por escrever a dedicatória desse seu livro. Extraordinariamente, Elizabeth levantaria uma das mais fortes objecções ao dualismo: como pode algo imaterial (como a alma) ser afectada causalmente por algo material (o corpo e o cérebro)?

Como poderia uma crença (mental) como “quero parar de ler este livro” levar a que o leitor feche o livro (um acto físico, corporal)? Esta objecção foi tão forte que fez com que vários filósofos tivessem de apresentar outras soluções para resolver o problema.

Uma forma de negar este dilema é aceitar a objecção e dizer que, de facto, não existe uma interacção entre o físico e o mental. Nicholas Malebranche, filósofo francês, argumentou neste sentido: o mental e o físico relacionam-se através de uma relação ocasional – chamamos a esta posição de “ocasionalismo”. Já Gottfried W. Leibniz argumentou também nesta linha,

mas de forma diferente de Malebranche: a relação mente-corpo dar-se-ia numa relação de “paralelismo”.

O problema com esta duas respostas à objecção de Elisabeth é que ambas postulam a existência de um deus para suportar o argumento. Por um lado, o ocasionalismo argumenta que deus é responsável por ordenar o mental e o físico constantemente em cada momento: quando o leitor forma a intenção de fechar o livro, deus entra em acção e garante que o seu corpo faça essa acção acontecer.

Por outro lado, o paralelismo defende que deus não actua em cada momento, mas que este criou, no início do universo, uma sincronização perfeita entre o mental e o físico para que *pareça* que haja uma causalidade entre eles, como se fossem dois relógios sincronizados na perfeição, numa harmonia pré-estabelecida.

Assim, sempre que o leitor forma a crença “é desta que vou mesmo fechar este livro!”, o seu corpo fecha o livro, e tal dá-se sem causalidade, mas por mera sincronização: imagine que a linha do mental e a linha do físico são duas linhas paralelas que nunca se tocam, mas que estão perfeitamente sincronizadas uma com a outra.<sup>72</sup>

Ora, estes problemas com o dualismo levantaram uma suspeita sobre o mesmo, levando os filósofos e

---

<sup>72</sup> Baseado em: Gouveia, S. (2018) *Reflexões Filosóficas: Arte, Mente e Justiça*, Braga: Editora Húmus.

pensadores a ponderar uma outra posição sobre o mental e o físico: a tese monista de que, talvez, estes sejam constituídos por uma e a mesma substância, a material.

Esta nova posição origina a possibilidade de estudarmos e investigarmos a natureza do cérebro e da mente de um ponto de vista científico, baseado em modelos e observações rigorosas e informadas, o que levou à fundação de uma nova disciplina: a Neurociência.

Os primeiros passos desta disciplina começaram em 1810 com o projecto de Frenologia de Franz-Joseph Gall e J. G. Spurzheim. A frenologia tentaria localizar, através das formas e irregularidades do crânio humano, diversas funções, desde a linguagem à percepção ou consciência.

Contudo, este projecto “localizacionista” foi criticado pouco depois pelo fisiologista Jean-Pierre Flourens que, a partir de França, rejeitaria a ideia fundamental de que certas funções estavam circunscritas a determinadas regiões do cérebro. Já Hughlings Jackson, neurologista em Inglaterra, viria a defender que, baseado no seu trabalho com pacientes epilépticos, diversas funções seriam responsáveis por variadas regiões do cérebro humano.

Um pouco mais tarde, em 1861, surge o famoso caso relatado por Paul-Pierre Broca que aparentava

corroborar a tese de que uma lesão numa área específica do cérebro – no lobo frontal esquerdo, que ficaria conhecido posteriormente como “Área de Broca” – estava relacionada com uma lesão particular chamada afasia, ligada à inabilidade de compreender e expressar a linguagem.

Já na Alemanha, Karl Wernicke viria a especializar-se nesse tema, doutorando-se em 1876 com uma tese focada numa vítima de derrame cerebral que teria afectado uma outra região relevante do cérebro para a linguagem, que viria a ficar conhecida como a “Área de Wernicke”.

Um pouco mais tarde, em Itália, surge Camillo Golgi que estava a desenvolver uma técnica específica para identificar a estrutura dos neurónios que, pela primeira vez, permitiu a visualização de um único neurónio.

Utilizando o mesmo método, Santiago Ramón y Cajal, em Espanha, contribua com a ideia de que os neurónios eram estruturas unitárias que transmitiam informação apenas dos dendritos para os axónios, ao contrário do que pensava até então. Estas – e outras – descobertas levaram filósofos e cientistas a formular uma vinculação cada vez mais precisa entre aspectos mentais e partes específicas do cérebro.

O estudo de doenças neurológicas também desempenhou um papel crucial no avanço da neurociência. Erwin Strauss, Kurt Goldstein e Hartmut

Kuhlenbeck dedicaram-se à análise dos cérebros "doentes" de diversos soldados da Primeira Guerra Mundial. Inicialmente, focaram-se na localização específica das lesões e examinaram o impacto no comportamento e as funções associadas a essas lesões nesses soldados.

Curiosamente, numa abordagem única, questionaram os pacientes sobre de que forma é que as suas lesões estaria a influenciar as experiências corporais e as suas relações com o mundo, estabelecendo assim uma conexão indirecta entre descrições fenomenais da consciência e características neuronais do cérebro.

O avanço no desenvolvimento de tecnologias específicas para explorar o cérebro de forma mais directa foi também fundamental para o desenvolvimento de vários conhecimentos centrais do sistema nervoso. As investigações do alemão Hans Berger permitiram registar pela primeira vez a actividade eléctrica no crânio de seres humanos.

Esta técnica viria a ser melhorada ao longo de vários anos até que, na década de 1930, Berger desenvolve a técnica da electroencefalografia (mais conhecida por EEG), que permitia a identificação de vários ritmos cerebrais particulares. Por exemplo, com menor frequência, encontramos a onda delta, que se encontra no intervalo dos 0.5-4 Hz, e está associada a estágios de sono profundo ou, com maior frequência, a onda gama,

que se encontra no intervalo dos 30-100 Hz, associada a processos superiores, como a memória, a percepção ou a aprendizagem.

Por fim, outras tecnologias relevantes para o estudo do cérebro foram-se desenvolvendo nas últimas décadas, que incluem a tomografia por emissão de positrões (PET) e a ressonância magnética funcional (fMRI). Estas tecnologias não medem, directamente, a actividade eléctrica, mas, em vez disso, avaliam a actividade bioquímica ou metabólica nos neurónios.

Por exemplo, se usarmos fMRI, podemos quantificar com alguma precisão a quantidade de moléculas de oxigénio libertadas no neurónio e, baseado nesse dado, inferir o valor da actividade eléctrica. Já se usarmos PET, estaremos a fazer uso de substâncias radioactivas para identificar receptores específicos ou diferentes equilíbrios bioquímicos nos neurónios.

A distinção fundamental entre as tecnologias de EEG e fMRI/PET reside nos tipos de resoluções que podem ser alcançados: no caso do EEG, conseguimos alcançar medidas de tempo de alta resolução na ordem de milissegundos, mas com baixa resolução espacial. Já com fMRI/PET, obtemos uma alta resolução espacial, mas uma baixa resolução temporal na ordem de segundos. Por causa desta distinção, muitos estudos actuais em neurociência tentam criar estudos mistos

utilizando tanto EEG como fMRI/PET para obter uma leitura mais rigorosa dos dados.<sup>73</sup>

Depois desta breve odisséia histórica sobre o cérebro e das reflexões filosóficas sobre a sua natureza, composição e relevância, passaremos, nesta segunda parte, a introduzir outras abordagens interessantes sobre a natureza do cérebro com mais quatro pensadores de renome internacional (Anil Seth, Karl Friston, Christof Koch e Joseph LeDoux) com que tive o privilégio de dialogar.

Começemos, assim, esta segunda parte – dedicada ao cérebro – a introduzir uma visão mais contemporânea deste importante órgão que argumenta que, ao contrário do que pensamos até então, o cérebro pode ser visto como uma “máquina” de previsão cuja função é criar hipóteses sobre o mundo.

---

<sup>73</sup> Baseado em: Gouveia, S. (2022) *Philosophy and Neuroscience: a Methodological Analysis*, New York: Palgrave Macmillan.



## II. O CÉREBRO PREDITIVO

O primeiro diálogo desta segunda parte dedicado ao cérebro conta com a participação do neurocientista que popularizou a ideia que considera o cérebro como uma máquina de previsão, tendo uma das *TEDx Talks* mais vistas do mundo com 10 milhões de visualizações e o sugestivo título “O seu cérebro alucina a sua realidade consciente”.<sup>74</sup>

Falamos de Anil Seth, Professor de Neurociência Cognitiva e Computacional na Universidade de Sussex em Inglaterra, onde também é codiretor do *Sackler Center for Consciousness Science*, que defende uma teoria que podemos chamar de “Processamento Preditivo” (PP).

Esta abordagem surge como crítica ou resposta a uma teoria que atribui ao cérebro um papel somente passivo: o de receber estímulos do exterior (da realidade!) e, a partir desses estímulos, construir um mapa ou um modelo mental dessa realidade, sendo que um papel activo do cérebro não seria considerado relevante.

---

<sup>74</sup> YouTube: <https://www.youtube.com/watch?v=lyu7v7nWzfo>.

Assim, esta nova teoria vem criticar essa postura passiva e defender que o cérebro é um mecanismo de previsão perfeitamente activo cuja função é prever os estímulos do mundo de forma a combiná-los entre si e formar o melhor “palpite” do que poderia ter causado esses mesmos estímulos.

Pense na sua percepção visual. Ao contrário do que se defendia até então, a percepção era vista como um mecanismo onde o leitor recebia estímulos visuais das palavras que está a ler deste livro. Somente depois de receber esses sinais é que o seu cérebro construía uma representação mental do livro.

Ora, o Processamento Preditivo defende que esta interpretação está equivocada: a percepção tem uma natureza construtiva que está numa dança constante com os sinais recebidos do mundo. Para esta teoria, a percepção é uma forma de “alucinação controlada” que tenta prever um modelo dos sinais visuais que o leitor irá ter e, somente depois, irá à realidade confirmar se esses sinais estão correctos ou se necessitam de ser atualizados por outros sinais.

Caso esta explicação seja um pouco confusa, peço ao leitor que olhe directamente para esta imagem:

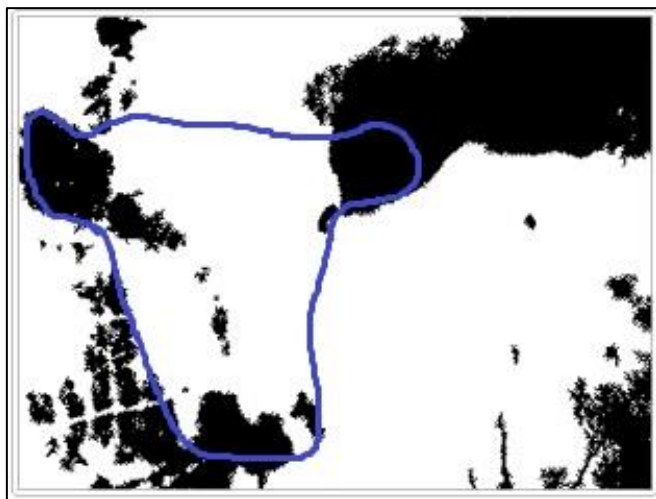


## 2. Ilustração do Processamento Preditivo (percepção visual)

|

Se o leitor nunca viu esta imagem, achará que faz pouco sentido sendo, no máximo, uma peça de gosto estético duvidoso que poderia ter sido produzida por um bebé pollockiano: não há nada que exista no mundo que possamos identificar com aquelas formas e sombras.

No entanto, apesar do seu cérebro não ser capaz de identificar algo, a abordagem defendida por Anil Seth torna-se mais plausível se olharmos agora para a segunda imagem que possui um mero detalhe diferente:



### 3. Ilustração do Processamento Preditivo (percepção visual)

II

Nesta segunda imagem, adicionando uma linha a contornar parte das formas e sombras que antes não tinham significado, o seu cérebro já é capaz de compreender o que está aí presente: a segunda imagem permite que o cérebro actualize o modelo anterior e crie um novo modelo onde consegue detectar a forma rudimentar do rosto de uma... vaca. O interessante desta “experiência” visual é que, se o leitor voltar a ver a primeira imagem, já será capaz de perceber o rosto da vaca sem o auxílio da linha, algo que antes o seu cérebro não foi capaz de ver.

Mas de que forma é que o Processamento Preditivo explica este fenómeno? No fundo, a ideia passa por

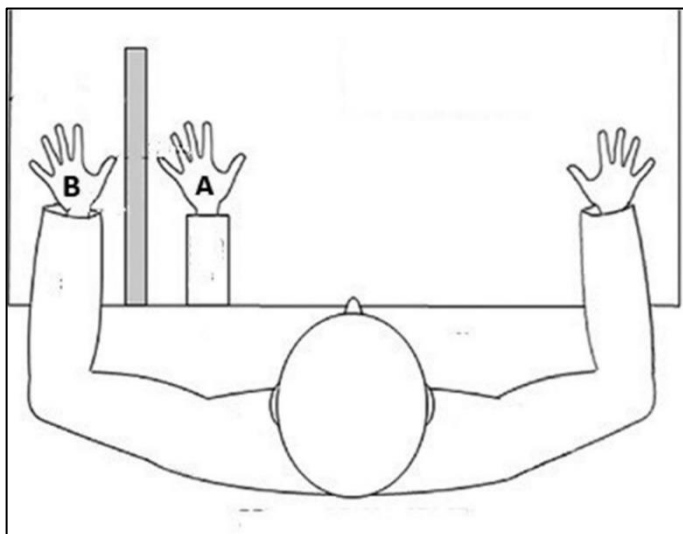
defender que, depois do seu cérebro ter percepcionado a segunda imagem, procedeu a uma actualização do modelo que antes não tinha significado, tendo colocado essa nova informação no modelo anterior, permitindo que algo que antes parecia não ter qualquer significado passasse a ter.

Repare que a informação sensorial que o seu cérebro possui tanto na primeira vez que olha para a primeira imagem, como quando olha para a mesma imagem a segunda vez (depois de ver a segunda imagem), é exactamente a mesma: o estímulo visual não mudou, manteve-se. O que mudou foi, portanto, o melhor “palpite” que o seu cérebro acha ser plausível baseado nas experiências anteriores.

Neste sentido, esta teoria defende que aquilo que vemos acontece tanto pela influência do mundo no cérebro, mas também, e talvez ainda mais, da influência do cérebro na previsão desse mundo.

É importante notar que o Processamento Preditivo pode ser aplicado à percepção e aos sentidos em geral, mas também a muitos outros aspectos do mental. No fundo, esta abordagem teórica tenta explicar a natureza de todos os aspectos mentais – desde a consciência, à memória ou até às doenças psiquiátricas – baseado no mesmo princípio-base. Veja-se outro exemplo interessante para mostrar o potencial deste *framework*, aplicado ao Eu (*self*).

A ilusão da mão de borracha é fenómeno estudado em psicologia focado na forma como o seu cérebro prevê, tal como a percepção, o seu Eu, isto é, aquilo que o seu cérebro considera ser parte do seu corpo. Esta experiência<sup>75</sup> tem a seguinte estrutura:



4. Ilustração do Processamento Preditivo (Eu);  
Mão A = de borracha/falsa; Mão B = verdadeira.

Imagine que o leitor é a pessoa sentada com uma mesa à frente pronta para participar nesta experiência. O que irá acontecer é o seguinte: a sua mão esquerda estará pousada na mesa, mas o leitor não poderá ter contacto visual com a mesma, dado que algo se encontra a tapar

---

<sup>75</sup> Um exemplo da experiência em vídeo pode ser encontrado no seguinte link: [https://www.youtube.com/watch?v=nzF\\_DfOafKw](https://www.youtube.com/watch?v=nzF_DfOafKw).

esse acesso (na imagem, estará na posição B). Um pouco mais à direita da sua mão esquerda estará uma mão de borracha acessível à sua percepção visual e que está posicionada no lugar onde a sua mão esquerda (real) deveria estar se não estivesse tapada (na imagem, na posição A).

O próximo passo da experiência é ter uma pessoa à sua frente que irá tocar, ao mesmo tempo, na sua mão real (que não consegue ver por estar tapada) e na mão de borracha (que consegue ver à sua frente). Finalmente, depois de alguns minutos de estimulação, algo de mágico parece ter acontecido: se Michael Myers aparecer de repente e espetar uma faca na mão de borracha que está a sua frente, o leitor irá, de modo automático e imediato, retirar a sua mão esquerda (real) da mesa, agindo como se tivesse sido essa sua mão que tivesse sido esfaqueada, em vez da mão de borracha.

O que aconteceu aqui? A ideia é argumentar que a informação sensorial do toque cruzada com a informação visual é suficiente para o seu cérebro criar um modelo do seu corpo onde a mão de borracha é assumida como sendo parte do mesmo. Na perspectiva desta teoria, até a experiência do que pertence ao nosso corpo é uma “alucinação controlada” pelo cérebro, é um palpite ou uma previsão.

Ora, tal como podemos ser enganados com ilusões visuais, a forma como experienciamos o nosso corpo

pode também levar a modelos equivocados: esta tese pode fornecer um contributo importante na compreensão de doenças psiquiátricas que lidam com perturbações do Eu como o transtorno bipolar ou a esquizofrenia.

Mas de que forma se dá esta previsão contínua que o nosso cérebro parece estar a fazer segundo esta teoria, tanto em relação aos sinais externos da realidade, como aos sinais internos do nosso corpo?

Um conceito fundamental para entender esta abordagem é o da “Inferência Bayesiana”, que permite explicar de que forma é que, por um lado, formamos os nossos modelos preditivos e, por outro, de que forma podemos actualizar-los constantemente e ajustar as nossas crenças a novas informações.<sup>76</sup>

Na prática, a inferência bayesiana<sup>77</sup> envolve combinar duas fontes de conhecimento: as previsões geradas internamente pelo cérebro, a que chamamos de “prior”, e as informações sensoriais provenientes do a realidade, a que chamamos de “verossimilhança”. Estas

---

<sup>76</sup> Baseado em: Gouveia, S. & Curado, M. (2020) (eds.) *The Philosophy and Science of Predictive Processing*, New York: Bloomsbury.

<sup>77</sup> Na sua forma mais formal, a probabilidade posterior  $P(H | E)$  é calculada pela conjunção da verossimilhança  $P(E | H)$  com a probabilidade anterior  $P(H)$  dividida pela verossimilhança marginal  $P(E)$ , onde E representa Evidência, ou seja, os novos dados do mundo, e H representa uma Hipótese, cuja probabilidade pode ser influenciada por E.

duas fontes são assimiladas para formar uma nova estimativa, a que chamamos de “posterior”, que representa o modelo actualizado do mundo. No fundo, o nosso cérebro procura equilibrar as suas expectativas internas com a evidência externa de modo a que possa adaptar-se constantemente à realidade.

Vejamos um exemplo simples que demonstra a utilidade deste método. Começamos pelo “prior”: antes de entrar em casa, o leitor assume que há uma probabilidade mínima (1%) de encontrar um gato, dado que sabe não ter nenhum gato. No entanto, ao abrir a porta, ouve um miar (uma observação, que se enquadra na “verossimilhança”) e, sabendo que os gatos miam, o leitor actualiza a sua crença de que poderá haver uma grande probabilidade (90%) de haver um gato em sua casa. Ao combinar esta crença inicial com a nova evidência, conclui-se o “posterior”: há uma grande probabilidade de haver um gato em sua casa.

Outros conceitos fundamentais centrais para compreender os fundamentos da teoria do Processamento Preditivo como uma teoria unificadora da mente humana, como a “Inferência Activa”, “Erro de Previsão” e “Princípio da Energia Livre”, serão introduzidos na próxima secção. Antes disso, vejamos também a utilidade desta abordagem para explicar a natureza de variadas doenças psiquiátricas.

Seguindo o que foi dito anteriormente, o cérebro é uma máquina de previsões que conjuga modelos internos com estímulos do mundo. Podemos argumentar que, na esquizofrenia, essa capacidade de realizar previsões de maneira mais eficaz fica comprometida por alguns motivos. Que motivos seriam esses?

Por exemplo, os indivíduos podem não conseguir distinguir as fontes internas e externas das informações, o que os leva a confundir determinados pensamentos internos com estímulos sensoriais externos.

Sabemos que um dos sintomas positivos mais presentes nesta doença são as alucinações auditivas, onde o sujeito pensa que as “vozes” geradas internamente têm uma origem externa no mundo – ou seja, que alguém está de facto a dizer-lhes algo.

Já indivíduos com autismo podem ter dificuldades na integração de informação sensorial externa, o que os pode levar a ficarem demasiado focados nos seus próprios modelos internos.

Dado que estes sujeitos possuem mais dificuldade em incorporar um novo estímulo sensorial, a sua capacidade de actualizar os modelos e, portanto, de adaptar-se de forma mais eficaz ao mundo, fica altamente comprometida.

É sabido que estes indivíduos preferem manter rotinas e padrões particulares e possuem grandes dificuldades

em relações sociais exactamente por terem uma capacidade reduzida de prever o comportamento dos outros.

Imagine agora um sujeito cujo cérebro tem uma tendência enviesada de forma negativa para antecipar os modelos do mundo: tal levaria essa pessoa a ter expectativas pessimistas constantes, distorcendo a sua capacidade de prever resultados positivos.

Tal pode estar baseado numa “rigidez” entre a relação do Eu com o mundo, fazendo com que o modelo relacional entre eles se mantenha negativamente rígido, isto é, não permitindo a entrada de novas informações positivas e perpetuando o viés negativo.

Nesse caso, estaríamos perante um paciente com.... Depressão, uma doença que afecta milhões de pessoas em todo o mundo. Sabemos, por exemplo, que pessoas depressivas têm uma propensão em olhar para si mesmas de forma negativa, processando apenas informações que confirmem uma baixa autoestima sobre elas mesmas.

E caso um cérebro dê uma importância preditiva excessiva a estímulos, em que está constantemente a gerar previsões de eventos futuros considerados ameaçadores, levando a uma antecipação exagerada que resulta em preocupações excessivas de perigo? Que problema do foro psiquiátrico teria um sujeito assim? Isso mesmo: ansiedade (no caso, crónica)!

Dado que existe uma interpretação distorcida de variadas ameaças no ambiente, o cérebro de um indivíduo ansioso terá dificuldade, por um lado, em actualizar os seus modelos internos de “segurança” no seu ambiente e, por outro, irá amplificar estímulos quotidianos como possíveis perigos, o que contribuirá para a sua sintomatologia de ansiedade.

Como podemos perceber através destes exemplos da Psiquiatria, a teoria do Processamento Preditivo tem um potencial explicativo enorme e poderá de facto contribuir para a descoberta de mecanismos específicos em diversas doenças mentais. Além disso, poderá também explicar como funciona a nossa percepção, o nosso Eu, a nossa memória, os sonhos ou até a consciência.

De seguida, e em continuação com esta abordagem, iremos introduzir algumas ideias que servem de base para esta teoria, desenvolvidas por um dos mais influentes neurocientistas do mundo, o Professor Karl Friston.

### III. O CÉREBRO LIVRE

O segundo diálogo conta com a participação do Professor Karl Friston, uma das maiores autoridades nas ciências do cérebro e um dos fundadores da perspectiva descrita na secção anterior. Depois de uma breve introdução do Processamento Preditivo, apresentarei de seguida alguns princípios que formam a base dessa abordagem ao cérebro e à mente humana, sendo que esta secção estará em diálogo directo com a antecedente.

Comecemos por introduzir o “Princípio da Energia Livre”. Este princípio tem uma longa história e foi desenvolvido por várias gerações de pensadores de variadas disciplinas, tendo sido formalizada mais recentemente na neurociência por Karl Friston, o cientista vivo mais citado do mundo. Na sua essência, sugere-se que o cérebro humano tem uma propensão intrínseca a minimizar o consumo de energia ao realizar as diversas tarefas cognitivas e mentais em geral.

A ideia é que o cérebro, de modo a economizar energia, procura organizar-se da forma mais eficiente possível, optimizando as tais previsões sobre o mundo que descrevemos anteriormente. Ora, tal significa que, segundo este princípio, o cérebro procura criar

modelos internos e precisos do mundo com o mínimo de “esforço” possível.

Em termos práticos, o Princípio da Energia Livre implica pensar o cérebro como estando, constantemente, a procurar um equilíbrio entre, por um lado, manter os modelos internos estabilizados e, por outro lado, ser flexível o suficiente para conseguir adaptar-se às novas informações do mundo, permitindo que tal aconteça de forma mais económica e eficaz possível, sendo uma abordagem que permite formular de que maneira os organismos lidam com a incerteza.

Em poucas palavras, podemos declarar que o Princípio da Energia Livre destaca a “economia” de energia como um princípio orientador na organização e funcionamento do cérebro humano com o objectivo de garantir a subsistência e sobrevivência do organismo ao longo do tempo.

Devemos notar que este princípio tem uma forma matemática complexa que não poderemos detalhar num livro introdutório desta natureza. Mas, na sua essência, trata-se de um princípio que até pode ser considerado “simples” de compreender. Imagine que o leitor está a conduzir numa estrada que lhe é familiar (por exemplo, do trabalho para casa, algo que faz todos os dias).

Nesse caso, o seu cérebro construiu, ao longo das várias viagens que fez, um modelo interno eficaz do

percurso que tem de percorrer, o que inclui a localização das curvas, dos semáforos, sinais de trânsito, e outros pontos de referência, numa tentativa de economizar a energia gasta para prever como será a estrada da próxima vez que a utilizar.

Esse modelo interno utiliza, portanto, informações que recolheu no passado e usa essas mesmas informações para prever o que poderá acontecer: por exemplo, se o leitor já souber que existe um semáforo ao fundo da rua, o seu cérebro irá antecipar esse facto e não irá ficar surpreendido se o sinal ficar vermelho. Isto permite a criação de um modelo “estável” da estrada, que é actualizado apenas se necessário, o que permite uma poupança de energia considerável.

No entanto, se por algum motivo, existir alguma alteração (e.g. um novo semáforo é instalado), o seu cérebro irá procurar adaptar-se e incorporar essa nova informação ao modelo já existente: em vez de reformular todo o modelo interno da estrada, actualiza o anterior contendo esse novo dado, estando constantemente a procurar diminuir os elementos de surpresa.

Apesar deste exemplo simples, é necessário compreender a força deste princípio: na sua essência, é ele que nos permite explicar de que forma é que um organismo se mantém vivo, sendo que tal é aplicado tanto a células simples como bactérias, como a

organismos altamente complexos como o ser humano e outros animais.

Tanto organismos simples como desenvolvidos são considerados sistemas dinâmicos na sua natureza, separados entre a sua existência e um meio ambiente onde essa sua existência acontece, o que implica uma interacção indirecta e troca de informações constantes. O propósito desta interacção é ir de encontro a um segundo conceito fundamental, o conceito de “Minimização do Erro de Previsão”.

O objectivo do que foi discutido até então é garantir que os modelos de previsão criados pelo cérebro minimizem a possibilidade de errar: um modelo preciso será mais próximo da realidade e, por isso, não exigirá um gasto energético tão dispendioso, quando comparado com um modelo menos preciso que precisará de ser actualizado.

Assim, todos os organismos tentam minimizar os erros das previsões feitas sobre o ambiente através do Princípio da Energia Livre. O que significa “energia livre” neste contexto? A energia livre pode ser considerada uma métrica relacionada com a surpresa ou incerteza nas suas previsões: quanto mais um organismo conseguir minimizar a sua energia livre, mais irá conseguir garantir a sua sobrevivência.

É importante notar que este princípio actua tanto na previsão dos estados externos do ambiente (como já

frisámos), mas também na previsão de estados internos do corpo: o organismo, além de estar a prever modelos de como será o ambiente em determinado momento, está, ao mesmo tempo, a prever as condições que os órgãos internos devem produzir para garantir a sobrevivência nesse ambiente.

Tal poderá implicar o seguinte: se uma previsão de determinadas condições externas (ambiente) ou internas (corporais) possuir erros, estaremos perante um cenário que resultará numa energia livre elevada, o que implicará estados de surpresa ou de imprevisibilidade elevada.

Se, no entanto, a previsão for precisa, o resultado será uma energia livre baixa. Portanto, todos os organismos – complexos ou simples, mamíferos ou aves, plantas ou bactérias – procuram minimizar essa incerteza ou surpresa.

É importante salientar outra característica deste princípio. Grande parte das teorias biológicas – as teorias que tentam explicar por que é que algum fenómeno biológico funciona de determinada forma – procuram responder a uma questão que esta abordagem tenta reverter: em vez de focar-se em descrever o que um organismo deve fazer para existir, procura-se compreender, assumindo que o organismo já existe, o que ele precisa de fazer para continuar a existir.

Tal é relevante para se compreender que o Princípio da Energia Livre em si mesmo não é uma teoria falsificável sobre como é que os organismos se comportam: apenas os seus postulados podem ser ou não falsificáveis, dependem da sua aplicabilidade.

Finalmente, para concluir esta breve introdução, resta-nos abordar outro conceito fundamental, o da “Inferência Activa”. Este conceito permite que o organismo, em casos em que o cérebro encontre algum problema nos modelos gerados em determinado contexto, possa procurar nova informação relevante para aquela situação em particular de modo a ajustar o modelo preditivo incorrecto ou impreciso.

Assim, a Inferência Activa pode ser considerada uma capacidade do cérebro de influenciar activamente a informação que vamos recolher ao ambiente, o que implica seleccionar determinados estímulos e dirigir a atenção para aqueles factores (que no caso, podem ser externos ou internos, novamente) que podem ser relevantes na redução da ambiguidade ou incerteza identificada.

Vejamos um exemplo relacionado com a previsão de estados internos do corpo, como os estados fisiológicos. Imagine que, de repente, é transportado da praia de Copacabana para os incríveis Alpes Suíços. O seu corpo, através da interocepção, identifica uma mudança de temperatura e faz produzir uma sensação de frio. Este sinal interoceptivo indica que a

temperatura do seu corpo irá diminuir se se mantiver naquele ambiente.

O que a Inferência Activa permite, neste cenário, é procurar informações sobre esse ambiente e também para outros sinais internos do corpo que poderão ter sido activados. Por exemplo, o seu cérebro poderá ter-se focado nos tremores que os seus músculos começariam a sentir, o que seria um indicador da temperatura real do seu corpo.

Baseando-se em todos estes elementos, o seu cérebro reformula o modelo de temperatura que tinha em Copacabana (onde, por exemplo estaria a suar, de modo a combater o calor sentido), para um modelo onde os outros elementos, como os tremores (que têm o propósito de aumentar o calor interno) passem a fazer parte do mesmo.

Portanto, é através da Inferência Activa que, neste exemplo, o cérebro optimizou activamente – focando a sua energia e atenção tanto nos sinais exteriores do ambiente como nos sinais internos do corpo – as previsões anteriores de modo a manter uma regulação da temperatura corporal adequada que, caso não tivesse acontecido, poderia levar a consequências... um pouco desagradáveis.

Alguns neurocientistas e filósofos criticaram esta abordagem<sup>78</sup> defendendo que nem tudo o que acontece no organismo parece acontecer com o objectivo de minimizar o consumo de energia, dado que há várias funções mentais, como a criatividade ou a imaginação, que parecem exigir um gasto bastante maior do que o normal.

A própria variação neuronal interindividual (entre pessoas) parece diferir, e essa diversidade parece ser difícil de explicar apelando à actuação do mesmo princípio. Mas o maior desafio talvez seja medir e quantificar de facto todo o mecanismo preditivo que acontece tanto no cérebro, como em cada célula.

Em seguida, abordaremos uma outra perspectiva de olhar para o cérebro e do seu papel no desenvolvimento de processos conscientes, apresentando uma teoria que pretende providenciar precisamente uma forma de quantificar a mente humana, desenvolvida pelo neurocientista e Presidente do *Allen Institute for the Brain* em Seattle, o Professor Christof Koch.

---

<sup>78</sup> Artigo do Professor Karl Friston em co-autoria: Friston, K. et al. (2023) “The free energy principle made simpler but not too simple”, *Physics Reports*, 1024: 1-29. A melhor introdução em língua portuguesa foi produzida por Maria Luzia Lennaco e colegas: Lennaco, M., Maia, T. e Sayeg, P. (2023) “Processamento Preditivo: uma Introdução à Proposta de Unificação da Cognição Humana”, *Principia*, 27 (3): 425-452.

## IV. O CÉREBRO INTEGRADO

Neste terceiro diálogo, iremos introduzir uma das teorias neurocientíficas sobre o cérebro e o seu papel no desenvolvimento da mente consciente mais em voga, desenvolvida pelo Professor Christof Koch, autor do livro *The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed*.

Esta teoria, apelidada de Teoria da Informação Integrada (TII),<sup>79</sup> tem vindo a revolucionar a forma como a consciência é pensada dentro da neurociência. Pelo seu carácter conceptual e teórico, tem impactado várias áreas adjacentes, como a Filosofia, a Física ou a Inteligência Artificial. Pela sua dimensão relevante, esta teoria tem também levado a considerações éticas interessantes.

Em primeiro lugar, é importante ressaltar que esta abordagem procura fornecer uma resposta científica e objectiva ao problema causado pela existência da consciência subjectiva no mundo físico, procurando

---

<sup>79</sup> O primeiro autor a desenvolver esta teoria foi Giulio Tononi em 2004, com quem Christof Koch, depois de ter trabalhado com o Prémio Nobel Francis Crick, uniu esforços para desenvolver mais profundamente esta teoria. Publicação original: Tononi, G. (2004) "An Information Integration Theory of Consciousness", *BMC Neuroscience*, 5 (1): 42.

identificar as propriedades essenciais – a que chamamos de “axiomas” – e inferir as condições necessárias e suficientes que qualquer substrato deve satisfazer para ser consciente – apelidados de “postulados” –, tudo isto expresso através de uma notação matemática específica que, pela sua complexidade, não poderemos abordar neste livro.<sup>80</sup>

A estratégia da TII procura reverter uma abordagem típica que podemos encontrar em neurociência na investigação da consciência: em vez de tentar identificar no cérebro os processos neuronais responsáveis pelos estados conscientes, Koch prefere começar pela fenomenologia (pela consciência!) e perguntar, depois, que tipo de mecanismos físicos poderiam explicar essa fenomenologia da experiência.

Para esta abordagem, a definição de experiência consciente está muito próxima daquela oferecida por Descartes, que considerava a consciência como uma propriedade fundamental do mundo. Partindo deste pressuposto, a TII avança com os seguintes 5 axiomas:

- Axioma da Existência (AExis): este axioma é, no fundo, uma reformulação do cogito cartesiano, substituindo o acto de “pensar” com o acto de

---

<sup>80</sup> Caso o leitor tenha curiosidade, poderá encontrar uma descrição matemática da TII no seguinte artigo (disponível gratuitamente online): Albantakis L. et al. (2023) “Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms”, *PLOS Computational Biology*, 19 (10): e1011465.

“existir”: “experiençio, logo existo”; tal significa que possuímos uma certeza indubitável e imediata da nossa própria perspectiva de primeira pessoa no mundo; a consciência é, seguindo este axioma, intrinsecamente real;

- Axioma da Informação (AInfo): este axioma indica que uma experiência tem de especificar sempre algo, sendo sempre distinta de outras experiências possíveis; assim, podemos afirmar que toda a experiência nos informa de forma particular através de um contraste com outras experiências particulares;
- Axioma da Integração (AInteg): este axioma defende que a consciência é unificada, e que o conteúdo da experiência é irreduzível a partes independentes, sendo antes integrada num todo; não possuímos, portanto, experiências isoladas umas das outras que somente depois formam um conjunto mental; antes, quando o leitor experiencia a leitura destas palavras, experiencia o tipo de letra e a cor numa experiência unificada, e não separadas entre si;
- Axioma da Composição (AComp): este axioma advoga que todas as experiências possuem uma estrutura, constituída de vários aspectos e várias combinações entre si; novamente, se o leitor estiver a experienciar estas palavras, tal contém diferentes aspectos fenomenológicos como as cores e as formas das letras;

- Axioma da Exclusão (AExcl): este último axioma indica que a experiência consciente exclui sempre outras experiências, dado que, quando especifica uma experiência (cf. AInfo), terá necessariamente de excluir outras experiências; por outras palavras, as experiências conscientes possuem fronteiras; além disso, possuem um “grão” temporal: ou seja, os conteúdos da experiência possuem um tempo de duração específico.

Vejamos de que forma estes axiomas podem ser aplicados a um exemplo do dia-a-dia, como o de ler um livro, por ordem dos axiomas apresentados.

Imagine que se encontra em sua casa, no sofá, no seu precioso tempo livre, a desfolhar com entusiasmo este livro, o que indica a presença de experiências conscientes intrínsecas (AExs). Durante essa leitura, a informação do livro é processada de forma única na sua consciência: os vários padrões e conhecimentos contidos nas palavras demonstram o carácter informativo dessa experiência (AInfo).

Ademais, essas informações presentes no livro formam uma experiência consciente unificada e integrada de maneira coesa entre as palavras, as imagens, e as emoções que poderá sentir ao ler o livro (AInteg). Claro, essa experiência consciente de leitura é composta de elementos particulares, como os vários debates e as introduções do autor: todos estes elementos

contribuem para uma composição única da sua experiência de leitura.

Finalmente, durante o tempo em que o leitor está a ler este livro, muitas outras informações poderão estar a ocorrer em simultâneo – como um jogo de futebol do seu clube favorito a dar na televisão. No entanto, caso esteja a considerar o conteúdo deste livro interessante, a sua experiência consciente poderá até excluir todas as outras experiências simultâneas, demonstrando a natureza única da consciência num determinado momento (AExcl).

Além destes axiomas – que são tomados como verdades auto-evidentes – a TII postula, baseando-se em cada um desses axiomas, que tipo de propriedades um sistema físico deve possuir para ser consciente. Vejamos, então, os postulados da TII:

- Postulado da Existência: este postulado afirma que a existência da consciência implica um sistema de mecanismos com poder de causa e efeito; para um substrato físico existir, uma condição necessária é possuir poder causal; por exemplo, o neurónio é um mecanismo de estados binários, dado que pode ser descrito como estando a “disparar” ou “desligado”;
- Postulado da Informação: este postulado declara que, se a consciência é de facto informativa, ela deve ter a capacidade de especificar ou diferenciar determinadas experiências de outras;

tal implica que qualquer mecanismo dentro de um sistema físico deve possuir poderes de causa-efeito; todos estes “reportórios” formam a estrutura causa-efeito que permite ao sistema especificar um determinado estado;

- Postulado da Integração: este postulado indica que, para que a integração da consciência aconteça num todo unificado, o sistema físico tem de ser necessariamente irreduzível: as partes desse sistema têm de ser interdependentes; todos os elementos mecânicos do sistema devem ser capazes de causar algo no todo, assim como serem afectados pelo mesmo; se determinado sistema físico pode ser dividido em partes sem que tal afecte a sua estrutura causa-efeito, então esse sistema perde a sua integração e não poderá ser consciente;
- Postulado da Composição: este postulado implica assumir-se que os elementos (mecanismos) de um sistema físico devem ter capacidade de se combinar entre si, garantindo que essas combinações possuem poder de causa-efeito; ou seja, se uma experiência tem uma estrutura determinada, o processo causal que gera essa experiência tem necessariamente de ser estruturado;
- Postulado da Exclusão: finalmente, este postulado envolve a ideia de que um estado consciente de um sistema físico tem de ser finito:

diversos mecanismos de um sistema podem ter variados poderes de causa-efeito; ora, para haver consciência num sistema, apenas um desses mecanismos poderá ter a estrutura causa-efeito maximamente irreduzível, o que representa o nível mais elevado de integração em termos de informação.

Vejamos um outro exemplo para compreender a relevância do que foi agora descrito. Imagine que o leitor acorda de repente do seu sono de beleza e vê, à sua frente, a parede azul do seu quarto. Além disso, vê também onde se encontra a sua cama, e percebe que, nesse momento, existe de forma consciente de si mesmo e do que se encontra ao seu redor.

Ora, o que a TII afirma é que essa experiência existe para si mesmo, e não para outra coisa qualquer; é uma experiência específica (e não uma genérica); é uma experiência unitária: o seu olho direito não está a experienciar algo diferente do seu olho esquerdo (ambos experienciam a mesma experiência); além disso, essa experiência é definida especificamente pelo que está a observar à sua frente e não o que está atrás de si ou ao seu lado; finalmente, essa experiência é estruturada por vários elementos que se relacionam entre si (e.g. o seu corpo, na sua cama, que está situada de forma a que possa ver a parede azul à sua frente) que lhe permite ter precisamente essa experiência e não outra.

Reparou no que acabámos de fazer? Isso mesmo: por meio deste exemplo, aplicámos todos os postulados da TII.

É importante ressaltar que, para os teóricos da TII, estes axiomas (e por consequência, os postulados) fazem parte de uma lista completa e finita: não existe mais nenhuma propriedade da experiência consciente que deva ser considerada essencial na sua existência. Se um sistema físico – qualquer um, biológico, artificial ou até alienígena – quiser ser consciente, terá de reproduzir, através dos postulados, todos os axiomas agora apresentados.

Um outro aspecto atraente desta teoria é fornecer uma tentativa de quantificar, numa medida particular, a capacidade de um sistema integrar informação. Esta quantificação é apresentada através da medição  $\Phi$  – a letra grega *Phi* – que representa a habilidade (entre zero e máxima) de um sistema poder integrar a informação processada num sistema particular.

Segundo a TII, a consciência “acontece” quando um sistema alcança um elevado nível de informação integrada: quanto mais intrínseca e irreduzível for essa informação, maior o seu  $\Phi$  e, portanto, maior o grau de consciência. O cálculo dá-se por intermédio da quantidade de informação que está globalmente integrada nesse sistema comparada com a soma de informação de cada uma das suas partes.

Assim, sempre que um sistema alcança um  $\Phi$  máximo, esse sistema será considerado consciente: abaixo desse valor máximo, defende-se que o sistema não terá informação integrada suficiente para alcançar um estado consciente.

Esta abordagem quantificável – em conjunto com os axiomas e os postulados da TII – fornecem uma teoria incrivelmente elegante e, além de tudo, com capacidade para testar os seus pressupostos. Vejamos algumas previsões interessantes, baseadas em dados empíricos reais, do poder explicativo desta teoria.

O cerebelo é uma região importante do sistema nervoso responsável pela função motora localizada na parte de trás da nossa cabeça. Apesar deste elemento pesar somente 150 gramas e constituir cerca de 10% do volume total do cérebro, possui 4 vezes mais neurónios que todo o resto do cérebro somado, estando altamente conectado com este.<sup>81</sup>

Ora, apesar disto, se este órgão for removido ao leitor, não terá de se preocupar demasiado: a sua consciência irá manter-se totalmente intacta! Como é que tal pode ser o caso? Segundo a TII, como o cerebelo é composto de módulos independentes entre si, a informação aí processada não é tão integrada como acontece em

---

<sup>81</sup> Publicação original: Lemon, R. & Edgley, S. (2010) "Life without a cerebellum", *Brain*, 133: 652–654.

outras zonas do cérebro e, por tanto, o seu  $\Phi$  é inferior ao mínimo necessário para haver consciência.

Lembra-se de algumas ideias que avançámos na primeira parte sobre o sono? A TII também prevê algo interessante sobre o que acontece ao nosso cérebro e à nossa consciência quando vamos dormir. Assumindo então que a consciência pode ser medida em  $\Phi$ , é intuitivo pressupor que, quando dormimos sem qualquer experiência consciente – sem entrarmos em sonhos lúcidos, por exemplo –, o nosso cérebro deverá ter uma diminuição acentuada de conectividade neuronal e, por consequência, estará menos integrado entre si.

Ora, estudos empíricos<sup>82</sup> em cérebros humanos durante o sono mostram que esta previsão é efectivamente precisa: o cérebro diminui a sua conectividade cortical.

Uma outra previsão intrigante desta teoria está relacionada com um tipo de procedimento neurocirúrgico conhecido por “calosotomia” (em inglês, *split-brain*).<sup>83</sup> Este procedimento faz algo curioso: divide

---

<sup>82</sup> Publicação original: Massimini, M., ..., Tononi, G. (2005) “Breakdown of cortical effective connectivity during sleep”, *Science*, 309: 2228–2232.

<sup>83</sup> Publicação original: Van Wagenen W. e Herren, R. (1940) “Surgical division of commissural pathways in the corpus callosum Relation to spread of an epileptic attack”, *Archives of Neurology and Psychiatry*, 44:740-9.

o cérebro em dois hemisférios desconectados entre si através de uma incisão precisa no corpo caloso.

Numa famosa experiência científica em pacientes com epilepsia<sup>84</sup>, depois da remoção do corpo caloso – para diminuir a sintomatologia –, algo previsto pela TII acontece.

Se um desses pacientes visualizar uma imagem apenas com o olho direito, processando a informação visual pelo hemisfério esquerdo do cérebro, e lhe perguntarmos o que via naquele momento, o mesmo dir-nos-á que não é capaz de se lembrar de ter visto alguma imagem.

No entanto, se pedirmos ao mesmo paciente para desenhar o que viu, este já será capaz de desenhar correctamente o que tinha observado. Como é que tal pode acontecer? Estaremos perante uma magia demoníaca?

Certamente que não: a razão para tal ocorrência deve-se ao facto de diferentes funções estarem localizadas em diferentes partes do cérebro. Enquanto que a linguagem é uma função cognitiva presente principalmente no hemisfério direito (que não processou qualquer informação que estava a ser assimilada no hemisfério esquerdo, dado que a conexão entre os dois estava interrompida), a função

---

<sup>84</sup> Publicação original: Gazzaniga, M. (1967) "The split brain in man", *Scientific American*, 217 (2): 24-29.

relacionada com desenhar está presente no hemisfério esquerdo.

Ora, de acordo com a TII, quando o corpo caloso foi desconectado, o cérebro deixou de ser, na sua totalidade, um sistema completo, criando-se dois sistemas independentes em cada hemisfério que poderão maximizar o seu  $\Phi$  e, assim, criar dois fluxos de consciência independentes um do outro.

Outro exemplo ligado ao mau funcionamento do cérebro está nas lesões cerebrais: esta teoria prevê que qualquer tipo de lesão que afecte a capacidade do cérebro integrar informação irá impedir a formação de estados conscientes.

Estudos recentes<sup>85</sup> mostram que esta previsão é exactamente fiel à realidade: através de estimulação magnética transcraniana, vários pacientes com lesões cerebrais tinham alterações substanciais no seu cérebro em relação a pacientes saudáveis.

Será que esta teoria, além de providenciar uma narrativa explicativa da forma como o cérebro biológico é consciente, pode também trazer alguma luz em relação ao desenvolvimento de sistemas artificiais conscientes? A resposta é mista: por um lado sim, e por outro, não.

---

<sup>85</sup> Publicação original: Casali, A. et al. (2013) "A theoretically based index of consciousness independent of sensory processing and behavior", *Science Translational Medicine*, 5: 198ra105–198ra105.

Do lado positivo, em teoria, a TII não especifica que um sistema tenha de ser biológico para ser consciente: apenas indica que tem de ser um sistema que possa integrar informação seguindo os axiomas avançados anteriormente.

No entanto, do lado negativo, para os defensores desta teoria, é claro que a Inteligência Artificial actual, baseada em estruturas de alimentação directa (*feed-forward*) em que a informação só é processada numa direcção (do *input* para um *output*), jamais poderá ser consciente, dado que um sistema deste tipo não terá qualquer  $\Phi$ , mesmo que seja um sistema que replique o comportamento consciente na perfeição: tal seria somente uma simulação. Antes, seria necessário que o sistema tivesse uma estrutura de reentrada (*re-entrant*) em que os *outputs* podem ser usados como *inputs* e vice-versa.

Para encerrar estas notas introdutórias à TII, faltará abordar uma crítica – que pode ser vista como uma vantagem – levantada a esta teoria. Para alguns autores, a TII pode ter consequências pampsiquistas contra-intuitivas. E o que é o pampsiquismo? O pampsiquismo é uma teoria desenvolvida por vários pensadores ao longo da história – de Espinoza a Leibiniz ou Bertrand Russell<sup>86</sup> e, mais recentemente,

---

<sup>86</sup> Publicação original: Russell, B. (1927) *The Analysis of Matter*, London: Kegan Paul, Trench, Trubner & Co.

Philip Goff<sup>87</sup> – que defende que tudo o que existe no mundo físico possui propriedades mentais fundamentais.

Ora, tal significa afirmar que a consciência é uma propriedade fundamental, o que vai de encontro à ideia avançada pela TII de que alguns sistemas complexos, quando atingem elevados níveis de informação integrada, passam a ser conscientes.

Além disso, tanto o pansiquismo com a TII concordam que a consciência não é um fenómeno de tudo ou nada: pelo contrário, a consciência deve ser vista como um espectro, entre o absolutamente inconsciente e o completamente consciente.

Apesar destas similitudes, parece algo radical identificar a TII como uma teoria pansiquista. Embora ambas tenham, de facto, compromissos cartesianos ao aceitar a consciência subjectiva como fundamental na sua natureza, o pansiquismo advoga que a consciência é fundamental para todas as entidades existentes, enquanto que a TII apenas sugere as condições necessárias para que a consciência possa emergir em determinados sistemas complexos.<sup>88</sup>

---

<sup>87</sup> Publicação original: Goff, P. (2019) *Galileo's Error: Foundations for a New Science of Consciousness*, London: Pantheon Book.

<sup>88</sup> Baseado em: Tononi, G. (2012) "Integrated information theory of consciousness: An updated account", *Archives italiennes de biologie*, 150: 56-90; Tononi, G. & Koch C. (2015) "Consciousness: Here,

Para finalizar estas notas introdutórias, abordaremos uma outra perspectiva de olhar para o cérebro através das emoções e do papel das mesmas no desenvolvimento da nossa mente, apresentando algumas das teses centrais desenvolvidas pelo neurocientista e Director do *Emotional Brain Institute* na Universidade de Nova Iorque, o Professor Joseph LeDoux.

---

There and Everywhere?", *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370 (1668): 20140167.



## V. O CÉRBRO EMOCIONAL

Neste quarto e último diálogo desta odisseia ao mundo do cérebro consciente, apresentaremos alguns dos conceitos-chave desenvolvidos pelo neurocientista e autor do livro *The Emotional Brain*, o Professor Joseph LeDoux. O primeiro passo para entendermos o que é este cérebro emocional é compreender quais são as estruturas anatómicas que compõem este sistema vital para a sobrevivência de cada um de nós.

Na neurociência, a parte do cérebro que processa as diversas emoções é apelidada de “sistema límbico”, que é composto por diversas partes, estando localizado na superfície medial do cérebro dos mamíferos, assumindo um papel basilar no funcionamento das nossas emoções e, por consequência, da forma como nos relacionamos conosco e com os outros.

Através do sistema nervoso autônomo, influencia valorativamente (de forma positiva ou negativa) o funcionamento visceral e a regulação metabólica do organismo, comandando os comportamentos que são considerados essenciais à sobrevivência de todos os mamíferos.

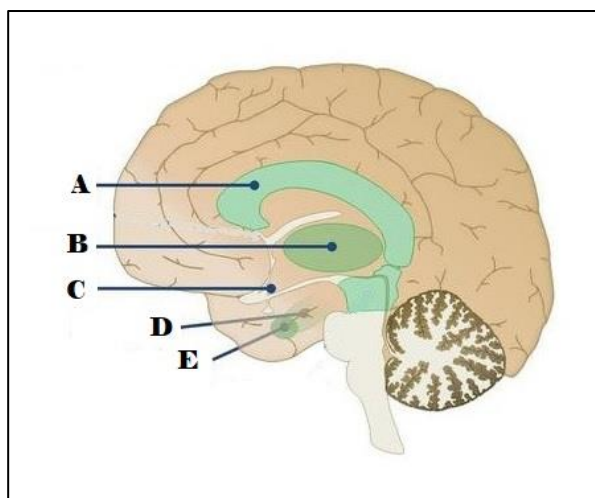
A resposta emocional apropriada será então produzida através da integração da informação sensorial num

estado mental particular, atribuindo-se um conteúdo afectivo aos estímulos que são registados e associados a memórias já existentes.

Vejamos algumas das principais estruturas anatómicas que pertencem ao sistema límbico:

- Amígdala: localizado no lobo temporal anterior, encontra-se conectada ao hipotálamo e é considerado o “centro” que nos orienta em situações de perigo, produzindo estados de alerta ou sensações de medo; a sua remoção em mamíferos causa uma postura sexualmente indiscriminada, indiferença a situações de risco e perda de sentido afectivo de estímulos exteriores; já a sua estimulação eléctrica gera comportamentos agressivos e violentos;
- Hipotálamo: localizado no centro do cérebro, encontra-se logo abaixo do tálamo e acima da hipófise, sendo considerado a parte mais fundamental do sistema límbico, dado ser responsável pelas funções vegetativas do cérebro (manutenção do equilíbrio interno, temperatura corporal, fome, sede, produção hormonal, etc.); além disso, está relacionado directamente com o comportamento, desempenhando um papel essencial nas emoções (a parte intermédia está ligada à aversão e as partes laterais ao prazer e raiva);

- **Tálamo:** localizado no diencefalo, logo abaixo do corpo caloso, é nesta parte do sistema límbico que as informações sensoriais captadas do ambiente são seleccionadas, sendo também responsável pela parte motora (activação muscular); está também relacionado com os estados de alerta e é considerado fundamental na percepção da dor, sendo importante na regulação dos estados de sono e vigília;
- **Hipocampo:** localizado na parte interna do lobo temporal medial, é responsável pela memória de curto e longo prazo, permitindo ao organismo comparar situações actuais com experiências passadas, aumentando assim a sua capacidade de sobrevivência; é também responsável pela orientação e navegação espacial.



5. Sistema Límbico: A = Corpo caloso; B = Tálamo; C = Hipotálamo; D = Hipocampo; E = Amígdala.

Partindo desta breve descrição neuroanatômica, poderemos compreender melhor as várias teses avançadas por Joseph LeDoux. Este neurocientista defende uma nova proposta conceptual que redefine alguns conceitos-chave no estudo das emoções, especialmente das funções e circuitos relevantes para a sobrevivência dos organismos.

Nesta nova abordagem, procura-se reverter o foco da forma tradicional de investigar as emoções: em vez de questionarmos se as emoções sentidas conscientemente pelos seres humanos estão também presentes em diversos animais – numa atitude clara de antropomorfismo –, tenta-se colocar a questão de compreender até que ponto as funções e circuitos relevantes para a sobrevivência e preservação dos animais estão, ou não, também presentes em nós, seres humanos.

É importante compreender que, para LeDoux, estes circuitos e funções não estão relacionados causalmente com as emoções, apesar de contribuírem indirectamente para estas. O que são, então, estes circuitos de sobrevivência? No fundo, este conceito amplo procura integrar todo o tipo de mecanismos relevantes para a preservação e sucesso do organismo no seu dia-a-dia.

E porquê introduzir este novo conceito e não falar de “emoções” directamente? LeDoux propõe aqui um argumento filosófico muito interessante para suportar

esta mudança: para este, o conceito de “emoção” é filosoficamente inconsistente, não existindo qualquer consenso na comunidade científica sobre o seu significado real: daí sugerir esta inversão metodológica, de modo a não ficar “preso” a esta indefinição conceptual.

Seguindo esta inovação filosófica interessante, LeDoux irá procurar focar-se nos fenómenos que são relevantes no estudo das emoções, evitando, todavia, a utilização de conceitos emocionais particulares. Em vez disso, o foco será nos circuitos específicos que instanciam determinadas funções que permitem ao organismo sobreviver através, por exemplo, da detecção e resposta a desafios e oportunidades, e serão estas instanciações particulares que poderão ser rotuladas com alguma emoção no fim (mas não vice-versa).

Esses circuitos passam por desenvolver capacidades de identificação do perigo, de identificações de possíveis parceiros de acasalamento, de presença de comida ou de água, de defesa e manutenção energética, de termorregulação, entre muitos outros. No fundo, o desafio do cientista será descrever todos estes processos evitando cair no uso de uma linguagem confusa e ambígua como é a linguagem dos conceitos ligados às emoções.

Repare que estes circuitos de sobrevivência estão conservados em todas as espécies de mamíferos (pelo

menos, e talvez em muitos outros animais e organismos): embora possam existir algumas diferenças, há componentes essenciais dessas funções que são partilhados por todos os animais.

Vemos, assim, de que forma a inversão *ledouxiana* é útil: como essas funções de sobrevivência estão instanciadas em circuitos conservados pela história evolutiva, podemos evitar antropomorfizar esta ao questionarmos “que emoções humanas estão presentes nos outros animais” e perguntar, em vez disso, “que circuitos presentes nos outros animais estão presentes nos seres humanos”.

A ideia desta inversão é pensar as emoções e conceitos adjacentes (como a motivação, reforço, inibição, excitação, etc.) como componentes de um processo maior, que não se esgota num sentimento específico.

Portanto, o que se segue desta abordagem não é uma tentativa de explicar ou definir as emoções: pelo contrário, o objectivo é disponibilizar uma estrutura para pensar alguns fenómenos associados às emoções – fenómenos relacionados com a sobrevivência – de uma forma que não seja confundida com a procura do significado demasiado abrangente de “emoção” e permitindo o foco em aspectos-chave que evitem debates infinitos sobre uma correcta definição de “tristeza”, “medo”, “felicidade”, entre todos os outros conceitos que usamos para descrever os nossos sentimentos (*feelings*).

Este foco na relação das funções de sobrevivência conservadas pela evolução com as emoções não é totalmente novidade, seguindo, pelo contrário, uma tradição que remonta pelo menos a Darwin (1872): não será inovação para os neurocientistas que estes circuitos estejam de facto ligados ao cérebro.

A estratégia – e inovação – de LeDoux é focar-se precisamente nestes circuitos de modo a substancializar as diferentes emoções, em vez da estratégia contrária tipicamente utilizada até então, de partir de definições *a priori* de emoções e só depois tentar identificar os circuitos relacionados com as mesmas. Para simplificar: o objectivo é identificar-se primeiro o circuito, e só depois a emoção, em vez de começar pela emoção para identificar o circuito.

Uma segunda problemática interessante na investigação das emoções é compreender se existem circuitos emocionais no cérebro que são inatos em vez de criações da mente humana ou dos aspectos sociais e culturais. A complexidade deste debate prende-se na questão de perceber se as chamadas “emoções básicas” são ou não naturais.

As emoções básicas são aquelas que, supostamente, são expressas por todas as pessoas que existem, estando também presentes nos animais mais próximos, em termos evolutivos, dos seres humanos. O exemplo canónico de emoções básicas é a lista de

Ekman (1972, 1992)<sup>89</sup> que inclui a felicidade, o medo, a tristeza, a raiva, a surpresa e o nojo, sendo que muitas investigações da neurociência actual procuram identificar a base neuronal dessas emoções.

Apesar destas múltiplas investigações neurocientíficas, o conceito de “emoções básicas” tem levantado vários desafios, dado que existem diversas teorias que incluem outras emoções ou excluem algumas da lista de Ekman. Claro, tal deve-se, provavelmente, à dificuldade em rotular um conceito emocional com uma diversidade de estados biológicos que podem ou não coincidir com esse conceito.

Além disso, há mais duas dificuldades centrais nesta abordagem das “emoções básicas”: (i) por um lado, há quem defenda que as emoções são construções psicológicas e sociais que surgem na interacção entre o ambiente físico ou social, sendo que, por isso, não existe uma determinação biológica à partida; (ii) a grande maioria das teorias ligadas às emoções básicas é baseada na investigação em cérebro de animais que

---

<sup>89</sup> Publicação original: Ekman, P. (1972) “Universals and cultural differences in facial expression of emotion” In J. Cole (Ed.), *Nebraska Symposium on Motivation*, Lincoln, Nebraska: University of Nebraska Press: pp. 207–283. Uma actualização interessante pode ser encontrada em: Ekman, Paul (1992). “An Argument for Basic Emotions”, *Cognition and Emotion*, 6 (3/4): 169–200.

não encaixam de todo na lista de Ekman ou outras emoções básicas (Panksepp, 1998, 2005).<sup>90</sup>

Por exemplo, Louise Barrett (2006)<sup>91</sup> considera não fazer sentido argumentar que as emoções básicas foram conservadas pela evolução através de circuitos neuronais por três razões particulares:

- alguns estudos imagéticos mostram que áreas activadas pelo mesmo estímulo foram associados a diferentes emoções básicas;
- grande parte das evidências a favor das emoções básicas foram recolhidas em cérebros de animais (e não de humanos) através de técnicas retrógradas (e.g. estimulação elétrica do cérebro) com uma falta de precisão acentuada;
- as emoções básicas identificadas nos estudos em seres humanos não coincidem com as emoções básicas identificadas nos animais.

Ora, esta controvérsia pode ser aclarada, novamente, pela abordagem de LeDoux: em vez de se falar em “emoções básicas”, os cientistas e filósofos deveriam optar por “circuitos de sobrevivência”, o que lhes

---

<sup>90</sup> Publicação original: Panksepp, J. (1998) *Affective Neuroscience*, New York: Oxford U. Press. Uma actualização relevante poder ser encontrada em: Panksepp J. (2005) “Affective consciousness: Core emotional feelings in animals and humans”, *Consciousness and Cognition*, 14: 30–80.

<sup>91</sup> Publicação original: Barrett, L. (2006) “Are Emotions Natural Kinds?”, *Perspectives on Psychological Science*, 1:28–58.

permitiria resolver estas e outras confusões conceptuais – que têm implicações directas na investigação empírica – causadas pelo uso da linguagem das emoções em primeira instância.

Estes circuitos de sobrevivência incluem circuitos ligados à manutenção nutricional, de fluídos, de defesa, de reprodução, de termorregulação e são circuitos que já estariam presentes nas primeiras formas de vida na Terra: por exemplo, as bactérias conseguem fugir de determinados ambientes tóxicos e procurar ambientes favoráveis ao seu desenvolvimento utilizando alguns destes circuitos.

Com o desenvolvimento dos organismos eucarióticos multicelulares, estes circuitos de sobrevivência foram ficando cada vez mais complexos, especialmente com o desenvolvimento de receptores sensoriais e especializados, e de um sistema nervoso central capaz de coordenar todas as funções corporais e diversas interacções com o ambiente.<sup>92</sup>

Deve ficar claro, no entanto, que o objectivo de LeDoux, ao descrever os diversos circuitos de sobrevivência, não é ligá-los directamente às “emoções básicas”: antes, o objectivo é libertar a linguagem neurocientífica de conceitos baseados em sentimentos subjectivos do ser

---

<sup>92</sup> Publicação original: Shepherd, G. (1988) *Neurobiology II*, New York: Oxford.

humano e focar, de forma neutra, nos tais circuitos conservados durante a história evolutiva.

Para concluir, vejamos um breve exemplo em relação à agressão para compreendermos a relevância desta abordagem. Sabemos que a agressão é uma emoção que não possui uma única “representação” fixa no cérebro: antes, existem diferentes tipos de agressão dependendo do contexto em que esta ocorre.

Por exemplo, a agressão pode ser desencadeada pelo circuito de defesa (na protecção do organismo); pode ser desencadeada pelo circuito alimentar (na caça por alimento); ou pode ser desencadeada pelo circuito reprodutivo (na competição por parceiros).

Note-se, portanto, e contra muita investigação em neurociência, que um circuito de sobrevivência não está conectado a uma emoção particular (muito menos a uma activação neuronal específica do cérebro): pelo contrário, diferentes circuitos de sobrevivência podem estar relacionados com uma e a mesma emoção e vice-versa.

Tal poderá significar que estes circuitos não têm uma relação directa nem um papel causal com as emoções. Claro, deve haver certamente uma influência indirecta: mas o foco destes circuitos é somente um, garantir propósitos adaptativos de forma a garantir a preservação do organismo nas diferentes

complexidades exigidas pela interação com o ambiente.

Portanto, fica agora claro de que forma a inversão proposta por LeDoux é útil: se olharmos para os circuitos de sobrevivência partilhados pelos animais e pelos seres humanos, e somente depois fizermos considerações sobre as chamadas “emoções”, estaremos mais perto de garantir explicações relevantes e perceber a natureza e o papel dessas emoções no ser humano.<sup>93</sup>

Concluídas estas segundas notas introdutórias, o leitor poderá usufruir dos próximos diálogos com mais quatro pensadores de renome internacional sobre a natureza do cérebro através de múltiplas perspectivas que irão certamente enriquecer o seu conhecimento sobre esta odisseia fascinante que ocorre dentro da sua cabeça.

---

<sup>93</sup> Baseado em: LeDoux, J. (1996) *The Emotional Brain*, New York: Simon and Schuster; LeDoux, J. (2003) “The Emotional Brain, Fear and the Amygdala”, *Cellular and Molecular Neurobiology*, 23 (4/5): 727-738; LeDoux, J. (2012) “Rethinking the emotional brain”, *Neuron*, 73 (4): 653-676.

# DIÁLOGOS II

## Cérebro

# DIÁLOGOS II

## Cérebro

## VI. Diálogo com Anil Seth



Anil Seth é Professor de Neurociência Cognitiva e Computacional na University of Sussex, onde também é Co-Director do Sackler Center for Consciousness Science. É também co-director do Programa do Instituto Canadense de Pesquisa Avançada (CIFAR) sobre Cérebro, Mente e Consciência e do Programa de Bolsas de Doutorado Leverhulme: "Da Sensação e Percepção à Consciência".

É o Editor-chefe da *Neuroscience of Consciousness* (Oxford University Press). Foi o Presidente da Conferência para o 16º Encontro da Associação para o Estudo Científico da Consciência (ASSC16, 2012).

A sua investigação foi financiada pelo EPSRC (Leadership Fellowship), o Conselho Europeu de Pesquisa (ERC, Advanced Investigator Grant), o Wellcome Trust e o Instituto Canadense de Pesquisa Avançada (CIFAR).

Publicou vários livros, sendo o mais recente *Being You: a New Science of Consciousness* (Faber & Faber), que foi considerado um dos 7 principais livros de ciência pelo *The Guardian* em 2021.

Mais informações: <https://www.anilseth.com/>

**Pergunta:** O Professor Seth é um dos principais neurocientistas que defende a Teoria do Processamento Preditivo, uma ideia com uma longa história e com pressupostos interessantes sobre o cérebro e a mente humana. Quando é que o Professor Seth se viu confrontado pela primeira vez com a ideia de que o cérebro é uma máquina de previsões (*prediction machine*)?

**Anil Seth:** A primeira vez que ouvi falar nessa ideia pareceu-me logo uma maneira realmente atraente de pensar não somente sobre como o cérebro funciona em geral, mas também nos temas que eu estava particularmente interessado na época – e ainda estou –, como por exemplo, como pensamos sobre os correlatos neuronais da consciência, ou como pensamos em relacionar o que acontece no cérebro com o que acontece nas nossas experiências de consciência, e também como entendemos as emoções no Eu.

Pensar estas ideias de codificação preditiva, das quais ouvi falar pela primeira vez provavelmente há cerca de vinte anos – e que depois comecei a trabalhar nelas há cerca de doze ou treze anos – parecia o tipo de linguagem pertinente a ser usada na investigação.

O outro aspecto que realmente me atraiu foi que a codificação preditiva é algo sobre o qual se pode abordar em vários níveis ou disciplinas: no nível muito

matemático, no nível computacional (pode-se escrever código preditivo) mas também no nível conceptual.

Além disso, podemos também abordá-la no nível filosófico, o que faz com que seja uma excelente forma de unir essas várias disciplinas e acho que essa característica multidisciplinar é muitíssimo relevante na neurociência actual.

**Pergunta:** Pode contar-nos, sucintamente, uma breve História sobre essa visão? Quem são os primeiros ou os principais defensores e proponentes da ideia de que o cérebro é um agente activo no mundo?

**Anil Seth:** Platão (risos)! Claro, depende de quão longe podemos ir atrás na história. Sabemos que, já na filosofia grega, podemos encontrar a ideia de que o que consideramos real, o que percebemos nos conteúdos da nossa experiência, não é necessariamente o que existe realmente.

É uma espécie de reflexão que Platão tem na Alegoria da Caverna, onde ele fala dos prisioneiros acorrentados à parede da caverna que tomam as sombras da caverna como o mundo real. Eles não conhecem e não podem conhecer nenhuma diferença além da realidade e da ilusão da caverna. Mas, claro, não há aqui nenhuma ideia explícita de codificação preditiva como um princípio da função cerebral.

Há também o conceito hindu de *Maya*, que enfatiza o processo pelo qual aquilo que percebemos tem o carácter de parecer real – tal está também intimamente relacionado com os conceitos modernos de processamento preditivo.

Podemos encontrar, no século XV, Ibn al-Haytham, um estudioso árabe que foi, provavelmente, a primeira pessoa a pensar sobre a percepção como inferência, esta ideia de que o que percebemos é um julgamento sobre o que está a acontecer no mundo, algo que o organismo estima ou prevê.

Outra grande influência é o cientista alemão do século XIX Herman Von Helmholtz, que abordou esta perspectiva com muito mais detalhes, apresentando a primeira formulação da ideia de que a percepção é um processo de inferência baseado no cérebro que é em si inconsciente. A formulação de von Helmholtz alinhou-se naturalmente com a matemática da inferência proveniente do raciocínio bayesiano.

Mas a codificação preditiva em si, como ideia e algoritmo específico, surgiu da engenharia e do processamento de sinais, onde os cientistas tentavam descobrir formas de comprimir sinais compridos sem perder informação. A ideia é: se se pode prever algo, então não precisamos de codificá-lo explicitamente.

Seguindo esta ideia, o algoritmo de codificação preditiva foi desenvolvido na década de 1950, no

âmbito da engenharia. Por volta de 1990, foi publicado o primeiro artigo, ou a referência clássica, focada no sistema visual: Rajesh Roh e Dana Balor defenderam a tese de que o sistema visual pode estar a incorporar um tipo de inferência preditiva, e usaram essa perspectiva para explicar as propriedades do sistema visual.

Desde então, o processamento preditivo tem sido considerado uma teoria muito mais geral de como o cérebro funciona e é usada por muitos pensadores e cientistas para elucidar variados aspectos do ser humano.

**Pergunta:** O Professor Seth tem uma posição crítica sobre a famosa distinção introduzida por David Chalmers entre os problemas fáceis (*easy*) versus difíceis (*hard*) da consciência. No fundo, não aceita essa divisão e acha que é mais interessante falarmos do problema *real* da consciência. Poderia esclarecer um pouco mais sobre essa estratégia de mudança conceptual para abordar a consciência de um ponto de vista científico e objectivo?

**Anil Seth:** Chamo-lhe de problema *real* da consciência para me “meter” com David Chalmers, dado que ele contribui tanto e tão bem para a filosofia da mente e considero que articulou muito bem essa distinção entre problemas difíceis e fáceis.

No entanto, acho que tal distinção tornou o problema da consciência um pouco obstrutivo. Se aceitarmos essa divisão, estamos a aceitar a intuição de que nenhuma explicação do cérebro ou de qualquer sistema físico poderia explicar que partes do universo são conscientes e que outras partes não o são.

É verdade que a consciência não parece ser o tipo de coisa que poderia ser explicada em termos físicos, o que é uma intuição realmente profunda. E, de uma perspectiva, esta intuição é precisamente aquilo que o problema difícil da consciência formaliza.

Mas se aceitarmos essa dificuldade, surge um efeito colateral: acabamos por tratar a consciência como um grande mistério e tentamos encontrar uma solução “mágica”, uma fonte especial – talvez algum tipo de mistério quântica – que esperamos poder produzir a magia da experiência consciente a partir de um mero mecanismo.

E talvez seja mesmo necessário algo mágico, no sentido de precisarmos de uma nova física ou de novo momento *eureka*. Porém, penso que uma abordagem mais interessante é focarmo-nos no problema *real* da consciência. Novamente, essa não é uma ideia nova, é apenas uma forma de dar uma ênfase maior dentro da linguagem que Chalmers tornou tão popular.

O problema real da consciência é a ideia de afirmar: Ok, podemos não saber de que forma os sistemas físicos

dão origem à consciência ou são idênticos a ela, mas sabemos que a consciência existe e que está intimamente relacionada, de várias formas, ao cérebro. A consciência é um facto do universo em que vivemos e que precisa de explicação.

A abordagem do problema real argumenta que não devemos abordar o problema “de frente” – como as soluções para o problema difícil tentam fazer – mas podemos dividir um pouco o problema em partes, e tentar entender quais são as propriedades da consciência e como podemos explicar essas propriedades.

A analogia que uso no meu recente livro *Being You* (2022) é a ideia de que os cientistas – não há muito tempo atrás – pensavam que a vida não poderia ser explicada em termos de mecanismo. Por causa dessa ideia, os pensadores da altura passaram muito tempo a discutir se havia uma centelha de vida, ou um Elan Vital ou outra coisa qualquer.

Mas no final, os cientistas não encontraram essa centelha mágica de vida, nem concordaram que a vida não existia. O que aconteceu foi que, simplesmente, perceberam que a vida não é algo único ou unitário, mas é composta de muitas propriedades diferentes, e que podemos explicar cada uma dessas partes através da química e da física.

Seguindo esta estratégia, o problema *difícil* da vida não foi resolvido, mas foi antes dissolvido: é esse tipo de abordagem que acho que pode ser muito útil para o estudo da natureza da consciência, mesmo que, eventualmente, não tenha sucesso em dissolver por completo o problema difícil. Mas só saberemos se funciona ou não se tentarmos, e esta abordagem real procura fazer exactamente isso: pelo menos tentar.

**Pergunta:** Portanto, essa abordagem ou mudança conceptual pode, no fim de contas, até criar novos problemas que não nos poderiam ter surgido inicialmente: quando conseguimos dissolver essa distinção – um tanto problemática – entre o problema fácil e difícil da consciência, podemos criar uma nova linha de investigação que pode causar algum progresso. Porque, antes disso, se afirmarmos que a consciência é *difícil* no sentido de Chalmers, então não podemos fazer muito em termos de tentar criar investigações empíricas sobre a natureza da consciência.

**Anil Seth:** Exactamente. Acho que esse é um ponto muitíssimo relevante, porque entende realmente por que é que essa mudança conceptual é pertinente. Acho que é um aspecto muito subestimado da interacção entre ciência e filosofia. Não é que tenhamos esse “menu” fixo de questões e apenas possamos ou não

resolvê-las uma por uma – embora seja assim que muitas vezes é apresentado ao público.

São tantos os artigos que vemos hoje em dia, em jornais e até em revistas científicas, como “10 maiores mistérios não resolvidos”. Ora, acho que tal abordagem negligencia totalmente a dialética entre teorias, dados e perguntas: é possível desenvolver algum progresso no conhecimento do mundo tanto pela forma como as perguntas que fazemos mudam, tal como pela forma como as nossas respostas se desenvolvem.

**Pergunta:** E podemos até descobrir que a questão ou o principal problema que estávamos a tentar resolver em primeiro instância não faz, afinal, nenhum sentido, certo?

**Anil Seth:** Certo. Voltando à analogia anterior, se alguém, hoje em dia, tentar ganhar uma bolsa de investigação para procurar o *Elan Vital*, não irá de certeza conseguir essa bolsa. É uma questão que não faz qualquer sentido. O mesmo aconteceu na física, com o antigo conceito de *éter*.

O que tal implica? Penso que isso implica que o aparato conceptual que herdamos de problemas ou mistérios específicos deve ser sempre questionado, uma vez que está sempre à nossa disposição, o que nos pode levar a utilizá-lo acriticamente e não ver os problemas que essa

aceitação neutra podem levantar para a nossa própria investigação.

**Pergunta:** Voltando à teoria do processamento preditivo, o Professor Seth acha que esta abordagem preditiva ao cérebro e à mente pode lançar alguma luz sobre o desenvolvimento de uma inteligência artificial consciente? Ouvimos tanto falar em “consciência artificial” actualmente, mas parece não fazer muito sentido procurar desenvolver uma consciência “artificial” sem primeiro compreender de que forma a consciência “natural” existe.

**Anil Seth:** Para começar, acho que já existe uma grande contribuição do processamento preditivo, como considerado na ciência cognitiva, na Inteligência Artificial. Como mencionei antes, foi na engenharia que surgiu o processamento preditivo como método de compressão de sinal.

Além disso, podemos encontrar objectos interessante em IA como as máquinas de Helmholtz, em homenagem ao mesmo Helmholtz que originou o conceito de percepção como inferência de inconsciência, e codificadores automáticos, e até algoritmos que implementam efectivamente aspectos do que chamaríamos de teoria do processamento preditivo nas ciências cognitivas – as principais computações que constituem essa teoria.

Esse tipo de sistemas já diferem um pouco das redes de aprendizagem profunda (*deep learning*) que são tão predominantes na investigação e nas aplicações de Inteligência Artificial actualmente.

Os principais ingredientes para o processamento preditivo e IA são os seguintes: precisamos de modelos generativos – algo que seja capaz de gerar previsões sobre os seus sinais de entrada (*inputs*). De seguida, precisamos de um modelo que compare esses sinais de entradas com essas previsões e actualize as previsões anteriores e, numa escala a longo prazo, o próprio modelo. A partir daqui, poderemos criar formas de percepção e de aprendizagem com o mesmo algoritmo.

Na verdade, um dos meus colegas – um talentoso ex-estudante de doutoramento fantástico chamado Beren Millidge – mostrou que, se implementarmos codificação preditiva, podemos realmente obter retropropagação (*back propagation*), que é o tipo de algoritmo de aprendizagem mais usado em IA.

Muitos investigadores criticaram a retropropagação porque afirmavam que não seria neurobiologicamente plausível: não seria o tipo de coisa que os cérebros fazem dado que requer sinais de erro globais que são propagados por toda a rede.

No entanto, o que descobrimos foi que a codificação preditiva devolve a propagação gratuitamente, de uma forma que é (potencialmente) biologicamente plausível.

Tal representa mais uma evidência forte de que os nossos cérebros realmente funcionam assim.

No presente, muita investigação em IA está a seguir esse caminho e a construir não somente redes cada vez mais profundas, mas redes que possuem modelos generativos: vemos este tipo de tecnologia em toda a parte, tais como as redes adversárias generativas (*adversarial networks*), algumas das quais são úteis para a neurociência, sendo que outras podem ser menos úteis. Seja como for, acho que há um território realmente fértil nessa intersecção.

**Pergunta:** O Professor Seth publicou o seu novo livro, *Being You: A New Science of Consciousness*, que foi muito bem recebido, tendo sido considerado pelo *Financial Times* e o *The Guardian* o livro do ano na categoria de ciência. Que ideia principal do livro gostaria de destacar?

**Anil Seth:** Acho que há três “mensagens” principais que tentei passar no meu livro. A primeira é que a consciência pode ser abordada científica e filosoficamente sem abordar o problema difícil de frente, ou seja, podemos lidar com este problema real com uma abordagem neurofenomenológica, e aqui devo dar os créditos a Francisco Varela, que foi o primeiro a propor esta abordagem para estudar a consciência.

A segunda mensagem é a ideia de que podemos usar a codificação preditiva para tentar explicar não apenas o que os organismos fazem – como eles recebem informações sensoriais para orientar o comportamento – mas também para mapear o conteúdo das previsões perceptivas. Nesta perspectiva, diferentes tipos de experiência visual podem ser explicados por envolverem diferentes tipos de previsão perceptiva.

A terceira e última mensagem é que todo esse mecanismo preditivo também se aplica ao Eu (*self*), e é esse o maior foco do livro. O Eu não é a “coisa” que tem percepções, nem temos um pequeno “mini-Eu” dentro da nossa cabeça que processa todas as percepções e decide o que fazer.

O que experienciamos como o Eu é, também, uma espécie de inferência perceptiva. Nesse caso, os sinais sensoriais vêm, em grande parte, de dentro do corpo, da frequência cardíaca ou do estômago, o que podemos chamar de interocepção.

Esta é a grande tese do meu livro: tudo o que aparece na consciência é algum tipo de previsão perceptiva, e todas as experiências perceptivas que temos estão, em última análise, fundamentadas e baseadas no corpo e na nossa natureza como máquinas vivas: há uma relação de conexão muito íntima entre o corpo vivido, o Eu, e a consciência em geral.

**Pergunta:** Esta abordagem pode moldar ou ser aplicada a muitas investigações em transtornos psiquiátricos? Podemos presumir que muitos dos distúrbios mentais de cérebros “doentes” podem ser baseados numa espécie de falta de equilíbrio entre o *input* previsto e o *input* real que deveria ser processado de alguma forma pelo modelo inferido?

**Anil Seth:** Espero sinceramente que tal seja o caso: é certamente algo que estamos a tentar abordar no meu centro de investigação na Universidade de Sussex ao longo dos últimos dez anos. Acho que, com todos estes temas ligados à doença mental e distúrbios psiquiátricos, existem muitas preocupações, mas também muitas promessas excessivas.

Considero que há muito a ser feito, dado que no tratamento de saúde mental e na medicina psiquiátrica, muitas das abordagens ainda se baseiam, primariamente, no tratamento dos sintomas, em vez de abordar as causas mecanicistas subjacentes.

Expandindo um pouco mais sobre o tema, penso que a maioria das abordagens actuais aborda os sintomas de transtornos mentais da mesma forma que o leitor toma um paracetamol para aliviar os sintomas de uma febre causada por uma doença subjacente.

Felizmente, para muitas doenças, temos já tratamentos que abordam estas causas subjacentes – como os antibióticos, por exemplo. Mas é verdade que, na

psiquiatria, em muitos casos, não temos realmente um equivalente: não temos “antibióticos psiquiátricos” que visem os mecanismos que dão origem aos sintomas.

Estou convencido de que tentar compreender as perturbações psiquiátricas através das lentes do processamento preditivo pode oferecer-nos um caminho melhor para identificar mecanismos plausíveis.

Mesmo assim, haverá ainda um longo caminho a percorrer antes de termos em mãos tratamentos eficazes. Existem sinais promissores por aí, mas não existe uma panaceia que resolva tudo relacionado aos transtornos mentais. Nem mesmo psicodélicos...

**Pergunta:** Para concluir, gostaria de saber a sua opinião sobre o debate actual sobre o livre arbítrio. Existem pensadores como Robert Sapolsky que defendem que não existe nenhum livre arbítrio, defendendo um determinismo radical. Mas outros filósofos como Dan Dennet dirão que temos graus de liberdade. Qual é a sua posição nesse debate?

**Anil Seth:** Tenho um capítulo no livro exactamente sobre esse tópico que, curiosamente, é o único tópico sobre o qual nunca havia escrito nada até então.

O conceito de livre arbítrio começou a fazer mais sentido para mim quando comecei a pensar nele da

mesma forma que penso sobre qualquer outro tipo de experiência: como uma espécie de previsão perceptual. Agora, o que isso significa em termos de como o debate é normalmente formulado? Há pessoas que – como Sapolsky – defendem fortemente a tese de que o livre arbítrio não existe – e normalmente afirmam isto por causa de alguma suspeita de que o universo é, de alguma forma, determinista, ou seja, que o seu curso está totalmente definido à partida. Este é certamente o caso de Sapolsky.

Parece-me, todavia, que argumentos como este caem muitas vezes na falácia do espantalho, argumentando contra uma versão do livre arbítrio que, de qualquer forma, não vale a pena manter. Esta versão indesejável do livre arbítrio é o que os filósofos chamariam de “livre arbítrio libertário”.

Sou muito céptico deste tipo de livre arbítrio e chamo-lhe “livre arbítrio fantasmagórico” – que é basicamente a ideia de que a consciência pode activar-se e alterar o curso dos eventos físicos, uma espécie de causa não-causada estrategicamente inteligente. Este tipo de livre arbítrio não faz sentido, mas não acho que seja o tipo de livre arbítrio que precisamos ou que deveríamos defender.

O principal problema com as discussões sobre o livre arbítrio é, a meu ver, o facto das pessoas tenderem a fazer suposições que não percebem que estão a ser pressupostas. Para mim, a pergunta que deveríamos

realmente fazer é a seguinte: é o cérebro um sistema suficientemente complexo para reagir a situações de maneiras muito flexíveis, que não são imediatamente determinadas pelo ambiente? E a resposta a essa pergunta, para organismos como nós, é: sim!

Podemos, de facto, realizar acções que têm as suas causas que vêm mais de dentro do corpo – e que remontam profundamente às nossas histórias – do que do mundo, da situação imediata do exterior, no ambiente.

Temos muitos graus de liberdade, novamente, num sentido de engenharia muito preciso. Posso fazer muitas coisas, mas faço apenas uma dessas coisas. E o nosso cérebro – altamente complexo – controla estes graus de liberdade, integrando muitas causas anteriores de formas muito sofisticadas e subtis num único “caminho final comum” comportamental.

Paralelamente, é um facto que há algumas ocasiões – quando realizo algumas acções – em que sinto uma experiência de volição e de agência. Às vezes, experiencio uma acção como sendo voluntária e outras vezes não. Essencialmente, o que faço é adoptar a mesma perspectiva de outros tipos de experiência perceptiva e perguntar: “por que é que essa experiência parece ser assim?” ou “qual é o objectivo dessa experiência?”.

Se seguirmos essa linha de pensamento, pode-se entender as experiências de livre arbítrio como sendo equivalentes às experiências de ver cores. Elas não reflectem a realidade como ela realmente é mas, ainda assim, são muito úteis para o organismo. A cor vermelha não existe objectivamente, como uma propriedade do mundo, mas é algo muito útil para experienciarmos subjectivamente, pois rastreia uma propriedade útil de como os objetos no mundo se comportam.

Na minha visão, as experiências de livre arbítrio rastreiam uma propriedade muito útil de como nós, organismos complexos, nos comportamos. Estas experiências rotulam acções que têm uma certa “liberdade do imediatismo”, para usar o termo de Mike Shadlen. E isto é útil – não porque as experiências de livre arbítrio realmente causem acções – mas porque permitem que o organismo aprenda para que, da próxima vez, se as coisas não funcionarem, uma acção diferente possa ser tomada.

O erro, do meu ponto de vista, é pensar que a experiência do livre arbítrio provoca uma acção: não é isso que acontece. Novamente, a experiência do livre arbítrio é útil porque rotula acções específicas como tendo as suas causas mais internas do que externas.

Se encararmos as coisas desta forma, todo o debate sobre o determinismo e a estocasticidade torna-se largamente irrelevante. Podemos pensar no livre

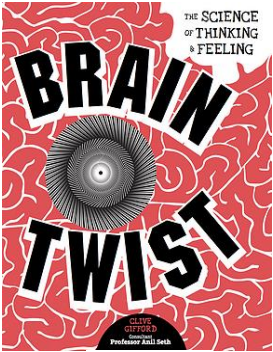
arbítrio como um fenómeno biológico natural, mas ele é real no mesmo sentido que experienciar as cores é real, e é útil no mesmo sentido em que a experiência das cores é útil.

Pensar no livre arbítrio desta forma mais subtil, como algo que vem em graus, significa que podemos falar sobre os graus em que ele varia, uma discussão que se torna claramente mais complicada quando falamos sobre responsabilidade moral e assim por diante, mas essa complexidade acompanha este tema fascinante e difícil de ser pensado, tal como a consciência.

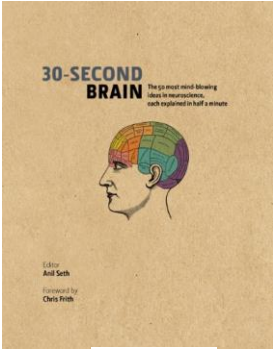
Livros de Anil Seth



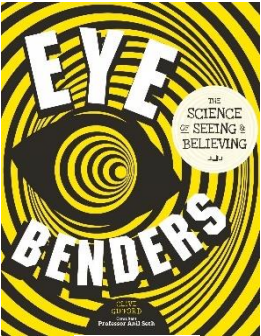
2021



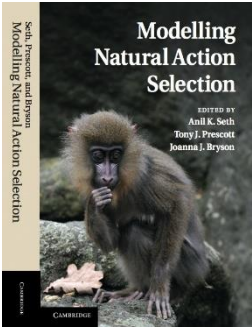
2015



2014



2013



2011

## VII. Diálogo com Karl Friston



Karl J. Friston é Professor no Instituto de Neurologia na University College London, sendo um neurocientista e uma autoridade em imagética cerebral, tendo inventado o mapeamento paramétrico estatístico, a morfometria baseada em voxel e a modelagem causal dinâmica (DCM).

Foi premiado com o “Minerva Golden Brain Award” e foi eleito “Fellow da Royal Society”. Recebeu uma “Medal College de France” (2008) e o prémio “Weldon Memorial” (2013) por contribuições à biologia matemática.

Em 2016, recebeu o prémio “Charles Branch” por descobertas incomparáveis em Investigação do Cérebro e o prémio “Glass Brain”. Friston possui nomeações honorárias da Universidade de Zurique e da Universidade Radboud.

Publicou vários livros, como o *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (2006), *Principles of Brain Dynamics: Global State Interactions* (2012) e mais recentemente *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior* (2022).

Mais informações: <https://www.fil.ion.ucl.ac.uk/~karl/>

**Pergunta:** O Professor Friston é considerado a maior autoridade do mundo no chamado Princípio da Energia Livre (PEL) e nas suas múltiplas consequências teóricas e empíricas. A minha primeira questão é, na verdade, uma metaquestão: como surgiu esta ideia do Princípio da Energia Livre na sua investigação?

**Karl Friston:** Podemos defender que a ideia do Princípio da Energia Livre (*Free Energy Principle*) existe desde os tempos de Platão, estando também presente no pensamento kantiano, e tendo sido muito bem articulado por Helmholtz, com a noção de inferências inconscientes.

Podemos também encontrá-lo nas ideias de Richard Gregory, no pressuposto da percepção como teste de hipóteses e formalizações, ou por pessoas como Peter Dine e Jeffrey Hinton e a Máquina Helmholtz, onde estes autores foram inspirados pela noção matemática de energia livre de Richard Feynman, que estava também a tentar usar o princípio para criar um problema de integração de inferência insolúvel num problema de optimização.

Tal significa que podemos considerar o cérebro como uma máquina de optimização, que tenta optimizar as suas crenças constantemente. Podemos então interpretar o cérebro como uma máquina de inferência, um órgão estatístico.

Estas ideias têm sido desenvolvidas há literalmente milénios, mas foram formalizadas pelo trabalho de pessoas como Richard Feynman, e em termos de inteligência artificial e aprendizagem automática, por pessoas como Jeffrey Hinton e Peter Dine.

As generalizações dessas ideias envolvem algumas etapas. A primeira generalização seria aplicar exactamente as mesmas ideias não apenas para fazer inferências, medições e observações, mas também para considerar que o nosso comportamento e as nossas acções podem ser entendidas como uma tentativa de minimizar a energia livre.

Se usarmos por exemplo *machine learning* como uma espécie de energia livre – que é uma energia livre negativa –, se eu estivesse a conversar com físicos, estes estariam a falar sobre minimizar a energia livre ou minimizar o erro de previsão, minimizar a surpresa.

Seguindo isto, podemos pensar agora em tudo o que fazemos como estando a tentar minimizar a surpresa, no sentido de que prevemos que nos vamos mover utilizando, por exemplo, uma teoria ideomotora para realizar essas previsões. Tudo isto pode ser entendido como minimizar exactamente a mesma energia livre ou erro de previsão ou surpresa.

Portanto, com estas e outras influências, fui formalizando aos poucos estas ideias, cristalizando-as ao longo de um longo período de tempo até à

actualidade, tentando desenvolver uma abordagem relevante para explicar o funcionamento do cérebro humano.

**Pergunta:** Tal como afirma o Princípio da Energia Livre, os organismos vivos visam minimizar a entropia da sua energia livre. Seguindo essa tese, acha que poderia haver um problema quando consideramos, por exemplo, estados alterados de consciência, uma vez que, geralmente, estes envolvem entropia anormalmente alta ou anormalmente baixa? Acha que estes casos podem ser problemáticos à luz da sua perspectiva? Estou a pensar, por exemplo, nos estados induzidos por drogas ou outros tipos de estados induzidos semelhantes.

**Karl Friston:** Não acredito que tal seja problemático. Penso que são janelas muito importantes e muito informativas para esta visão do cérebro enquanto tenta dar sentido a este mundo e fá-lo avançar no sentido da sensação activa.

Penso que, tanto a psicopatologia como os diferentes estados mentais que podemos ter quando consumimos determinadas drogas psicadélicas, ambas revelam aspectos mecanicistas muito importantes desta perspectiva.

O Steven falou em entropia. A entropia é apenas uma descrição de uma distribuição de probabilidade, e por

isso, é muito importante descobrir qual distribuição de probabilidade se está a referir nesse caso.

O Princípio da Energia Livre, quando aplicado a qualquer tipo de auto-organização – desde pequenas partículas até seres humanos –, a entropia de que as pessoas falam e que está a ser focada é a entropia dos resultados: são as coisas que eu realmente experiencio como organismo.

Essa seria a perspectiva da homeostase: o organismo está apenas a tentar manter a sua exposição ao meio ambiente na sua troca com o mundo vívido dentro de limites viáveis.

Nesse caso, o organismo está a tentar minimizar essa entropia, mas, por outro lado, para fazer isso, tem também de maximizar a entropia das suas crenças de acordo com coisas como o princípio de Occam: queremos encontrar essas explicações muito simples, sem nos comprometer muito com uma explicação particular.

Podemos ver que existe uma dialética entre tentar fazer o melhor tipo de inferência, o que exigiria maximizar a entropia, e a consequência disso, que é minimizar a entropia física dos meus resultados.

O mesmo argumento também se aplica à entropia da nossa actividade neuronal que codifica as nossas crenças sobre as causas das nossas sensações. Então, ao mesmo tempo, o cérebro está a maximizar a

entropia das crenças, mas estará também a minimizar, em grande parte, a entropia da dinâmica neuronal, de modo que sejam minimamente complexas, tentando assim torná-las as mais eficientes possíveis.

Acho que devemos ter um certo cuidado ao falar de entropia, pois existem diferentes tipos de entropias, como a entropia da actividade geral, a entropia das nossas crenças, e a entropia das observações. Mas todos estes diferentes tipos de entropias são formas muito úteis, como o Steven destacou, de resumir diferentes estados de consciência.

A história em relação a condições psiquiátricas e/ou drogas que alteram a mente e, na verdade, a diferentes estados alterados de consciência que são fisiológicos, como estados de sono, por exemplo, ou até sonhos lúcidos – todos estes estados focam-se em aspectos muito interessantes desse processo de optimização da energia livre.

Às vezes, com a sua codificação preditiva, se o organismo fizer algumas suposições simplificadoras, poderá ler-se a energia livre como um erro de previsão, mais precisamente, um erro de previsão ponderado com precisão. Nesse caso, não se está a tentar minimizar todos os erros de previsão, mas apenas aqueles que são considerados muito precisos, informativos e confiáveis.

Quando pensamos no cérebro como se estivesse a ser movido pelo imperativo de minimizar erros precisos de previsão ponderada, percebemos que não basta fazer previsões sobre o conteúdo das observações, mas também sobre a precisão das mesmas.

É necessário também quantificar a sua incerteza, e o que isso significa, do ponto de vista de um fisiologista, é que tem de se estimar a excitabilidade de várias estruturas neuronais ao relatar erros de previsão para níveis mais elevados que são responsáveis por acumular a evidência e a actualização das crenças.

Dito de forma muito simples, o que tal significa é que a modulação da excitabilidade das estruturas neuronais que reportam erros de predição, reportando a energia livre, tem ela própria de ser prevista. E se o cérebro errar nessas predições, obterá algumas crenças e inferências muito estranhas e anormais.

Tal é importante por duas razões centrais: em primeiro lugar, liga-o aos aspectos psicológicos da ponderação da decisão, que seria a atenção. Dar atenção a algo é basicamente permitir que se privilegie aqueles que preveem erros, tendo acesso privilegiado à actualização de certas crenças, mas não de outras, algo como: “vou ignorar isso, não é novidade, dado que tem baixa precisão; são notícias falsas; mas isto já parece ser interessante: estou a estimar que esta seja uma informação importante ou precisa; vou aumentar o

ganho de atenção desses erros de previsão e fazer a minha actualização.”

Do ponto de vista farmacológico, os medicamentos que apresentam esses efeitos são exactamente os mesmos que afectam a excitabilidade dessa ponderação de precisão. Então, podemos agora fazer uma ligação interessante entre o papel dos psicadélicos como a psilocibina e/ou o LSD e os relatos de atenção visual sobre o porquê de obtermos inferências perceptivas anormais sob efeito desses psicadélicos.

Mas o interessante é que não precisamos de tomar psicadélicos para ter essas inferências anormais. Podemos observar precisamente as mesmas falhas de neuromodulação como fenótipos de certas condições psiquiátricas que se prestam a falsas inferências.

Com esta “inferência”, quero apenas dizer erros clássicos do “tipo um” e do “tipo dois”: inferir algo que está lá quando não está, como uma alucinação, ou inferir algo que não está lá, quando está, o que seria uma síndrome dissociativa ou síndrome de negligência. Ambos são características de muitas condições neurológicas e psiquiátricas e que podem ser explicadas pela minha abordagem.

**Pergunta:** A minha próxima questão é exactamente sobre como podemos aplicar essa abordagem aos transtornos mentais, como a esquizofrenia ou o

autismo. Acho que o poder das suas ideias é exactamente o facto de podermos explicar tanto o cérebro “normal” ou saudável, mas também o cérebro anormal, os estados alterados de consciência e até os sonhos lúcidos. Acha que essa profusão explicativa é um sinal que está no caminho certo?

**Karl Friston:** Sim, o caso específico dos sonhos lúcidos é muito interessante.

**Pergunta:** Penso que o Professor Friston tem um artigo publicado onde aborda os sonhos lúcidos através da sua perspectiva de inferência activa e energia livre, de 2018,<sup>94</sup> se não me engano?

**Karl Friston:** Sobre sonhos, posso mencionar brevemente o meu colega e amigo Allan Hobson, com quem estudei há alguns anos e que adorava o fenómeno dos sonhos lúcidos, mas que também estava focado em salientar que, se quisermos evidências de que o cérebro é um órgão construtivo no sentido que gera fantasias ou hipóteses sobre como o mundo funciona, então não precisamos ir muito mais além do que focarmo-nos nos sonhos.

Durante o sonho temos esse desligar neuroquímico/farmacológico/neuromodulador da

---

<sup>94</sup> Publicação original: Hobson, A. & Friston, K. (2014) “Consciousness, Dreams, and Inference”, *Journal of Consciousness Studies*, 21 (1-2): 6–32.

precisão de todas as nossas sensações, com exceção dos olhos, porque os olhos fazem parte do sistema nervoso central.

E mesmo assim ainda temos compreensão, ainda sonhamos. Então, acho que esse é um exemplo maravilhoso do cérebro como uma espécie de órgão activo, criativo e construtivo, que apenas se contextualiza ao obter essa ponderação precisa e esse tipo de controlo de ganho no equilíbrio certo.

**Pergunta:** Tenho uma provocação filosófica para o Professor. Em alguns artigos, parece defender uma visão corporificada ou anti-representacionalista. Mas, noutros artigos, parece realmente argumentar a favor de uma visão representacionalista. Qual é a sua posição real sobre o que é conhecida como a “guerra da representação” na Ciência Cognitiva e Filosofia da Mente? O Professor é a favor ou contra as representações em geral?

**Karl Friston:** A minha resposta é a seguinte: depende um pouco de quem é o meu interlocutor. No outro dia, ouvi uma frase bonita numa comédia australiana sobre querer “agradar” às pessoas.

Se estou a falar com um céptico – alguém que é anti-realista – o Princípio da Energia Livre pode ser usado para defender essa posição: “sim, é absolutamente verdade!” Porquê? Porque o organismo nunca estará

directamente exposto ao que está “lá fora”, assumindo que haja alguma coisa fora.

Tudo a que temos acesso é o *sensorium* da experiência. Somos um cérebro “fechado” numa caixa e tudo o que acreditamos sobre o mundo é apenas uma fantasia criada por ele. Então, esta argumentação seria muito consistente com a abordagem céptica da representação.

Por outro lado, se estiver a falar com um realista ou um externalista, a leitura da inferência activa ou a aplicação do Princípio da Energia Livre celebrará isso, porque todos esses aspectos estão relacionados com a sensação activa corpórea: trata-se de como EU me envolvo fisicamente com o meu mundo para obter o tipo certo de sensações que tornem as minhas crenças internas mais consistentes com esse mundo.

Portanto, podemos ler a auto-evidência, literalmente, como a recolha de evidências do mundo real para o modelo-Eu desse mundo. E se EU sou esse modelo, estou a recolher provas da minha existência, e essa é uma forma de olhar para os imperativos existenciais que são implicados por este tipo de auto-evidência.

Mas o Steven levantou uma noção interessante que é o representacionalismo. Se considerarmos o Princípio da Energia Livre, posso afirmar que, independentemente de com quem esteja a falar, que ele é essencialmente representacionalista. Absolutamente. Esse é o objectivo

de separar o cérebro do mundo: o mundo passa agora a ter crenças sobre ele mesmo. E isso é apenas uma representação probabilística.

Assim, o Princípio da Energia livre e a inferência activa não admitiriam um enativismo radical. Mas se for um tipo mais suave de enativismo, penso que a teoria pode acomodá-los alegremente.

**Pergunta:** Que consequências tem o Princípio da Energia Livre em relação àquilo que os filósofos realmente querem ter respostas, isto é, em relação à existência das qualidades subjectivas da experiência, os chamados *qualia*? A sua abordagem avança algum tipo de conhecimento relevante para lidar com esta noção filosófica?

**Karl Friston:** Esta é uma questão muito interessante e desafiadora, que se tornou um foco recente entre vários colegas meus que se dedicam à fenomenologia da subjectividade.

O meu melhor resumo do pensamento actual relacionado com as pessoas que conheço que se focam na filosofia da mente e tentam enquadrar a subjectividade dentro do Princípio da Energia Livre e da inferência activa seria que, para se ter uma experiência qualitativa, é necessário que tal seja uma acção interna (*covert action*).

Se eu fosse um psicólogo, seria o mesmo que dizer que tenho que poder atentar a algo particular: se não posso escolher no que focar a minha atenção, então não posso ter a experiência qualitativa disso.

Muitas vezes, tal está ligado às noções de transparência e opacidade fenomenais que o filósofo Thomas Metzinger aborda, por exemplo. Então, para que algo seja tornado opaco – e agora estou a “ler” essa opacidade como uma espécie de isomórfico com uma qualidade de experiência de uma percepção – é necessário ter a capacidade de alterar a precisão da evidência sensorial que são relevantes, ou seja, o erro de previsão.

Mas se eu fosse um neuroanatomista, diria que ter experiências qualitativas – ou pelo menos ter esses tipos de percepções que podem ser tornadas opacas – significa que eu deveria ser capaz de encontrar projecções neuronais anatómicas no cérebro que possuem um efeito neuromodulador em algumas partes mais internas do mesmo numa determinada hierarquia.

E, se eu voltasse a ser psicólogo, esperaria que víssemos essas projecções apoiadas na capacidade de atender, endogenamente, a diferentes partes do sensorio. Assim, o que temos aqui é uma história, uma história que liga uma espécie de qualidade da experiência com atenção, com a selecção activa de fontes de evidência das minhas fantasias e actualização

das minhas crenças. Para mim, a selecção activa não é um tipo de comportamento evidente.

Podemos ver paralelos, certamente, com a teoria pré-motora da atenção de Rizzolatti, no sentido de que agir no mundo para obter o tipo certo de informação precisa pode, na verdade, envolver uma acção aberta à medida que apalpo visualmente o mundo.

Mas se eu internalizar isso, e agora pensar naquela palpação do sensório a internalizar-se por meio de várias fontes selectivas de erros de previsão ou evidências sensoriais – datando-as com o controlo de precisão – tal só pode ser um tipo de acção mental. Para mim, tal é um teste decisivo para a qualidade da experiência, ou a leitura dos *qualia* como experiência qualitativa.

Se concordarmos com essa leitura, há algumas condições que precisam de estar satisfeitas antes que possamos envolvermo-nos nesse tipo de acção mental: é necessário ter um modelo generativo que possa gerar as previsões da precisão que se está a usar como base da acção mental, o que imediatamente nos informa que é necessário ter um modelo generativo do futuro.

Ora, isso diz muito sobre o tipo de criaturas, ou sistemas, ou partículas, que poderiam possivelmente ter experiências qualitativas: têm de ser aquelas criaturas bastante sofisticadas que têm modelos generativos do futuro, porque é necessário ter um

modelo das consequências da acção, mesmo que essa acção seja secreta ou interna (*covert*).

Acho que tal é muito importante. Há uma espessura temporal que está implicada neste tipo de explicação mecanicista da experiência qualitativa. Existem outros argumentos que colegas, em particular pessoas como Maxwell Ramstead, Mark Solms e outros, que escreveram recentemente a defender uma espécie de formulação mais abstracta da senciência.

Estes colegas tentam pensar a experiência qualitativa como sendo projectada numa “tela” como um Manto de Markov (*Markov Blanket*), do ponto de vista da teoria clássica da informação, ou como uma tela holográfica do ponto de vista da teoria quântica da informação.

A ideia básica é que talvez um sistema físico possivelmente distribuído no cérebro possa ser subdividido ainda mais através do *Markov Blanket* que envolve esse sistema – ou, se adoptar a visão de um teórico da informação quântica, a tela holográfica que contém a informação clássica que o aspectos internos desta estrutura produzem – seja capaz de separar essa estrutura interna do resto do cérebro.

Há argumentos que defendem que, para algo poder ser consciente e ter experiências conscientes, é necessária a existência deste Cobertor de Markov ou de uma tela holográfica única, o que, curiosamente, nos leva de volta a um Teatro Cartesiano.

No entanto, não há nenhum dualismo essencial implícito nesta visão, mas tal traz-nos de volta, de forma interessante, à noção de uma tela interna. O que é crucial nesta leitura é que os estados internos que observam esta tela nunca poderão ver-se a si mesmos, porque não é possível inserir outros fluxos internos.

Essa história pode ser desvendada no contexto da atenção e acção interna ou de uma história de acção mental, simplesmente porque a única maneira pela qual o interior pode mudar as projecções nesta tela interna é através desta acção mental. Mas temos sempre de voltar a essa noção de acção mental no final do dia, não podemos escapar dela.

**Pergunta:** Para concluir, não tenho certeza se o Professor Friston conhece o novo livro *The Model of the Mind* de Grace Lindsay, uma neurocientista computacional da Universidade de Nova York, que argumenta que a sua abordagem, baseada no Princípio da Energia Livre, é infalsificável do ponto de vista da filosofia da ciência, ou seja, que não é possível compreender, de um ponto de vista científico, se o princípio é ou não verdadeiro. O que acha dessa crítica?

**Karl Friston:** Acho que essa crítica pode fazer todo o sentido: o Princípio da Energia Livre é, de facto, infalsificável. Porquê? É muito simples. Este princípio é um princípio de mínima acção, como o princípio de

mínima acção de Hamilton. Na física e na matemática, ou mesmo na ciência da computação, estes tipos de princípios são apenas métodos ou ferramentas: não são teorias ou hipóteses; eles são apenas ferramentas úteis.

O Princípio da Energia Livre é apenas isso, uma ferramenta: não está nem certo, nem errado – ou o investigador aplica esse princípio, ou não o aplica. O seu propósito é ser aplicado, simulado ou reproduzido, como criação de sentido ou de comportamento senciante.

Por vezes pode passar por combinar simulações ou observar o comportamento, de modo a poder realmente reproduzir o comportamento observado. No fundo, é um princípio ou método que permite simular ou reproduzir, ou realizar um comportamento senciante do tipo básico.

Mas, ao fazê-lo, é necessário comprometer-se a aplicá-lo a um modelo generativo específico. E, nesse ponto, se se está a afirmar que determinado modelo generativo é capaz de explicar este tipo de comportamento, este tipo de criatura ou este paciente em particular, então tal torna-se uma hipótese e tal é dificilmente falsificável.

O que se faz neste caso é, portanto, aplicar o Princípio da Energia Livre a diversos modelos generativos de

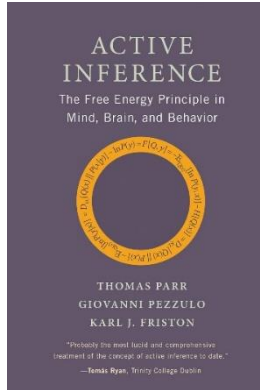
determinada hipótese para descobrir qual é, efectivamente, o melhor modelo para essa hipótese.

Mas se se interpretar o modelo generativo, aplicando o Princípio da Energia Livre como uma teoria, então tal já se torna certamente falsificável: pode haver um melhor modelo de julgamento que explicará o comportamento de determinada pessoa ou um tipo de comportamento particular. Mas o próprio princípio é apenas isso, um princípio: não pode ser falsificado em si mesmo.

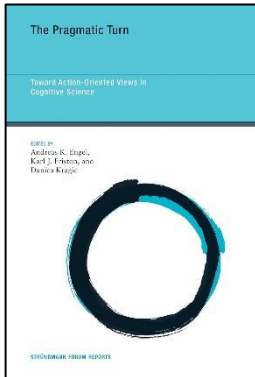
Para concluir, o que o Princípio da Energia Livre oferece é uma estrutura teórica formal na qual podemos compreender a noção de falsificação: trata-se apenas de comparar as evidências da hipótese nula em relação à hipótese alternativa.

É por isso que não há nenhum problema em afirmar que, de um ponto de vista particular, o Princípio da Energia Livre é infalsificável, e é por isso que digo que a autora está absolutamente correcta.

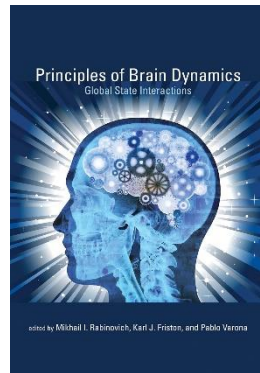
## Livros de Karl Friston



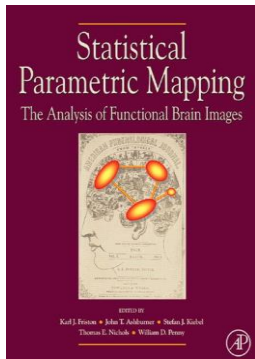
2022



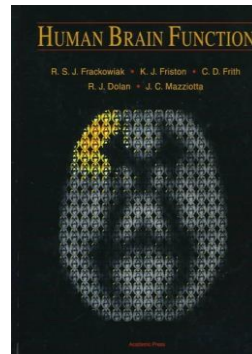
2016



2012



2011



1997



## VIII. Diálogo com Christof Koch



Christof Koch é um neurofisiologista e neurocientista computacional mais conhecido pelo seu trabalho sobre a base neural da consciência.

É o Presidente e Cientista-Chefe do Allen Institute for Brain Science em Seattle e é o Cientista-Chefe da The Tiny Blue Dot Foundation.

De 1986 a 2013, foi Professor do Instituto de Tecnologia da Califórnia. O principal colaborador de Koch no esforço em localizar os correlatos neuronais da consciência foi Francis Crick, o Prémio Nobel de Medicina.

Mais recentemente, Koch trabalhou em estreita colaboração com o psiquiatra e neurocientista Giulio Tononi e tem vindo a desenvolver a Teoria da Informação Integrada da consciência.

Koch é o autor de vários livros tal como *The Quest for Consciousness: a Neurobiological Approach* (2004), *Consciousness: Confessions of a Romantic Reductionist* (2012) e *The Feeling of Life Itself - Why Consciousness is Widespread but Can't be Computed* (2019).

Mais informações: <https://christofkoch.com/>

**Pergunta:** Sabemos que a consciência tem colocado grandes dificuldades à ciência contemporânea. O Professor Koch acha que precisamos de uma espécie de redefinição da natureza física fundamental para explicar a consciência? Ou a física actual e o fisicalismo são suficientes para lidar com a consciência?

**Christof Koch:** Depende do que entendermos por fisicalismo. Apoio uma reformulação da Teoria da Informação Integrada (TII) que afirma que não é apenas o poder causal extrínseco que é descrito pela física convencional, mas também é necessário investigar o poder causal intrínseco, porque é isso que a consciência é. Tal não significa que precisamos de uma nova teoria integrada da mecânica quântica e da gravidade, como é defendido pelo Sir Roger Penrose.

Temos que ver para que caminhos a física irá evoluir, mas neste momento, e dado o facto de o cérebro funcionar a 300 graus Kelvin, não creio que, mesmo com uma reforma onde conseguimos atingir uma teoria única e completa das leis físicas, tal faça diferença para o cérebro, uma vez que os cérebros não funcionam nessa escala. Posso estar errado, mas essa é minha intuição.

**Pergunta:** A Teoria da Informação Integrada (TII) é composta por cinco axiomas, mas alguns autores criticam a inclusão do “Princípio da Exclusão”, por

considerá-lo um tipo de princípio meramente arbitrário ou um mecanismo *ad hoc*. O Professor Koch acha que esse princípio da exclusão é um ponto fraco da TII que precisa de ser melhorado, ou acha que o facto de qualquer experiência consciente ser definida por si só é suficiente para apoiar a inclusão do princípio da exclusão?

**Christof Koch:** Discordo totalmente dessa crítica porque acho que esse princípio é essencial. Porquê? Porque, caso contrário, iríamos obter uma multiplicidade de experiências de consciência e, por consequência, uma multiplicidade de poderes causais. Portanto, é uma parte essencial da teoria.

Este princípio da exclusão não foi acrescentado posteriormente ou *ad hoc*, mas fez sempre parte da teoria pois, em última análise, dada toda a combinatória possível, existe um número muito grande de mecanismos possíveis que existem por si só. A TII diz que só existe um possível, aquele que é o máximo de toda a granularidade espacial e temporal.

Tal significa que existe apenas uma consciência e não pode haver um número infinito de consciências na minha cabeça. Por que é que esta escala espacial e temporal específica está a ter essa relação particular no meu cérebro?

A resposta para essa pergunta é a seguinte: porque é essa relação particular que maximiza o poder causal

intrínseco. Se tivesse neurónios adicionais, ou se tivesse menos neurónios, teria menos poder causal. Se olharmos para uma escala de tempo diferente, terá menos poder causal. Essa é a afirmação da teoria.

A teoria pode estar errada, mas não se pode remover arbitrariamente um axioma e ainda assim fazer com que toda a teoria sobreviva.

**Pergunta:** Houve uma pequena mudança ou actualização da versão inicial da TII a que chamamos no presente da versão 3.0 da TII, onde há uma tentativa de redefinir a consciência como o poder de causa e efeito maximamente irreduzível de qualquer rede. O Professor acha que seria mais correcto pensar na teoria como sendo sobre poderes causais integrados, em vez de informação integrada em si mesma?

**Christof Koch:** Talvez, tal parece-me um argumento linguístico. Em última análise, trata-se de consciência. Bem, se pensar um pouco mais, podemos afirmar que são as duas coisas. A afirmação é que o poder causal intrínseco é o que a consciência é. Como afirmei, é uma identidade.

Na TII, rotulámos essa informação integrada como  $\Phi$ , mas a teoria é mais do que apenas sobre  $\Phi$ , porque qualquer teoria da consciência não tem apenas de explicar como a consciência se enquadra na ordem natural das coisas, mas também tem de explicar por

que é que a consciência da dor é diferente da consciência do prazer ou é diferente de ver o espaço.

Uma boa teoria da consciência tem de conseguir explicar a fenomenologia e tem de conseguir explicar por que determinados estados fazem-nos sentir o que sentimos, e acho que esta teoria da TII faz exactamente isso.

**Pergunta:** Recentemente, houve uma polémica no estudo científico da consciência onde vários cientistas e filósofos assinaram uma carta pública a defender que a TII era uma teoria pseudocientífica, que não era falsificável. Que tipo de experiência acha que poderia falsificar a sua teoria da TII?

**Christof Koch:** Muitos delas, na verdade: a TII faz testes e previsões concretas. Por exemplo, podemos observar uma rede específica e perguntar se ela maximiza o poder causal intrínseco. É uma previsão muito particular: depois de ter a matriz de probabilidade de transição daquela rede neural podemos, em princípio, calculá-la. E aí podemos perguntar: tal maximiza a potência? A fenomenologia desse sistema explica realmente isso?

Refiro-me brevemente a este tema num artigo sobre a fenomenologia do espaço: e a fenomenologia do fluxo do tempo, e a fenomenologia da cor? Existe no

momento um consórcio<sup>95</sup> que procura colaborar testando previsões entre a TII e a Teoria do Espaço de Trabalho Global (TETG).<sup>96</sup> A TETG diz que o substrato da consciência é o parietal pré-frontal, e a TII diz que está na parte de trás do cérebro.

Portanto, estamos neste momento a testar cientificamente várias previsões da teoria. A TII diz que um organismo estará consciente enquanto estiver a experienciar: se experienciar algo por 10 segundos, haverá um substrato físico dessa experiência consciente por 10 segundos. A TETG diz que não: o substrato só está presente quando envia a transmissão pela primeira vez e depois desaparece – novamente, há uma grande variedade pelas quais essa teoria pode ser testada.

Outro exemplo: há um método de teste de complexidade de perturbação que permite avaliar, num contexto clínico, quando é que o paciente está consciente ou não. Tal vem directamente da TII e essa previsão está a ser testada agora mesmo em ensaios clínicos com seres humanos.

---

<sup>95</sup> Publicação original: Melloni L, et al. (2023) “An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory”, *PLOS ONE*, 18 (2): e0268577.

<sup>96</sup> Em inglês, Global Neuronal Workspace.

**Pergunta:** Tal é um bom sinal, certo? Porque muitas teorias da consciência não podem ser testadas, então tal parece-me ser uma vantagem da TII em relação a outras teorias da consciência.

**Christof Koch:** Correcto. É essencial até, diria.

**Pergunta:** Em relação à Inteligência Artificial, o Professor Koch argumenta que os computadores, pelo menos aqueles que usam uma arquitectura de Von Neumann, nunca poderão ser conscientes. No entanto, concordaria que, se os poderes causais intrínsecos do cérebro fossem reproduzidos num substrato artificial de silício, por exemplo, então poderíamos assumir que eles seriam conscientes?

**Christof Koch:** Parece-me que tal é plausível: não há nada de mágico no cérebro, não há nada de sobrenatural no cérebro. Um transístor típico pode estar conectado com dois ou três receptores. Já um neurónio está conectado com 50.000 outros neurónios.

Além disso, há uma enorme sobreposição entre estes 50.000 neurónios. Assim, considerando dois neurónios próximos, um projecta 50.000 neurónios, o outro projecta também 50.000 neurónios, mas eles sobrepõem-se em grande medida um ao outro.

Novamente, tudo isto é muito diferente da conectividade, que em última análise determina o seu

poder causal. Então, se incorporarmos essa força causal num hardware, seja qual for a matéria desse hardware, poderemos obter consciência, sim.

Mas temos que ter muito cuidado aqui: o conceito de “inteligência” é diferente do de “consciência”. Na verdade, são duas coisas muito diferentes. Não há dúvida de que iremos atingir inteligência “artificial”, isto é, uma Inteligência Artificial Geral (AGI), dado que, em última análise, trata-se apenas de *fazer* algo. Mas a consciência não é algo que se *faça*, mas é algo que *é*: a consciência é um estado de *ser*. Já a inteligência, em última análise, é *fazer*. E não há problema em colocar inteligência humana ou sobre-humana em máquinas, mas tal não significa que elas sejam necessariamente conscientes.

**Pergunta:** O Professor Koch mencionou uma relação entre Inteligência Artificial e consciência, afirmando que a Inteligência Artificial é inevitável, mas não temos certeza se será consciente ou não. A minha pergunta é: é mesmo possível que a inteligência exista sem consciência? A inteligência não vem depois de ser consciente? Como pode a Inteligência Artificial existir sem ser consciente?

**Christof Koch:** Essa é uma boa pergunta. O único exemplo que temos neste momento de verdadeira

inteligência flexível é a inteligência humana, e somos de facto conscientes.

No entanto, muitos de nós acreditam que esta é apenas uma forma de nos tornarmos inteligentes, é uma forma que a evolução “escolheu”, mas que existem outras formas – talvez através de silício e software. Até agora, não vejo nenhuma evidência que essas outras formas possam funcionar.

Vamos colocar a questão de outra forma: posso facilmente imaginar quão mais poderosos os computadores, ou as máquinas de Turing, podem ser do que eu, certamente muito mais rápidos do que eu, que são inteligentes ou até superinteligentes. Não vejo, no entanto, *a priori*, por que a consciência seria necessária para isso.

Se pensarmos em *deep learning*, ou se pensarmos em redes generativas ou redes de transformadores, etc., essas tecnologias parecem estar a desenvolver-se muito bem; quanto maior a escala, melhor será o seu desempenho. Não vejo por que é que a consciência seja relevante para essas tecnologias.

Quem sabe se, entretanto, essas tecnologias vão encontrar alguns obstáculos. Mas a minha suposição é que podemos chegar à inteligência artificial geral (AGI) – incluindo à superinteligência – sem envolver a consciência necessariamente.

Não foi desta forma que a evolução produziu inteligência ou consciência, mas agora, como seres humanos, podemos fazer as coisas de forma diferente do que a evolução fez. Se tal é bom para nós, é uma questão totalmente diferente.

Existirem essas superinteligências é realmente bom para o Homo Sapiens-Sapiens? Sou muito céptico quanto a isso, mas é essa a viagem que estamos a fazer no presente, para o bem ou para o mal.

**Pergunta:** Existem vários métodos para estudar o cérebro consciente. Podemos concentrar-nos nos distúrbios da consciência, ou em metodologias baseadas em psicadélicos, mas podemos também concentrar-nos na estimulação cerebral, etc. Quais destes métodos acha que é o mais promissor para estudar a natureza da consciência?

**Christof Koch:** Diria que todos esses métodos são relevantes: não há nenhum que é especificamente melhor, pois tal depende de qual é o histórico e formação do/a investigador/a, qual é a sua expertise, etc.

Todos os diferentes métodos têm desvantagens e vantagens. Alguns deles são mais objectivos, alguns podem incorporar dados da primeira pessoa, enquanto que outros são mais difíceis de incluir esses dados.

Na verdade, podemos perguntar coisas relevantes às pessoas quando estamos a tentar estudá-las, mas, por outro lado, não podemos intervir nos cérebros das pessoas como fazemos com cérebros dos animais na experimentação científica.

Portanto, cada uma dessas técnicas apresenta vantagens e desvantagens: depende, de facto, do interesse de cada um, da sua formação, de que ângulo deseja estudar a consciência.

Os psicadélicos, por exemplo, têm a vantagem de poderem impactar enormemente a consciência, mas também têm desvantagens: a maioria é ilegal e, mais importante que isso, é que não sabemos muito sobre a sua natureza.

Muitos estudos focaram-se nos psicadélicos em ratos, mas é muito difícil saber realmente o que eles estão a experienciar subjectivamente quando consomem algum psicadélico específico. Portanto, diria que depende da sua formação e dos seus interesses profissionais.

**Pergunta:** Pensando nas teorias do processamento preditivo, o que pensa das influências de “cima para baixo” (*top-down*) no cérebro? Acha que são relevantes para pensar a natureza da consciência?

**Christof Koch:** Na maioria das pessoas – deixando de lado os esquizofrênicos – o que cada pessoa experiencia é bastante estável: podemos ser confrontados com essas ilusões visuais por alguns minutos e continuaremos a ver essa mesma ilusão.

Na minha visão, tal diz-me que as influências de “cima para baixo” da minha expectativa não são tão fortes: definitivamente existem, não há dúvida, principalmente quando tenho muito pouco tempo para processar informação, quando tenho 100 milissegundos e tenho que relatar o que vejo nessas condições – essas influências de “cima para baixo” são mais importantes.

Mas, em geral, acho que as pessoas enfatizam demais esse papel: posso olhar para uma parede e ver a parede. Não há nada para prever, nada muda, apenas vejo uma parede vazia.

Então, tal diz-me que a abordagem de “cima para baixo” não é tão importante, pelo menos nestas condições, como as pessoas pensam que é. E não acho que tenha algo a ver com codificação preditiva na maior parte dos casos.

**Pergunta:** Que tipo de consequências éticas acha que a TII pode ter? Especificamente, com a sua relação ao pampsiquismo, que afirma que todas as partículas no mundo possuem algum grau de consciência. Tomemos, como exemplo, a Ética Animal em geral. Podemos

assumir que, pelo menos filosoficamente, uma das previsões da teoria é que muitos animais são capazes de sentir dor e prazer, especialmente criaturas com cérebros desenvolvidos. Acha que esta teoria da consciência tem implicações éticas directas?

**Christof Koch:** Sim, absolutamente. Há muitos anos, justamente por isso, tornei-me vegetariano. E, na verdade, nem mato abelhas e procuro não matar mais insectos porque eles também sentem. Acredito realmente que muitos dos animais que achamos não terem capacidade de sentir dor ou prazer sentem de facto alguma coisa. Mas eles não têm voz.

Uma abelha, por exemplo, não tem voz nem uma mente, como eu. Claramente, o cérebro dela é muito mais simples que o meu, mas ela também sente felicidade quando está apenas a beber um pouco de néctar dourado e a voar sob o sol quente.

E, assim como eu, ela também está presa entre duas eternidades, o início e o final da sua vida: ambos somos jogados neste universo e devemos minimizar o sofrimento de todas as criaturas, incluindo criaturas como abelhas e similares.

Portanto, acho que a TII tem implicações éticas claras e devemos agir de forma coerente com essas consequências, mesmo que tal requiera uma mudança radical do nosso modo de vida.

**Pergunta:** Com todo o conhecimento que o Professor acumulou nas ciências do cérebro, poderíamos dizer que está numa posição vantajosa para corrigir, por exemplo, um mau hábito que tenha no seu dia-a-dia? Esse conhecimento incrível e complexo dá-lhe acesso a uma porta de entrada para qualquer mau hábito que possa querer corrigir, ou esse conhecimento não lhe dá propriamente nenhuma vantagem para a sua vida quotidiana?

**Christof Koch:** Não, não me dá absolutamente nenhuma vantagem: tenho que lutar a mesma luta que todas as pessoas têm. Estou a tentar beber menos e a ser mais compassivo com os outros, seguindo uma tradição específica de sabedoria antiga.

Acontece que sou um grande fã do estoicismo e especificamente de Marco Aurélio, e tal exige disciplina e esforço mental constante. A TII não dá a ninguém nenhum superpoder que possa dizer “Ah, agora percebi! É assim que o cérebro funciona; portanto, posso livrar-me magicamente de todos os meus maus hábitos!”. Não, não é assim que funciona.

Pelo contrário, requer muita disciplina mental e força de vontade: é necessário muito poder causal para evitar activamente as coisas que nos podem prejudicar. Na verdade, é aqui que reside a verdadeira liberdade da TII.

A teoria também levanta algumas implicações sobre o livre arbítrio, afirmando que o livre arbítrio existe no sentido original, no sentido de que posso fazer a diferença para mim, “sou o dono da minha fé, sou o capitão da minha alma”.<sup>97</sup>

Tenho de pensar no que faço, penso ou reflicto e chego a uma conclusão particular de que algo é um mau hábito ou um bom hábito, e que o que quero reforçar ou que quero evitar é isto ou aquilo. É preciso disciplina consistente, dia após dia, e isso, claro, mudará, em última análise, o meu cérebro e, por consequência, a minha vida.

Portanto, tenho a liberdade de determinar o meu próprio destino, dadas as condições de fundo sobre as quais, evidentemente, posso fazer muito menos.

**Pergunta:** Está muito na moda uma abordagem “espiritual” da ciência, como a sincronicidade quântica, a cura quântica, etc., e isso tem impactado a vida de cada vez mais pessoas. Qual é a sua opinião sobre essas abordagens? Considera que sejam imorais? Existe alguma evidência sobre a relevância desse tipo de ideias, ou é apenas algo que vende bem e as pessoas que não têm capacidade para compreender esses

---

<sup>97</sup> Citação do poema “Invictus” do poeta britânico William Ernest Henley, escrito em 1875 e publicado em 1888 no seu *Livro de Versos*.

conhecimentos simplesmente caem nessas armadilhas?

**Christof Koch:** Não há provas de que, neste momento, a mecânica quântica microscópica, em particular o emaranhamento – que é fundamental para a computação quântica – desempenhe qualquer papel no cérebro.

Se olharmos para os computadores quânticos actuais, a temperatura operacional de um computador quântico é de 25 ou 35 mili-kelvin, ou seja, cerca de 50 mil vezes mais fria do que a do cérebro humano.

Portanto, é muito improvável que estes efeitos desempenhem qualquer papel na biofísica ou na bioquímica do cérebro. É claro que muitas pessoas vendem todo tipo de coisas, como curas com cristais, mas isso não nos diz nada, é algo totalmente divorciado de qualquer coisa que o cérebro faça ou que constitua algo de real.

**Pergunta:** Acho que o Professor Koch teve um debate interessante com o Dalai Lama sobre espiritualidade e cérebro. Podemos terminar com um breve resumo do que concluiu desse debate?

**Christof Koch:** Fui duas vezes à Índia e encontrei-me e debati com Sua Santidade, o Dalai Lama. Concordámos em muitas coisas, como, por exemplo, em minimizar o

sofrimento de todas as criaturas conscientes, ao contrário de outras abordagens religiosas, como a do catolicismo que não tem essa preocupação (cresci numa fé católica).

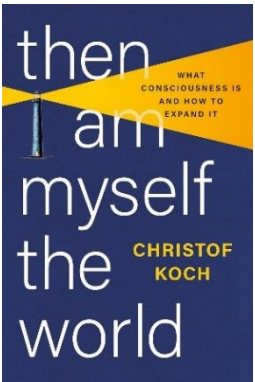
O Budismo, particularmente o Budismo Tibetano, enfatiza que a consciência é comum a todas as criaturas, e não apenas aos seres humanos que podem falar sobre ela. Mas depois divergimos quando falamos sobre encarnação.

O mantra que argumentei é constituído por quatro palavras: *sem cérebro, nenhuma mente* ("no brain, never mind"). O que significa que, uma vez que o seu cérebro se dissolva, ou morra, então o substrato físico da consciência desaparece e, nesse sentido, não há mais consciência sem haver algum portador físico.

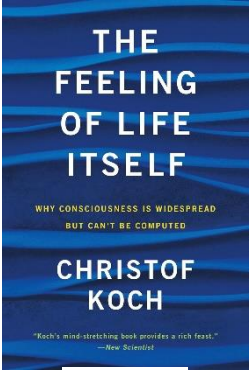
Mesmo que seja uma física exótica, tem de haver algum portador do cérebro, das suas memórias e das suas ideias. Os budistas falam sobre o "bardo", este espaço liminar entre uma vida e o seu renascimento na próxima vida.

Novamente, se há algo da minha vida que eu carrego na próxima vida, tem que haver um lugar no espaço e no tempo que carregue algum substrato das minhas memórias. Caso contrário, não creio que possa existir. E então acabámos o debate com o Dalai Lama a rir-se com suas gargalhadas profundas, afirmando "bem, veremos isso em breve".

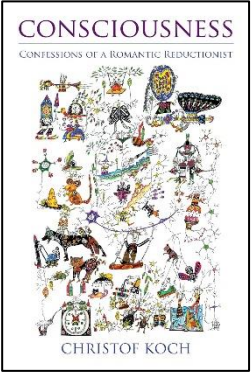
Livros de Christof Koch



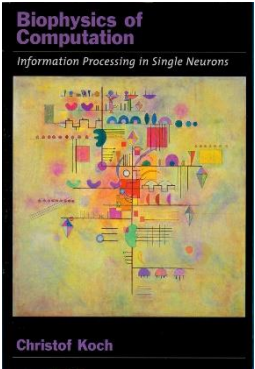
2024



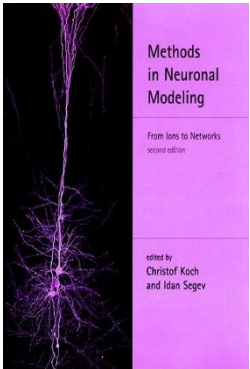
2020



2017



2004



2003

## IX. Diálogo com Joseph LeDoux



Joseph LeDoux é o Professor de Neurociência, Psicologia, Psiquiatria e Psiquiatria Infantil e Juvenil na Universidade de Nova York. É o Director do Emotional Brain Institute e do Nathan Kline Institute for Psychiatric Research.

É vice-diretor do Centro de Linguagem, Música e Emoção no Max Planck-NYU e membro da Academia Nacional de Ciências dos Estados Unidos.

LeDoux é autor de vários livros importantes sobre o cérebro: é autor de *Anxious* (2015), *Synaptic Self* (2002) e *The Emotional Brain* (1996), entre muitos outros. O Professor LeDoux é também cantor e compositor da banda de folk-rock *The Amygdaloids* e da dupla acústica *So We Are*.

Ganhou diversos prémios como o "Distinguished Scientific Contributions to Psychology" pela American Psychological Association (2010), o "Karl Spencer Lashley Award" pela American Philosophical Society (2011) e a "Medalha de Gantt" pela Pavlovian Society (2012).

Mais informações: [www.joseph-ledoux.com/](http://www.joseph-ledoux.com/).

**Pergunta:** O Professor LeDoux é um dos maiores especialistas nos estudos das emoções e da forma como o nosso cérebro, em conjunto com o corpo e o ambiente, cria determinados comportamentos. No entanto, considera-se céptico quanto à utilidade da linguagem vernacular ou popular (o que chamamos em inglês de *folk language*) em relação às emoções em geral. Poderia explicar-nos mais especificamente em que sentido aborda essa problemática?

**Joseph LeDoux:** Não é bem essa posição que defendo: acho que a linguagem popular (*folk*) tem um papel importante quando falamos de estados mentais, uma vez que a nossa vida mental é basicamente vivida nessa linguagem popular. Porém, traço o limite quando aplicamos essa linguagem especificamente ao comportamento.

Por exemplo, o cérebro humano pode responder a um estímulo perigoso como um reflexo, ou pode ter um padrão de acção fixo mais complicado, como a resposta de congelamento, ou pode ainda – num nível mais alto – ter aprendido um hábito, ou pode até ter uma espécie de modelo cognitivo da situação direccionado a um objectivo e responder dessa maneira, e tudo isso pode acontecer inconscientemente.

Mas pode também, enfim, responder conscientemente a esse estímulo perigoso. Ou seja, o meu ponto é que

não podemos pegar num comportamento simples como “congelar ou fugir” (*flight or fight*) e dizer que é, por exemplo, um indicador da emoção de medo, uma vez que há muitos comportamentos pelos quais o medo pode ser expresso, e nem todos o são em termos de uma experiência consciente de medo.

**Pergunta:** Na sua investigação, tem vindo a investigar uma parte específica do cérebro chamada de amígdala, responsável por detectar e responder a ameaças, mas defende que esse órgão anatómico não é um sujeito apropriado para predicar com “medo”. Por que decidiu avançar com uma mudança conceptual do que normalmente chamamos em neurociência de “circuito do medo” para o que o Professor LeDoux chama de “circuito de sobrevivência de defesa”?

**Joseph LeDoux:** Defendo essa mudança conceptual porque realmente não considero que o medo esteja a “borbulhar” na amígdala. Consideremos o circuito de defesa e sobrevivência em geral: cada animal bilateral – ou seja, cada vertebrado que já viveu – tem de ter um circuito que possa detectar e responder ao perigo.

É um tipo de mecanismo muito primitivo e, em cima disso, tem experiências conscientes associadas, onde uma pessoa sabe que é ela mesma que está a ter essa experiência consciente.

Mande até fazer umas t-shirts estampadas que dizem “Sem Eu, Sem Medo” (no *self, no fear*), o que significa que, se não se está pessoalmente envolvido numa espécie de forma autooética e autorreferencial, então não pode haver emoção.

A pessoa tem de fazer parte da experiência para que algo seja uma emoção: responder apenas de forma reflexiva ou instintiva não vai explicar uma experiência emocional. Se vimos um cão a ser atropelado, caído na beira da estrada, a contorcer-se de dor, todos nós projectamos as nossas emoções e sentimentos naquele cão.

Mas o que estamos de facto a observar são “apenas” os reflexos do animal quando se contorce e rosna e assim por diante. Ora, na minha visão, não estamos perante indicadores de dor efectiva. Não estou, claro, a defender que o cão não esteja a sentir dor naquele momento: antes, estou a afirmar que temos de traçar uma linha e separar essas respostas automáticas e reflexivas das respostas associadas às experiências conscientes de dor ou medo, dado que há uma diferença profunda entre elas.

**Pergunta:** Essa visão parece criticar as linhas tradicionais de investigação científica sobre as emoções. Por exemplo, quando fazemos investigação com animais, os cientistas tendem a observar o

comportamento e tentam rotular esse comportamento com uma emoção humana específica. O Professor LeDoux acha que esta é uma forma incorrecta de fazer investigação sobre emoções e animais?

**Joseph LeDoux:** Esse é um ponto interessante. Estava a escrever sobre essa questão há algum tempo atrás e estava a lembrar-me da referência bibliográfica que estava a usar, um capítulo de um livro sobre antropomorfismo da Elizabeth Knoll, onde aborda a perspectiva de Darwin.

A época em que Darwin viveu era a época da Inglaterra vitoriana e o antropomorfismo fazia parte daquela cultura, era uma espécie de “modo de vida” daquele tempo.

Darwin estava a ter muitos problemas para fazer com que a sua teoria da evolução/selecção natural fosse aceite devido às implicações religiosas e tomou uma decisão explícita de falar sobre mentes animais em termos humanos, em vez de mentes humanas em termos animais, porque ele achava que a segunda opção não seria bem recebida.

As teorias de Darwin são realmente o ponto de partida para o estudo moderno das emoções, uma vez que os seus acólitos e seguidores no final do século XIX viam o comportamento como um embaixador da mente.

Embora isso seja certamente verdade até certo ponto, não é uma indicação clara, porque os comportamentos

que estudamos nos animais tendem a ser tipos de respostas reflexivas inatas, e não respostas que são necessariamente produtos da experiência consciente do medo.

A atribuição livre de explicações conscientes sem qualquer evidência foi o que desencadeou toda a revolução behaviorista na psicologia. Não se pode dizer que algo “parece consciente, logo é consciente”, a menos que possamos realmente testar isso de alguma forma. E é muito difícil fazer isso em animais.

Às vezes, sou acusado de negar as emoções dos animais, mas isso não é verdade. Só acho que é metodologicamente muito difícil testar isso nos animais em geral.

**Pergunta:** O Professor LeDoux também afirma que o que é de facto conservado pela evolução não é o comportamento, como a maioria dos cientistas e filósofos argumenta, mas são os circuitos ligados aos comportamentos que são conservados. Portanto, na sua visão, o mesmo comportamento pode variar para o mesmo circuito particular. Pode dar-nos um exemplo apropriado dessa distinção?

**Joseph LeDoux:** Não estou a defender que o comportamento não esteja envolvido nesse processo evolutivo. O que acredito é que os comportamentos, em geral, estão em um nível diferente de análise. Por

um lado, o sistema nervoso controla o comportamento e, por outro, a interacção entre o comportamento do animal e o seu ambiente é controlado pelo seu sistema nervoso.

À medida que as espécies evoluem, mudam e tornam-se em outros grupos, outras espécies e todas estas mudanças estão a contribuir de forma complexa para o desenvolvimento dos comportamentos.

Mas o que o Steven afirmou sobre não ser o comportamento que foi transmitido pela evolução é um ponto que fiz quando pensamos no início da vida. A primeira célula que viveu o tempo suficiente para se reproduzir e dar origem a outras células tinha de ser capaz de detectar o perigo e responder ao perigo no seu ambiente.

Agora, o que isso significava para uma célula que viveu há 3,7 ou 4 mil milhões de anos é que tinha de satisfazer várias necessidades essenciais para se manter viva: identificar e afastar-se de elementos perigosos (ou seja, tóxicos) do ambiente. Mas ela também teve de ter a capacidade identificar e incorporar nutrientes e equilibrar fluidos e electrólitos. E para que a espécie continuasse, a replicação era essencial.

Se uma célula bacteriana primitiva encontra um elevado nível de acidez, ela usa os seus flagelos para se afastar. Caso contrário, não sobreviverá o suficiente para se reproduzir. Estes tornaram-se requisitos

fisiológicos fundamentais da vida, de todas as coisas que já existiram. Mas cada espécie resolve os “seus” problemas à sua maneira.

Assim, o que defendo é que aquilo que herdamos ao longo de quatro mil milhões de anos de vida não é o comportamento em si, mas a exigência de responder ao perigo, identificar nutrientes, equilibrar fluidos e iões e reproduzir-se. Estes são requisitos de sobrevivência de qualquer organismo vivo, qualquer coisa viva, seja uma única célula ou um organismo gigantesco, como nós seres humanos.

Portanto, não estava a tentar dizer que herdámos a nossa amígdala das bactérias: o que herdamos das bactérias foi a capacidade de sobreviver, a necessidade de sermos capazes de detectar o que é prejudicial e útil, e permitir qualquer outra coisa que o animal ou o organismo tem que fazer para sobreviver.

Chamo a isso de “necessidades de sobrevivência” ou “estratégias de sobrevivência” que têm implementações biológicas específicas de maneiras únicas para cada tipo de grupo de animais e para cada tipo de espécie de animal e, até certo ponto, para cada animal individual.

Em suma, todos os animais têm estes requisitos de sobrevivência, mas apenas os organismos com sistema nervoso possuem circuitos de sobrevivência.

**Pergunta:** Contra muitos neurocientistas, o Professor LeDoux não concorda que as emoções estejam biologicamente ligadas ao cérebro, mas argumenta que elas surgem do processamento cognitivo inconsciente. Por que acha que tantos neurocientistas têm uma visão tão diferente da sua sobre a natureza das emoções?

**Joseph LeDoux:** Depende de quem é o neurocientista e no que ele está interessado. A teoria cognitiva da emoção não é nova. William James tinha uma versão. Mas a versão moderna remonta à teoria da dissonância cognitiva de Leon Festinger na década de 1950 e à teoria cognitiva da emoção de Stanley Schachter e Jerome Singer na década de 1960. Baseei-me nisso na minha investigação com pacientes com cérebro dividido (*split-brain*) na década de 1970.

Eu e o meu mentor Mike Gazzaniga observámos que, quando o hemisfério direito de um paciente com cérebro dividido produzia um comportamento e perguntávamos ao hemisfério esquerdo “por que fez isso”, ele inventava, confabulava – gerava uma narrativa – para tentar fazer sentido daquele comportamento.

Ora, tal é consistente com a teoria da dissonância cognitiva de Leon Festinger, que diz que quando se tem informação discordante, é preciso resolvê-la de alguma forma. Os comportamentos gerados a partir de um sistema não-consciente – neste caso, o hemisfério direito – seriam uma fonte de stress ou ansiedade, uma

vez que todos acreditamos que temos livre arbítrio, quer o tenhamos ou não, mas acreditamos que temos.

Se o nosso corpo está a produzir comportamentos pelos quais não somos responsáveis, tal é muito perturbador e temos que encontrar uma maneira de contornar isso.

Assim, eu e Michael Gazzaniga levantámos a hipótese de que talvez os sistemas emocionais sejam aqueles sistemas no cérebro que produzem esses comportamentos inconscientemente e que podem exigir algum tipo de interpretação cognitiva para ajustá-los à avaliação mental e ao nosso autoesquema sobre quem nós somos.

É por isso que me foquei nos estudos com ratos, porque queria compreender comportamentos inconscientes que nos humanos podem desencadear interpretações cognitivas.

A teoria da emoção de Schachter e Singer surgiu directamente da teoria da dissonância cognitiva de Festinger. Evoluiu bastante desde a década de 1960, quando Schachter e Singer a propuseram, mas é uma abordagem muito viável para a emoção. Para mim, é uma maneira muito mais realista de pensar sobre as nossas emoções.

Imagine que o leitor se encontra no topo de uma montanha e a sua mochila caiu do penhasco, o sol está a pôr-se, não tem comida nem água nem agasalho: de

repente, entra num estado de medo e ansiedade sobre o que vai acontecer consigo. Mas o sistema da amígdala, o chamado sistema do medo, ou o que eu chamo de sistema de ameaça, evoluiu como um sistema de defesa predatório.

Não há nenhum predador no topo da montanha que esteja a deixar o leitor com medo. O medo não é uma emoção ligada a um determinado tipo de estímulo: é uma interpretação. O leitor passa a monitorizar os sinais corporais do estômago que indicam que está com pouco suprimento de energia; ao mesmo tempo está a começar a receber sinais termorreguladores de que não está aquecido o suficiente.

Deste cenário todo, começa a preocupar-se com a possibilidade de morrer de fome, congelar ou desidratar até a morte: pode sentir medo por todos os tipos de razões na vida que não têm nada a ver com a existência de predadores no seu contexto.

**Pergunta:** Isso é interessante. Que tipo de implicações clínicas tem a sua abordagem em relação, por exemplo, aos transtornos de ansiedade? Do ponto de vista teórico, o que há de novo na sua perspectiva?

**Joseph LeDoux:** O meu grupo de investigação publicou recentemente um artigo na revista *Molecular Psychiatry* de título “Colocando o ‘mental’ de volta em ‘mental’” (“Putting the ‘mental’ back into ‘mental’”). A ideia é que

toda a abordagem da psicoterapia, o tratamento dos transtornos mentais que começou na década de 1950, foi impulsionada por uma agenda behaviorista que marginalizou a mente consciente subjectiva.

Na década de 1950, abordagens como a terapia comportamental e a terapia psicofarmacológica tornaram-se abordagens padrão. A terapia comportamental veio obviamente directamente de Skinner, Watson e assim por diante, e abriu o caminho para a terapia cognitiva, que começou com um ângulo mental (Arron Beck, o fundador, era psicanalista), mas lentamente tornou-se mais baseada em métricas objectivas e um tanto behavioristas.

Mas e a psicofarmacologia? Quem está a trabalhar nesses laboratórios das indústrias psicofarmacêuticas, a testar animais para descobrir medicamentos que ajudem as pessoas? Esses investigadores foram treinados por behavioristas nas décadas de 1940, 1950 e 1960.

A suposição era que, se colocar um rato em algum tipo de situação ameaçadora ou stressante e lhe der uma droga que o torne menos tímido comportamentalmente, assumir-se-á que ele é menos tímido porque é menos medroso ou ansioso, e que quando damos a droga a um ser humano, a pessoa ficará menos assustada ou ansiosa.

Mas a indústria farmacêutica começou a abandonar esta abordagem da “anti-ansiedade” porque esta não conseguiu gerar nenhum conhecimento novo. A maioria das classes de medicamentos para lidar com ansiedade foram encontradas acidentalmente na década de 1960, como os inibidores da recaptação de norepinefrina ou benzodiazepínicos.

O problema é que, em poucas palavras, um medicamento que faz os ratos congelarem ou evitarem o frio não vai alterar significativamente a angústia humana, a ansiedade, a preocupação e o medo que uma pessoa experiencia: precisamos de ajudar o paciente a controlar a fisiologia do comportamento, mas isso não basta, também precisamos de levar a sério a experiência subjectiva do paciente se queremos resolver os seus problemas.

**Pergunta:** O Professor LeDoux pode dar algum crédito a Sigmund Freud por alguma ideia relacionada com o seu trabalho?

**Joseph LeDoux:** O que diria é que os desenvolvimentos que ocorreram na década de 1950 para criar novas formas de terapia foram esforços para escapar de Freud, porque ele era considerado demasiado subjectivo na sua abordagem.

Porém, ao se livrarem da subjectividade que Freud trouxe para a psiquiatria, os cientistas da mente

“deitaram fora o bebê junto com a água do banho”. Em vez de dizer “Ok, talvez o profundo e obscuro inconsciente não seja toda a história da mente e nem tudo seja repressão sexual” e todas as outras coisas pelas quais Freud foi criticado – mas isso não significa que devemos deitar fora toda a subjectividade.

Podemos não conseguir corrigir tudo originando mudanças somente no comportamento. Como disse anteriormente, até mesmo a Terapia Cognitivo-Comportamental (TCC) tornou-se uma espécie de abordagem de relato do comportamento.

Quando Beck e Ellis iniciaram a TCC nas décadas de 1950 e 1960, havia elementos mais subjectivos nela. Estes foram afastados mais tarde, uma vez que toda a indústria se voltou para o pagamento de seguros. Os terapeutas tinham critérios específicos para assinalar – métricas objectivas – que poderiam ser usados para identificar ou categorizar qual era o problema do paciente.

Se olharmos para o NIMH RDoc em Inglaterra, por exemplo, deparamo-nos com todas essas métricas que usamos para identificar e classificar aspectos objectivos, e em algum lugar no meio dessa longa lista, podemos encontrar o relato verbal – que é o que mais se aproxima da experiência subjectiva do paciente.

No entanto, o relato verbal é tratado exactamente como qualquer outro sintoma. É neste detalhe que

devemos realmente perguntar: mas, afinal, por que é que as pessoas procuram um terapeuta em primeiro lugar? Porque eles se sentem mal e querem se sentir melhor. Portanto, na minha opinião, temos de corrigir a forma como os pacientes se sentem, e não apenas a forma como se comportam, e isso implica levar a subjectividade a sério.

**Pergunta:** Imagine o exercício em que teria de programar um novo tipo de ser humano desde o início e os seus respectivos circuitos de sobrevivência. O Professor LeDoux acha que poderíamos programar esses circuitos para os seres humanos procurarem a felicidade ou o bem ou o prazer?

**Joseph LeDoux:** Definitivamente, acho que poderíamos usar uma nova forma de programação humana, o que é uma boa ideia porque não estamos num bom caminho como espécie no momento.

Se planear esse tipo de projecto, devemos sempre começar com o positivo e o negativo e tentar construir algo a partir daí. Mas temos de nos perguntar, em termos dos mecanismos básicos, o seguinte: quando um rato está a ser reforçado com um comportamento, é o prazer que está a causar esse reforço?

Toda esta ideia de centros de prazer no cérebro, criada por Olds e Milner na década de 1950, surgiu por acidente. Olds escreveu um artigo científico chamado

“Centros de Prazer do Cérebro”. Alguns anos mais tarde, perguntei a Peter Milner, quando ainda estava vivo: “O que vocês estavam realmente a pensar quando estavam a investigar esses centros de prazer?”

E ele respondeu-me: “Bom, antes de mais nada, não estávamos a estudar o prazer; o que estávamos a tentar fazer era encontrar mecanismos de reforço no cérebro”. Se lermos o artigo que Olds publicou, veremos que a palavra “prazer” nunca aparece no artigo e está apenas presente no título. Os editores da revista talvez tenham acrescentado o título ao artigo, e Olds deixou-se seduzir por ele e começou a promover a ideia de “centros de prazer”.

Logo depois, Rory Wise começou então a afirmar que a dopamina é a substância química do prazer, mas tanto Wise quanto Olds rejeitaram a ideia de que estavam a estudar o prazer e que a dopamina era a substância química do prazer.

Mas uma vez que “o gato sai do saco” com essas coisas, “ele nunca mais volta”. Agora, todos se referem ao reforço em termos de recompensa e prazer, mas, em última análise, o que estamos a falar é dos efeitos da dopamina na actividade celular que causa uma mudança comportamental.

Tal não significa que um animal não esteja a sentir prazer, ou que a pessoa não esteja a sentir prazer, mas isso não é a mesma coisa que o reforço de um

comportamento, e esta distinção deve ser clara quando procuramos estudar estas emoções positivas.

**Pergunta:** Para encerrarmos, gostaria de saber o que pensa sobre o problema que mais tem ocupado os filósofos. Acha que estamos mais perto de descobrir como é que os *qualia* são formados ou sustentados no cérebro humano? Ou ainda estamos longe de conseguirmos fornecer uma resposta sólida ao problema difícil da consciência?

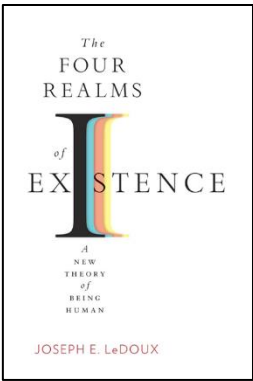
**Joseph LeDoux:** Uma resposta que posso avançar é baseada no seguinte conselho: os cientistas precisam de ter cuidado com os conceitos que importam da filosofia.

A filosofia é um sistema de regras e raciocínio que é óptimo, mas tal não significa que seja assim que o cérebro realmente funciona. Não estou a dizer que os *qualia* não existem, mas acho que o chamado problema difícil da consciência foi criado de uma forma que nunca poderá ser resolvido porque pressupõe o dualismo mente-corpo.

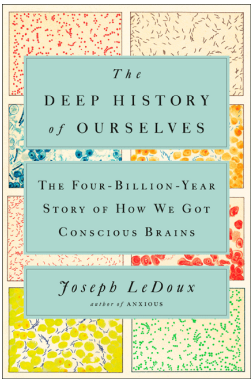
Tal pode ser interessante filosoficamente, mas ter uma perspectiva dualista sobre o cérebro é um fracasso científico. Acho que a maioria dos neurocientistas são materialistas e queremos realmente encontrar alguns mecanismos que façam com que os *qualia* aconteçam na mente.

Tenho um novo livro intitulado *The Four Realms of Existence: A New Theory of Being*, onde proponho uma maneira possível de pensar sobre *qualia*, então talvez seja um bom lugar para o leitor procurar mais respostas.

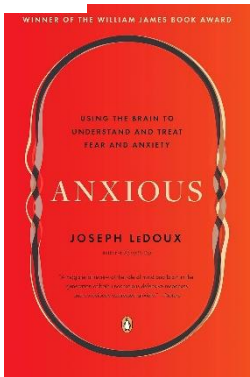
Livros de Joseph LeDoux



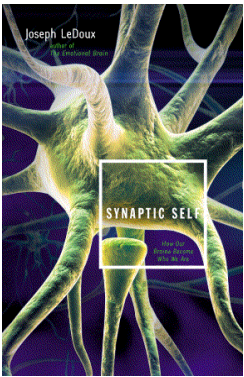
2023



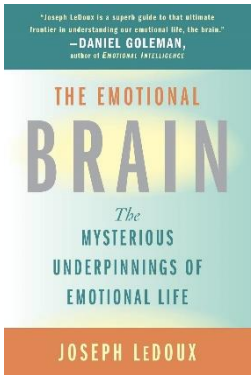
2019



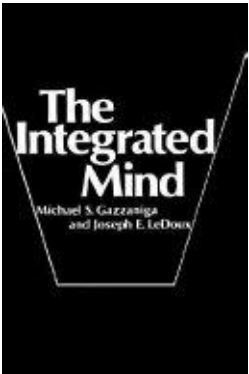
2015



2001



1996



1978



# Conclusão

Este livro procurou introduzir o leitor a alguns dos debates mais actuais e fascinantes da filosofia e da ciência da mente. Assim como os destemidos navegadores lusitanos que ousaram cruzar o cabo das Tormentas, enfrentámos as complexidades dos mistérios da mente e das inúmeras dificuldades de navegação que esta levanta.

O cabo das Tormentas situa-se na Cidade do Cabo, na África do Sul, sendo um dos pontos terrestres mais a sul do planeta. Devido à sua posição peculiar, era considerado intransponível dadas as constantes tempestades intensas e condições impossíveis de navegação. Até 1488, nenhum ser humano teria conseguido passar esse cabo.

Tudo mudou com o navegador português Bartolomeu Dias que, pela primeira vez, conseguiu dobrar o cabo das Tormentas, que viria a ser rebaptizado pelo Rei D. João II como cabo da Boa Esperança, dado que se instaurava, também pela primeira vez, uma rota marítima directa entre o oceano Atlântico e o oceano Índico, entre o Ocidente e o Oriente – especificamente

com a Índia –, um acontecimento que iria mudar para sempre a história económica, cultural e política do nosso mundo.

Como Bartolomeu Dias, espero que alguns dos progressos indicados nesta obra possam vir a desenvolver-se com maior robustez nos próximos anos, esperando que consigamos transformar as Tormentas da Mente num cabo da Boa Esperança.

Cada página deste livro é como mais um dia no alto mar, procurando a terra que talvez nunca chegue. Quiçá a curiosidade de cada um de nós seja o vento necessário para impulsionar as próximas descobertas que possam emergir dos faróis desta escuridão.

Será que valeu a pena? Confiemos nas palavras oferecidas pelo Poeta:

“Valeu a pena? Tudo vale a pena  
Se a alma não é pequena.  
Quem quer passar além do Bojador  
Tem que passar além da dor.  
Deus ao mar o perigo e o abismo deu,  
Mas foi nele que espelhou o céu.”<sup>98</sup>

Que esta *Odisseia da Mente* ofereça ao leitor um ponto seguro nos portos do entendimento e da reflexão. Com o horizonte à vista, desejo que cada leitor encontre

---

<sup>98</sup> Fernando Pessoa, Segunda Parte: X. Mar Português, da obra *Mensagem*.

neste livro uma bússola confiável para as suas próprias explorações e que a sabedoria adquirida nesta expedição ilumine caminhos para futuros desconhecidos, onde a mente humana possa ter as suas actuais fronteiras mais exploradas.

Para encerrar esta conclusão e o livro, uma nota menos poética (novamente). Este livro nasce inspirado em diversos cursos online que organizei como Professor Principal nos últimos 2 anos. Nesse âmbito, tive o privilégio de chegar a mais de 600 estudantes de cerca de 35 países diferentes.

A todos eles, um agradecimento profundo por me terem ensinado muito mais do poderei ter transmitido. Ademais, a perspicaz curiosidade dos estudantes que, com as suas questões e intervenções, inspiraram muitos dos debates e tornaram os cursos muito mais estimulantes e fascinantes para todos, principalmente para mim.

Que a *Odisseia da Mente* continue!



# Agradecimentos

Marco Aurélio, filósofo estóico e Imperador Romano, inicia a sua célebre obra *Meditações* a agradecer a todas as pessoas que se cruzaram na sua vida e com quem aprendeu algo particular. Ora, uma obra desta magnitude não poderia acontecer sem a contribuição directa e indirecta de um variadíssimo número de pessoas, pelo que sigo os passos do Imperador, esperando não deixar ninguém de fora.

Começo, claro, por agradecer aos 8 incríveis pensadores que me deram a honra e a confiança de partilhar o palco deste livro: a David Chalmers, Susan Blackmore (e Alison Seldon), Sir Roger Penrose (e Helen McGregor), Nicholas Humphrey, Joseph LeDoux, Anil Seth, Christof Koch e Karl Friston, um enorme obrigado.

Na execução deste livro, Tássia Vianna foi fundamental nas transcrições dos debates orais em língua inglesa, transformando-os em diálogos escritos que foram depois aprovados por cada um dos intervenientes e traduzidos para língua portuguesa. Ana Monteiro foi elementar na produção criativa da capa deste livro, cujos rostos foram desenhos à mão pelo seu puro talento e criatividade. A ambas, uma sincera gratidão.

Aos colegas da Universidade do Porto, principalmente à Professora Doutora Sofia Miguens, Coordenadora do grupo de investigação Mind, Language and Action Group onde realizo a minha investigação, que tem sido a maior influência para pensar a filosofia de forma aberta ao mundo, e ao Professor Doutor José Meirinhos, Director do Instituto de Filosofia e à Senhora Directora, Professora Doutora Paula Pinto Costa, por apoiarem os diversos projectos sem colocarem qualquer obstáculo. À Isabel Marques, gestora de ciência do Instituto, um obrigado por toda a paciência.

De seguida, agradecer a todos aqueles que têm contribuído para que as minhas ideias cheguem ao mundo, servindo de estímulo constante na minha investigação.

No Brasil, Gabriel Mograbi e Paulo Taddei, da Universidade Federal do Rio de Janeiro; Nythamar de Oliveira, da Pontifícia Universidade Católica de Porto Alegre; e Maria Luiza Ilenaco em conjunto com Osvaldo Pessoa Jr., da Universidade de São Paulo.

Na Roménia, Florin Piscociu e a sua equipa da Mindlifeline (Léa Chibany, Luana Aldea e os restantes membros) têm sido um constante apoio, cuja ética de trabalho me inspiram todos os dias.

Em Malta, Ian Rizzo, Valdeli Pereira, Francois Zammit e os restantes amigos da Philosophy Sharing Foundation têm-me mostrado a importância de uma filosofia

pública, desconfinada da torre de marfim e acessível a todos os cidadãos. No Chipre, o entusiasmo e a sede de conhecimento de Panayiotis Stavrou é uma motivação constante.

No Chile, a amizade e admiração do trabalho feito por Ricardo Ramirez, Camilo Garcia, Marcelo Martín, Nicole Nakousi e restantes colegas, que me deram o privilégio de uma vida de receber, aos 30 anos, a distinção de Professor Honorário da Faculdade de Medicina Andrés Bello, distinção essa que espero honrar ao longo da vida.

No âmbito do meu projecto de Ética da Inteligência Artificial na Medicina, vários colegas devem ser nomeados: Pekka Mäkelä, Raul Hakli e Pii Telakivi (Uni. Helsínquia); Radu Uzskai (Uni. Bucareste); Simona Tiribelli (Uni. Macerata); Heidi Mertes (Uni. Gante); Inês Dutra (Uni. Porto), Sabina Leonelli (Uni. Exeter) e, mais recentemente, Antonio Chella e o grupo dos Estudos Avançados da Universidade de Palermo.

Aos amigos da Universidade Católica de Braga, com quem tive uma curta, mas intensa colaboração, cujo foco na multidisciplinaridade foi um alento: um agradecimento a Augusto Soares Da Silva, Carlos Morujão, Padre João Onofre, Paulo Dias, João Duque, António Melo, Padre Bruno Nobre, Elton Marques, Ângela Leite, Tânia Oliveira e Padre José Lopes – um Obrigado por me receberem na vossa Casa.

Nunca devemos esquecer quem viu em nós algo que mais ninguém viu: ao filósofo Manuel Curado (Uni. Minho) e ao neurocientista Georg Northoff (Uni. Ottawa), supervisores do meu Doutorado, um agradecimento por todas as orientações nesta ainda breve, mas energética carreira académica.

Aos estudantes de doutoramento que tenho/tive o privilégio de orientar e que confiaram na minha orientação, grato a Jaroslav Malík, Romeu Ivoleta e Maria Luiza Ienacco por todas as colaborações.

De forma aleatória, a todos os colegas espalhados pelas várias instituições que têm dado valor à partilha do conhecimento: Klaus Gärtner, Robert Clowes, Glorinda Andrada, Rui Vieira da Cunha, Sâmara Costa, Sara Fernandes, Bárbara Sousa e Brito, James Grayott, João Cordovil, João de Fernandes Teixeira, Leonel Moura, Ralph Bannell, Roberto Pereira, Diogo Gurgel, Diana Tavares, Dina Mendonça, Joana Rita Sousa, Tomás Magalhães Carneiro, Ângelo Milhano, Björn Lundgren, Jorge Gonçalves, Luísa Neto, Peter Singer, Luiz Meirelles, Noam Chomsky, Yinchun Wang, entre muitos outros cujos neurónios do meu hipocampo não me permitem recordar, mas que sempre surgem na mente.

Para concluir esta já longa nota de agradecimentos – devo muito a muita gente –, Marco Aurélio atribui aos deuses “o facto da minha constituição ter sobrevivido tanto tempo a este tipo de vida”: resta-me agradecer aos meus grandes amigos e família, que me têm

ajudado neste “tipo” de vida dedicada ao conhecimento: Sr. Joaquim, Maria José e família (Ana, Manuel, João e Maria Augusta); Rafael, Pedro, João Pedro e Carolina; irmãos Kevin e Dylan; Carlos e Zaida (os verdadeiros responsáveis desta “constituição”); avó Arminda, primos Sandra, Nuno, António, Ana e tios e tias; Mónica e Kiara; Isabel e Isaura; Roberta e Bruno; Ana Maria, Gilberto e Ana Mafalda; a Luna e os gatos Miró, Tareco, Mia, Didi, Manchinhas e Amarelinho (nomes criativos, bem sei).

A Todos, um profundo Obrigado.



# Biografia do Autor

Steven S. Gouveia (Sion, 30/09/1992) é Investigador Contratado no Mind, Language and Action Group do Instituto de Filosofia da Universidade do Porto (Portugal), financiado pela Fundação da Ciência e Tecnologia (CEECIND.02527.2022), onde lidera um projecto de 6 anos sobre Ética da Inteligência Artificial na Medicina, que conta com parcerias das Universidade de Exeter, Yale e Helsínquia (cf. <https://trustaimedicine.weebly.com/>).

Licenciou-se em Filosofia e aos 22 anos inicia o Doutoramento em Filosofia da Mente pela Universidade do Minho (Portugal), sob supervisão do filósofo Manuel Curado, tendo sido investigador-visitante do Minds, Brain Imaging and Neuroethics do Royal Institute of Mental Health, University of Ottawa (Canadá), sob supervisão do neurocientista e psiquiatra Georg Northoff.

Depois de terminar o Doutoramento (2021), foi Investigador do Centro de Estudos Filosóficos e Humanísticos da Universidade Católica Portuguesa, em Braga (Portugal).

Desde junho de 2023, é Professor Honorário da Faculdade de Medicina da Universidade Andrés Bello, Viña Del Mar, Chile, título atribuído na mesma ocasião do Prémio Nobel Sir Roger Penrose. É ainda Consultor Ético de uma empresa de Altruísmo Eficaz "Carreiras Altruísticas".

Publicou diversos livros académicos. Em 2016, organizou a obra "Filosofia e as Artes", tendo convidado alguns dos artistas mais relevantes da prática artística nacional, como a artista plástica Joana Vasconcelos, o rapper Valete, o músico Fernando Ribeiro (Moonspell), Leonel Moura (IA), a poeta Ana Luísa Amaral, entre outros.

Em 2017, co-editou a obra colectiva "Pensar a Democracia" com prefácio de Noam Chomsky e, com Ana Figueiredo Sol, editou "Bioética no Século XXI". Nesse mesmo ano, editou ainda "Philosophy of Mind: Contemporary Perspectives", a sua primeira obra internacional com Manuel Curado na Cambridge Scholars Publishing.

No ano seguinte, publicou a sua primeira obra autoral com a Editora Húmus: "Reflexões Filosóficas: Arte, Mente e Justiça", com prefácio do filósofo da mente João de Fernandes Teixeira.

Em 2019, publicou três obras colectivas internacionais: "Perception, Cognition and Aesthetics" e "Film and Philosophy: Bridging Divides", pela influente editora

internacional Routledge e, pela Vernon Press, co-editou, com Manuel Curado, “Automata’s Inner Movie: Science and Philosophy of Mind”.

Edita ainda a obra colectiva “The Age of Artificial Intelligence: an Exploration” pela mesma editora, que conta com a participação de alguns dos mais influentes pensadores e transhumanistas da actualidade, como Daniel Dennett, Ben Goertzel (criador da famosa robot Sofia), David Pearce, Natasha Vita-More, Roman V. Yampolskiy e Vernor Vinge, entre outros.

No ano seguinte, publicou uma obra colectiva pela editora internacional Bloomsbury – a mesma editora responsável pela publicação de *Harry Potter* – como título “The Philosophy and Science of Predictive Processing”.

Ademais, publica o seu livro autoral mais procurado até ao momento: “Homo Ignarus: Ética Racional para um Mundo Irracional”, pela Editora Minerva, que conta com prefácio de Peter Singer, uma obra focada em diversos problemas éticas como o voto, o humor, a eutanásia, ou a inteligência artificial.

Em 2022, alcança reconhecimento internacional ao publicar a sua terceira obra autoral em língua inglesa pela editora nova iorquina Palgrave, do grupo Springer Nature: com o título “Philosophy and Neuroscience: a Methodological Analysis”.

Ainda no mesmo ano, publica uma obra colectiva de título “Artificial Intelligence: Conversations about the New World” com participação de vários pensadores como Peter Singer, Wulf Loh ou Sabina Leonelli. Em 2023, volta a co-editar uma obra com Manuel Curado de título “Predictive Minds: Old Problems and New Challenges” pela Vernon Press.

Além de uma intensa produção académica, é o anfitrião e produtor do documentário internacional “The Age of Artificial Intelligence: a Documentary”, que conta com a participação de investigadores internacionais e está acessível gratuitamente no YouTube, contando com mais de 55.000 visualizações.

Foi orador em dezenas de conferências científicas com revisão por pares em Portugal e no estrangeiro. Além disso, tem sido convidado como conferencista para palestrar em universidades espalhadas pelo mundo, no Brasil, Argentina, Chile, Malta, Itália, República Checa, Coreia do Sul, Chipre, Roménia, entre outros. Por duas vezes, foi orador na *Science of Consciousness Conference*, a conferência internacional mais relevante dos estudos da consciência. Foi ainda orador na TEDx NOVA, em setembro de 2023.

Tem sido requisitado para variadas entrevistas, tendo participado nos programas de televisão e rádio como “A Minha Geração” (RTP3), “Sociedade Civil” (RTP2), “Linha da Frente” (RTP1), “Muito Barulho para Nada” (RTP2), “Filosofia na Rua” (Antena 2), além de diversos podcasts,

como o “45 Graus”, “Despolariza”, “Pergunta Simples”, “Desassossego”, “Smart Meat” entre muitos outros.

É o professor principal de vários cursos online sobre temas como a democracia, a consciência ou a ética, com a participação de professor convidados como Sir Roger Penrose, Peter Singer, Noam Chomsky, Slavoj Zizek, Paul Bloom, entre outros.

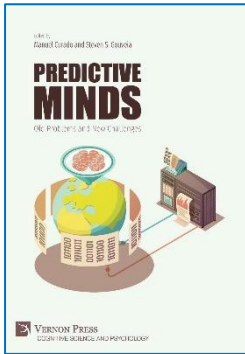
Foi fundador e editor principal da “Apeiron – Student Journal of Philosophy” (2012-2016), que contou com a participação de pensadores como Noam Chomsky, Peter Singer, Daniel Dennett ou Noël Carroll.

Finalmente, tem sido o organizador principal de diversas conferências internacionais, trazendo a Portugal pensadores como Peter Singer (Princeton), Luciano Floridi (Yale), William Child (Oxford), Shaun Gallagher (Memphis), Dan Zahavi (Copenhagen), Karl Friston (London), David Papineau (New York), Tim Crane (Cambridge), entre outros.

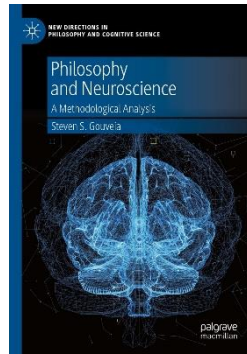
Mais informações podem ser encontradas em [www.stevensgouveia.weebly.com](http://www.stevensgouveia.weebly.com).

Projecto financiado pelo CEEC Individual 2022.02527.CEECIND da Fundação da Ciência e Tecnologia, no Mind, Language and Action Group, Instituto de Filosofia, Universidade do Porto (Faculdade de Letras, Via Panorâmica s/n, P-4150-564 Porto, Portugal).

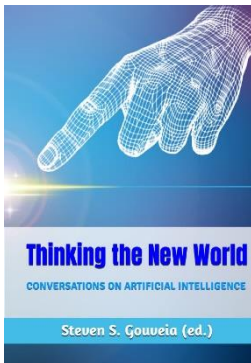
## Livros do Autor



2023



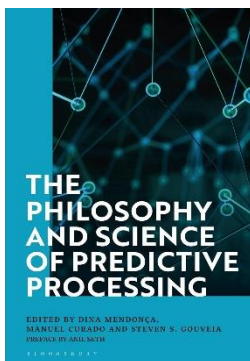
2022



2022



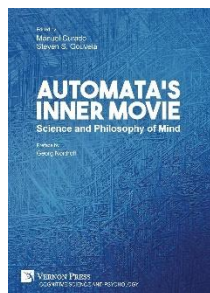
2020



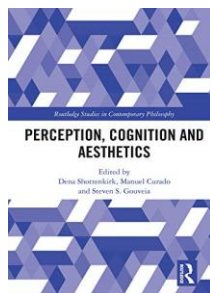
2020



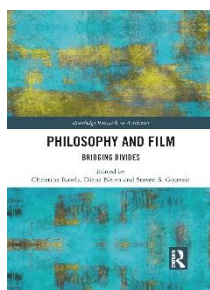
2020



2019



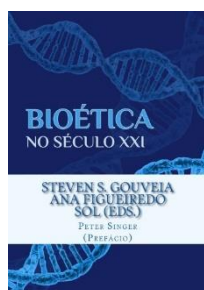
2019



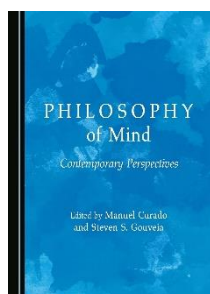
2019



2018



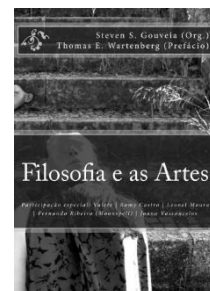
2018



2017



2017



2016