



ΠΡΟΜΕΘΕΥΣ
Journal of Philosophy
n. 33 maio - agosto de 2020
Dossiê Linguagem e Cognição



Editores: Marcus Souza, Marcos Silva & Maxwell Lima Filho

ENFRENTANDO O PROBLEMA DA CONSCIÊNCIA¹

David J. Chalmers
Australian National University/ New York University

Tradução: Pedro H. G. Muniz
Universidade Estadual do Ceará

RESUMO: Para progredir no problema da consciência, temos que confrontá-lo diretamente. Neste artigo, primeiro isolo a parte realmente difícil do problema, separando-a das partes mais tratáveis e esclarecendo por que é tão difícil explicá-la. Faço uma crítica de trabalhos recentes que usam métodos redutivos para tratar da consciência, e argumento que inevitavelmente esses métodos falham em lidar com a parte mais difícil do problema. Uma vez que essa falha é reconhecida, a porta para um subsequente progresso é aberta. Na segunda parte do artigo, argumento que se passarmos para um novo tipo de explicação não redutiva, podemos fornecer uma descrição naturalista da consciência. Apresento minha própria candidata para tal descrição: uma teoria não redutiva baseada em princípios de coerência estrutural e invariância organizacional e em uma visão de duplo aspecto da informação.

PALAVRAS-CHAVE: Consciência. Dualismo. Experiência. Explicação. Funcionalismo. Problema difícil. Informação. Materialismo. *Qualia*. Redução.

ABSTRACT: To make progress on the problem of consciousness, we have to confront it directly. In this paper, I first isolate the truly hard part of the problem, separating it from more tractable parts and giving an account of why it is so difficult to explain. I critique some recent work that uses reductive methods to address consciousness, and argue that such methods inevitably fail to come to grips with the hardest part of the problem. Once this failure is recognized, the door to further progress is opened. In the second half of the paper, I argue that if we move to a new kind of nonreductive explanation, a naturalistic account of consciousness can be given. I put forward my own candidate for such an account: a nonreductive theory based on principles of structural coherence and organizational invariance, and a double-aspect theory of information.

KEYWORDS: Consciousness. Dualism. Experience. Explanation. Functionalism. Hard problem. Information. Materialism. *Qualia*. reduction.

¹ Artigo publicado originalmente no *Journal of Consciousness Studies*, 2 (3), p. 200-19, 1995. Agradeço a Francis Crick, Peggy DesAutels, Matthew Elton, Liane Gabora, Christof Koch, Paul Rhodes, Gregg Rosenberg e Sharon Wahl por seus comentários.

1. Introdução

A consciência impõe os mais desconcertantes problemas à ciência da mente. Não há nada que conheçamos mais intimamente do que a experiência consciente, mas não há nada que seja mais difícil de explicar. Vários tipos de fenômenos mentais se submeteram à investigação científica nos últimos anos, mas a consciência resistiu obstinadamente. Muitos tentaram explicá-la, mas as explicações sempre ficam aquém do objetivo. Alguns foram levados a supor que o problema é intratável, e que nenhuma boa explicação pode ser dada.

Para progredir no problema da consciência, temos que confrontá-lo diretamente. Neste artigo, primeiro isolo a parte realmente difícil do problema, separando-a das partes mais tratáveis e esclarecendo por que é tão difícil explicá-la. Faço uma crítica de trabalhos recentes que usam métodos redutivos para tratar da consciência e argumento que inevitavelmente esses métodos falham em lidar com a parte mais difícil do problema. Uma vez que essa falha é reconhecida, a porta para um subsequente progresso é aberta. Na segunda parte do artigo, argumento que se passarmos para um novo tipo de explicação não redutiva, podemos fornecer uma descrição naturalista da consciência. Apresento minha própria candidata para tal descrição: uma teoria não redutiva baseada em princípios de coerência estrutural e invariância organizacional e em uma visão de duplo aspecto da informação.

2. Os problemas fáceis e o problema difícil

Não há apenas um problema da consciência. “Consciência” é um termo ambíguo, que se refere a muitos fenômenos diferentes. Cada um desses fenômenos precisa ser explicado, mas alguns são mais fáceis de explicar que outros. No começo, é útil dividir os problemas associados da consciência em problemas “difíceis” e “fáceis”. Os problemas da consciência fáceis são os que parecem diretamente suscetíveis aos métodos padrão da ciência cognitiva, através dos quais um fenômeno é explicado em termos de mecanismos computacionais ou neurais. Os problemas difíceis são os que parecem resistir a tais métodos.

Os problemas da consciência fáceis incluem os de se explicar os seguintes fenômenos:

- A habilidade de discriminar, categorizar e reagir a estímulos do ambiente;
- A integração de informações por um sistema cognitivo;
- A relatabilidade de estados mentais (*the reportability of mental states*)²;
- A habilidade de um sistema de acessar seus próprios estados internos;
- O foco da atenção;
- O controle deliberado do comportamento;
- A diferença entre vigília e sono.

Todos esses fenômenos estão associados à noção de consciência. Por exemplo, às vezes dizemos que um estado mental é consciente quando ele é verbalmente relatável (*reportable*), ou quando é acessível internamente. Às vezes se diz que um sistema está consciente de alguma informação quando ele tem a habilidade de reagir com base nessa informação ou, de maneira mais forte, quando ele lida com essa informação, ou quando pode integrar essa informação e explorá-la no sofisticado controle do comportamento. Às vezes dizemos que uma ação é consciente precisamente quando ela é deliberada. Com frequência dizemos que um organismo está consciente como outro jeito de dizer que ele está acordado.

Não existe um real problema no que diz respeito a se *esses* fenômenos podem ser explicados cientificamente. Todos eles são diretamente vulneráveis a uma explicação em termos de mecanismos computacionais ou neurais. Para se explicar o acesso e a relatabilidade, por exemplo, precisamos apenas especificar o mecanismo através do qual a informação sobre estados internos é recuperada e disponibilizada para um relato verbal. Para se explicar a integração de informações, precisamos apenas exibir os mecanismos pelos quais as informações são reunidas e exploradas por processos posteriores. Para uma descrição do sono e da vigília, uma descrição neurofisiológica

² A noção de *reportability* tem difícil tradução para o português. Seu equivalente direto é “relatabilidade”, embora esta não seja uma palavra muito usual. O termo está relacionado a *report*, que pode ser traduzido como “relato” ou, em certos contextos, “relatório”. O verbo “*to report*” equivale em português a “relatar”. Quando falamos de *reportability* em filosofia da mente, referimo-nos à habilidade que seres conscientes têm de relatar ou comunicar seus próprios estados mentais ou, mais especificamente, do fato de que estados mentais são passíveis de serem relatados por tais seres. Ainda que a palavra “relatabilidade” possa causar estranhamento, preferimos utilizá-la aqui por dois motivos principais. Primeiro, trata-se de um termo técnico recorrente em investigações sobre a consciência. Mas também porque, ao usar uma única palavra em vez de uma expressão como “habilidade de relatar” ou algo similar, não nos vemos obrigados a impor adaptações demasiadamente extensas ao discurso original do autor (Nota do Tradutor [N. do T.]).

apropriada dos processos responsáveis pelo comportamento contrastante do organismo nesses estados é suficiente. Em cada caso, um modelo cognitivo ou neurofisiológico apropriado pode claramente realizar o trabalho explicativo.

Se a consciência se resumisse a esses fenômenos, ela não seria um problema muito grande. Embora ainda não tenhamos nada próximo a uma explicação completa desses fenômenos, temos uma ideia clara de como poderíamos proceder para explicá-los. É por isso que eu os chamo de problemas fáceis. É claro, “fácil” é um termo relativo. Obter os detalhes de maneira correta provavelmente vai levar um século ou dois de um trabalho empírico difícil. Ainda assim, sobram motivos para se acreditar que os métodos da ciência cognitiva e da neurociência serão bem-sucedidos.

O problema da consciência, que é realmente difícil, é o problema da experiência. Quando pensamos e percebemos, há um trabalho de processamento de informação, mas também há um aspecto subjetivo. Como formulou Nagel (1974), há *algo que é como ser* um organismo consciente. Esse aspecto subjetivo é a experiência. Quando vemos, por exemplo, *experienciamos* sensações visuais: a qualidade sentida da vermelhidão, a experiência de escuro e brilhante, a qualidade de profundidade em um campo visual. Outras experiências acompanham a percepção em diferentes modalidades: o som de um clarinete, o cheiro de bolinhas de naftalina. Há também sensações corporais, variando de dores a orgasmos; imagens mentais que são conjuradas internamente; a qualidade sentida da emoção e a experiência de um fluxo de pensamento consciente. O que unifica todos esses estados é que há algo que é como estar neles. Todos eles são estados de experiência.

É inegável que alguns organismos são sujeitos de experiência. Mas a questão de como esses sistemas são objetos de experiência é desconcertante. Por que é que, quando nossos sistemas cognitivos se envolvem no processamento visual e auditivo de informações, temos experiência visual ou auditiva (a qualidade do azul profundo, a sensação do dó médio)? Como podemos explicar por que há algo que é como ter uma imagem mental ou ter a experiência de uma emoção? É amplamente aceito que a experiência surge a partir de uma base física, mas não temos uma boa explicação de porquê e como ela surge assim. Por que o processamento físico deveria dar origem a uma rica vida interior? Objetivamente, parece irracional que ele deva fazer isso e, mesmo assim, ele o faz.

Se algum problema se qualifica como o problema da consciência, é esse. Nesse sentido central de “consciência”, um organismo é consciente se há algo que é como ser

esse organismo, e um estado mental é consciente se há algo que é estar nesse estado. Às vezes, termos como “consciência fenomênica” e “*qualia*” também são usados aqui, mas acho mais natural falar em “experiência consciente” ou simplesmente “experiência”. Outra maneira útil de evitar confusão (usada, por exemplo, por Newell 1990; Chalmers 1996) é reservar o termo “consciência” para os fenômenos da experiência, usando o termo menos carregado “estado de ciência” (*awareness*)³ para os fenômenos mais diretos descritos anteriormente. Se tal convenção fosse amplamente adotada, a comunicação seria muito mais fácil. Como as coisas estão hoje, frequentemente, aqueles que falam sobre “consciência” estão conversando sem se entender.

A ambiguidade do termo “consciência” é frequentemente explorada tanto por filósofos quanto por cientistas que escrevem sobre o assunto. É comum vermos um artigo sobre consciência começar com uma invocação do mistério da consciência, notando a estranha intangibilidade e inefabilidade da subjetividade e preocupando-se que até então não tenhamos nenhuma teoria do fenômeno. Aqui, o tópico é claramente o problema difícil – o problema da experiência. Na segunda metade do artigo, o tom se torna mais otimista, e a teoria da consciência do próprio autor é delineada. Ao examiná-la, essa teoria acaba sendo uma teoria de um dos fenômenos mais diretos – o da relatibilidade, do acesso introspectivo, ou qualquer outro. No fim das contas, o autor declara que a consciência acabou se revelando tratável, mas, ao leitor, resta o sentimento de ter sido vítima de uma propaganda enganosa (*bait-and-switch*). O problema difícil permanece intocado.

3. Explicação funcional

Por que os problemas fáceis são fáceis, e por que o problema difícil é difícil? Os problemas fáceis são fáceis precisamente porque eles concernem à explicação de *habilidades* cognitivas e *funções*. Para explicar uma função cognitiva, precisamos apenas especificar um mecanismo que pode desempenhar a função. Os métodos da

3 Em português, é comum traduzir ambos os termos “*consciousness*” e “*awareness*” como “consciência”. No entanto, em alguns contextos, como no presente texto, a palavra “*awareness*” é usada com o sentido de “estado de ciência”, assim como poderíamos usar “*aware*” para dizer que alguém está *ciente* de algo, sendo este algo, geralmente, algum tipo de informação (muitas vezes, de natureza fenomenológica) que pode ser usada para guiar o comportamento. Chalmers faz uso da distinção entre *consciousness* e *awareness* e desse sentido específico de “estado de ciência” em outros escritos seus, como em seu livro *The conscious mind: in search of a fundamental theory*. Philosophy of Mind Series. Oxford University Press, 1996 (N. do T.).

ciência cognitiva são bem adequados para esse tipo de explicação e, assim, são bem adequados para os problemas fáceis da consciência. Em contraste, o problema difícil é difícil precisamente porque não é um problema sobre o desempenho de funções. O problema persiste mesmo quando o desempenho de todas as funções relevantes é explicado. (Aqui “função” não é usada no sentido teleológico estreito de algo que um sistema é projetado para fazer, mas no sentido mais amplo de qualquer papel causal que um sistema pode desempenhar na produção de comportamento).

Explicar a relatabilidade, por exemplo, é apenas explicar como um sistema poderia executar a função de produzir relatos sobre estados internos. Para explicar o acesso interno, precisamos explicar como um sistema poderia ser afetado de maneira apropriada por seus estados internos e usar a informação sobre esses estados no direcionamento de processos posteriores. Para explicar a integração e o controle, precisamos explicar como os processos centrais de um sistema podem reunir os conteúdos de informação e usá-los na facilitação de vários comportamentos. Todos esses são problemas que dizem respeito à explicação de funções.

Como explicamos o desempenho de uma função? Especificando um mecanismo que desempenha a função. Aqui, modelagens neurofisiológicas e cognitivas são perfeitas para a tarefa. Se quisermos uma explicação detalhada de baixo nível, podemos especificar o mecanismo neural que é responsável pela função. Se quisermos uma explicação mais abstrata, podemos especificar um mecanismo em termos computacionais. De uma forma ou de outra, o resultado será uma explicação completa e satisfatória. Uma vez que tenhamos especificado o mecanismo neural ou computacional que desempenha a função do relato verbal, por exemplo, a maior parte do nosso trabalho de explicação da relatabilidade chega ao fim.

De certa forma, o ponto é trivial. É um fato *conceitual* sobre esses fenômenos que sua explicação envolve apenas a explicação de várias funções, uma vez que os fenômenos são *definíveis funcionalmente*. A relatabilidade estar instanciada em um sistema *significa* apenas que o sistema é capaz de gerar relatos verbais de informações internas. Um sistema estar acordado significa apenas que ele está adequadamente receptivo a informações do ambiente e que é capaz de usá-las para guiar o comportamento de maneira apropriada. Para entender que esse tipo de coisa é um fato conceitual, note que alguém que diz “você explicou o desempenho da função do relato verbal, mas não explicou a relatabilidade” está cometendo um erro conceitual trivial sobre a relatabilidade. Para explicar a relatabilidade, a única coisa *possivelmente*

necessária é uma explicação de como a função relevante é executada. O mesmo vale para os outros fenômenos em questão.

Nas ciências de nível mais alto, as explicações redutivas funcionam exatamente dessa maneira. Para explicar o gene, por exemplo, foi preciso especificar o mecanismo que armazena e transmite informações hereditárias de uma geração para a outra. Acontece que o DNA desempenha essa função, e uma vez que explicamos como a função é desempenhada, explicamos o gene. Para explicar a vida, precisamos, em última instância, explicar como um sistema pode se reproduzir, adaptar-se ao seu ambiente, metabolizar e assim por diante. Todas essas são questões sobre o desempenho de funções e, portanto, são adequadas para explicações redutivas. O mesmo vale para a maioria dos problemas na ciência cognitiva. Para explicar o aprendizado, precisamos explicar como as capacidades comportamentais de um sistema são modificadas à luz de informações do ambiente, e como novas informações podem ser usadas para adaptar as ações de um sistema a seu ambiente. Se mostrarmos como um mecanismo neural ou computacional faz esse trabalho, explicamos o aprendizado. Podemos dizer o mesmo para outros fenômenos cognitivos, como a percepção, a memória e a linguagem. Às vezes as funções relevantes precisam ser caracterizadas de maneira bastante sutil, mas é claro que, na medida em que a ciência cognitiva explica esses fenômenos, ela o faz explicando o desempenho de funções.

Quando se trata de experiência consciente, esse tipo de explicação falha. O que faz com que o problema difícil seja difícil e quase único é que ele está para *além* dos problemas relacionados ao desempenho de funções. Para entender isso, observe que mesmo que tenhamos explicado o desempenho de todas as funções cognitivas e comportamentais próximas à experiência (a discriminação perceptiva, a categorização, o acesso interno, o relato verbal), ainda pode restar uma questão adicional não respondida: *Por que o desempenho dessas funções é acompanhado da experiência?* Uma simples explicação das funções deixa essa questão em aberto.

Não existe uma pergunta adicional análoga na explicação dos genes, da vida ou do aprendizado. Se alguém diz: “Estou vendo que você explicou como o DNA armazena e transmite informações hereditárias de uma geração à outra, mas você não explicou como é um *gene*”, essa pessoa está cometendo um erro conceitual. Ser um gene significa apenas ser uma entidade que desempenha a função relevante de armazenamento e transmissão. Mas se uma pessoa disser “Estou vendo que você explicou como informações são discriminadas, integradas e relatadas, mas você não

explicou como elas são *experienciadas*”, ela não está cometendo um erro conceitual. Essa é uma questão adicional não trivial.

Essa questão adicional é a questão chave no problema da consciência. Por que todo esse processamento de informações não ocorre “no escuro”, livre de qualquer sensação interior? Por que acontece que, quando formas de onda eletromagnéticas colidem com uma retina e são discriminadas e categorizadas por um sistema visual, essa discriminação e categorização são experienciadas como uma sensação de vermelho vivo? Sabemos que a experiência consciente *realmente* surge quando essas funções são desempenhadas, mas o fato mesmo de ela surgir é o mistério central. Há uma *lacuna explicativa* (um termo que devemos a Levine 1983) entre as funções e a experiência, e precisamos de uma ponte explicativa para atravessá-la. Uma mera descrição das funções permanece em um lado da lacuna, de forma que os materiais para a ponte devem ser encontrados em outro lugar.

Isso não quer dizer que a experiência não *tenha* função. Talvez se revele que ela desempenha um papel cognitivo importante. Mas para qualquer papel que ela possa desempenhar, haverá mais na explicação da experiência do que uma simples explicação da função. Talvez se revele mesmo que, ao longo da explicação de uma função, seremos levados ao *insight* chave que nos permitirá explicar a experiência. No entanto, se isso acontecer, a descoberta será uma recompensa explicativa *extra*. Não há nenhuma uma função cognitiva tal que possamos dizer antecipadamente que a explicação dessa função explicará *automaticamente* a experiência.

Para explicar a experiência precisamos de uma nova abordagem. Os métodos explicativos usuais da ciência cognitiva e da neurociência não são suficientes. Esses métodos foram desenvolvidos precisamente para explicar o desempenho de funções cognitivas e eles são bons para isso. Mas, na forma em que eles são, são equipados *apenas* para explicar o desempenho de funções. Quando se trata do problema difícil, a abordagem padrão não tem nada a dizer.

4. Alguns estudos de caso

Nos últimos anos, vários trabalhos abordaram os problemas da consciência no âmbito da ciência cognitiva e da neurociência. Isso pode sugerir que a análise acima é falha. Na verdade, porém, um exame atento do trabalho relevante só lhe dá mais suporte. Quando investigamos exatamente em quais aspectos da consciência esses estudos se

focam e quais aspectos eles acabam explicando, descobrimos que o alvo final da explicação é sempre um dos problemas fáceis. Vou ilustrar isso com dois exemplos representativos.

O primeiro é o da “teoria neurobiológica da consciência”, esboçado por Crick e Koch (1990; ver também Crick, 1994). Essa teoria está centrada em certas oscilações neurais de 35 a 75 hertz no córtex cerebral. Crick e Koch levantam a hipótese de que essas oscilações são a base da consciência. Isso ocorre em parte porque as oscilações parecem estar correlacionadas com o estado de ciência (*awareness*) em várias modalidades diferentes (nos sistemas visual e olfativo, por exemplo) e também porque elas sugerem um mecanismo pelo qual a *ligação* (*binding*) de conteúdos de informação pode ser alcançada. A ligação é o processo pelo qual parcelas de informação sobre uma única entidade representadas separadamente são reunidas para serem usadas em um processamento posterior, como quando informações sobre a cor e a forma de um objeto percebido são integradas a partir de rotas visuais separadas. Seguindo outros (por exemplo, Eckhorn et al., 1988), Crick e Koch levantam a hipótese de que a ligação pode ser alcançada pelas oscilações sincronizadas de grupos neuronais que representam o conteúdo relevante. Quando duas informações devem ser unidas, os grupos neurais relevantes oscilam com a mesma frequência e fase.

Os detalhes de como essa ligação pode ser alcançada ainda são pouco compreendidos, mas suponha que eles possam ser resolvidos. O que a teoria resultante poderia explicar? Claramente, ela poderia explicar a ligação dos conteúdos de informação e poderia, talvez, gerar uma descrição mais geral da integração da informação no cérebro. Crick e Koch também sugerem que essas oscilações ativam os mecanismos da memória de trabalho, de forma que no futuro pode haver uma descrição dessa e, talvez, de outras formas de memória. Em algum momento, a teoria poderia levar a uma descrição geral de como a informação percebida é vinculada e armazenada na memória para uso em processamento posterior.

Essa teoria seria valiosa, mas não nos diria nada sobre por que os conteúdos relevantes são experienciados. Crick e Koch sugerem que essas oscilações são os *correlatos* neurais da experiência. Essa afirmação é discutível – a ligação também não ocorre no processamento inconsciente de informação? Mas mesmo que ela seja aceita, a questão *explicativa* permanece: por que as oscilações dão origem à experiência? A única base para uma conexão explicativa é o papel que eles desempenham na ligação e no armazenamento, mas a questão de por que a ligação e o armazenamento devem eles

próprios ser acompanhados pela experiência nunca é abordada. Se não soubermos por que a ligação e o armazenamento devem dar origem à experiência, contar uma história sobre as oscilações não pode nos ajudar. Por outro lado, se *soubéssemos* por que a ligação e o armazenamento deram origem à experiência, os detalhes neurofisiológicos seriam apenas a cereja do bolo. A teoria de Crick e Koch ganha seu terreno *assumindo* uma conexão entre a ligação e a experiência e, portanto, não pode fazer nada para explicar esse *link*.

Não creio que, em última instância, Crick e Koch estejam abordando o problema difícil, embora alguns tenham lhes interpretado como afirmando isso. Uma entrevista publicada com Koch fornece uma declaração clara das limitações das ambições da teoria.

Bem, esqueçamos primeiro os aspectos realmente difíceis, como sentimentos subjetivos, pois eles podem não ter uma solução científica. O estado subjetivo de brincar, da dor, do prazer, de ver azul, de cheirar uma rosa – parece haver um grande salto entre o nível materialista, de explicar moléculas e neurônios, e o nível subjetivo. Concentremo-nos em coisas que são mais fáceis de estudar, como o estado de ciência visual (*visual awareness*). Você está falando comigo agora, mas não está olhando para mim, está olhando para o cappuccino e, portanto, está ciente dele. Você pode dizer: “É um copo e há um pouco de líquido nele”. Se eu o der a você, você vai mover seu braço e pegá-lo – você vai responder de uma maneira significativa. É isso que eu chamo de estado de ciência (*awareness*) (What is Consciousness, *Discover*, novembro de 1992, p. 96).

O segundo exemplo é uma abordagem no nível da psicologia cognitiva. Trata-se da teoria da consciência do espaço de trabalho global de Bernard Baars, apresentada em seu livro *A cognitive theory of consciousness*. De acordo com essa teoria, os conteúdos da consciência estão contidos em um *espaço de trabalho global*, um processador central usado para mediar a comunicação entre uma série de processadores não conscientes especializados. Quando esses processadores especializados precisam transmitir informação para o resto do sistema, eles o fazem enviando essa informação para o espaço de trabalho, que atua como uma espécie de quadro-negro comum para o resto do sistema, acessível a todos os outros processadores.

Baars usa esse modelo para abordar muitos aspectos da cognição humana e explicar diversos contrastes entre o funcionamento cognitivo consciente e inconsciente. Contudo, em última análise, trata-se de uma teoria da *acessibilidade cognitiva*, explicando como certos conteúdos de informação são amplamente acessíveis dentro de

um sistema, bem como uma teoria da integração informacional e da relatabilidade. A teoria se mostra promissora como uma teoria do estado de ciência (*awareness*), o correlato funcional da experiência consciente, mas ela não oferece uma explicação da experiência em si.

Pode-se supor que, de acordo com essa teoria, os conteúdos da experiência sejam precisamente os conteúdos do espaço de trabalho. Mas mesmo que seja assim, nada interno à teoria *explica* por que as informações no espaço de trabalho global são experienciadas. O melhor que a teoria pode fazer é dizer que a informação é experienciada porque é *acessível globalmente*. Mas agora a questão surge de uma forma diferente: por que a acessibilidade global deveria dar origem à experiência consciente? Como sempre, essa questão de ponte não é respondida.

Quase todos os trabalhos que adotam uma abordagem cognitiva ou neurocientífica da consciência nos últimos anos poderiam ser submetidos a uma crítica semelhante. O modelo do “Darwinismo Neural” de Edelman (1989), por exemplo, aborda questões sobre o estado de ciência perceptual (*perceptual awareness*) e o autoconceito (*self-concept*), mas não diz nada sobre por que também deveria haver experiência. O modelo dos “esboços múltiplos” de Dennett (1991) é amplamente orientado a uma explicação da relatabilidade de certos conteúdos mentais. A teoria do “nível intermediário” de Jackendoff (1988) fornece uma descrição de alguns processos computacionais subjacentes à consciência, mas Jackendoff enfatiza que a questão de como eles “se projetam” na experiência consciente permanece misteriosa.

Pesquisadores que usam esses métodos geralmente não são explícitos sobre suas atitudes em relação ao problema da experiência consciente, embora às vezes eles adotem uma posição clara. Mesmo entre os que são claros, as atitudes diferem amplamente. Ao posicionar esse tipo de trabalho em relação ao problema da experiência, várias estratégias diferentes estão disponíveis. Seria útil se essas escolhas estratégicas fossem explicitadas com mais frequência.

A primeira estratégia é simplesmente *explicar outra coisa*. Alguns pesquisadores são explícitos ao dizer que o problema da experiência é difícil demais no momento e que talvez esteja até mesmo fora do domínio da ciência. Esses pesquisadores escolhem abordar um dos problemas mais tratáveis, tais como o da relatabilidade ou o do autoconceito. Embora eu tenha chamado esses problemas de os problemas “fáceis”, eles estão entre os mais interessantes problemas não resolvidos na ciência cognitiva, de forma que esse trabalho certamente vale a pena. O pior que se pode dizer dessa escolha

é que, no contexto da pesquisa sobre a consciência, ela é relativamente pouco ambiciosa e, às vezes, o trabalho pode ser mal interpretado.

A segunda escolha é seguir uma linha mais difícil e *negar o fenômeno*. (Variações dessa abordagem são empregadas por Allport, 1988; Dennett 1991; Wilkes, 1988.) De acordo com essa linha, uma vez que tenhamos explicado funções como a acessibilidade, a relatabilidade e outras, não há mais nenhum fenômeno chamado “experiência” para explicar. Alguns negam explicitamente o fenômeno, defendendo, por exemplo, que o que não é verificável externamente não pode ser real. Outros alcançam o mesmo efeito admitindo que a experiência existe, mas somente se equipararmos “experiência” a algo como a capacidade de discriminar e de relatar. Essas abordagens levam a uma teoria mais simples, mas, em última instância, são insatisfatórias. A experiência é o aspecto mais central e manifesto de nossas vidas mentais e, de fato, talvez seja o *explanandum* chave na ciência da mente. Por causa desse status de *explanandum*, a experiência não pode ser descartada como o espírito vital quando surge uma nova teoria. Antes, é o fato central que qualquer teoria da consciência deve explicar. Uma teoria que nega o fenômeno “resolve” o problema se esquivando da questão.

Em uma terceira opção, alguns pesquisadores *afirmam estar explicando a experiência* no sentido pleno. Estes pesquisadores (diferentemente dos anteriores) desejam levar a experiência muito a sério: eles expõem sua teoria ou seu modelo funcional e afirmam que ele explica toda a qualidade subjetiva da experiência (e.g., Flohr, 1992; Humphrey, 1992). Contudo, geralmente se passa depressa pela etapa relevante na explicação e, frequentemente, ela acaba parecendo mágica. Depois que alguns detalhes sobre o processamento de informações são dados, a experiência entra em cena de repente, mas continua obscuro *como* esses processos devem, de repente, dar origem à experiência. Talvez simplesmente se tome como certo que eles dão, mas então temos uma explicação incompleta e uma versão da quinta estratégia abaixo.

Uma quarta abordagem mais promissora apela a esses métodos para *explicar a estrutura da experiência*. Por exemplo, pode-se argumentar que uma descrição das discriminações feitas pelo sistema visual pode explicar as relações estruturais entre diferentes experiências de cores, bem como a estrutura geométrica do campo visual (veja, por exemplo, Clark 1992; Hardin 1992). Em geral, certos fatos sobre estruturas encontradas no processamento correspondem e, pode-se argumentar, explicam fatos sobre a estrutura da experiência. Essa estratégia é plausível, mas limitada. Na melhor das hipóteses, ela toma como certa a existência da experiência e esclarece alguns fatos

sobre sua estrutura, fornecendo uma espécie de explicação não redutiva de seus aspectos estruturais (falarei mais sobre isso adiante). Isso é útil para muitos propósitos, mas não nos diz nada sobre por que, antes de mais nada, a experiência deve existir.

Uma quinta estratégia que é razoável é *isolar o substrato da experiência*. Afinal, quase todo mundo admite que a experiência *surge* de um modo ou de outro a partir dos processos cerebrais, e faz sentido identificar o tipo de processo do qual ela surge. Crick e Koch apresentam seu trabalho como isolando o correlato neural da consciência, por exemplo, e Edelman (1989) e Jackendoff (1988) fornecem argumentos relacionados. A justificação desses argumentos requer uma análise teórica cuidadosa, especialmente porque a experiência não é diretamente observável em contextos experimentais, mas, quando aplicada criteriosamente, essa estratégia pode lançar uma luz indireta sobre o problema da experiência. Ainda assim, a estratégia é claramente incompleta. Para uma teoria satisfatória, precisamos saber mais do que *quais* processos geram a experiência: precisamos de uma descrição do porquê e do como. Uma teoria da consciência completa deve construir uma ponte explicativa.

5. O ingrediente extra

Vimos que existem razões sistemáticas de por que os métodos usuais da ciência cognitiva e da neurociência falham em explicar a experiência consciente. Eles são simplesmente o tipo errado de método: nada do que eles nos dão pode produzir uma explicação. Para explicar a experiência consciente, precisamos de um *ingrediente extra* na explicação. Isso cria um desafio para aqueles que levam a sério o problema difícil da consciência: qual é o seu ingrediente extra e por que *isso* deveria explicar a experiência consciente?

Não faltam ingredientes extras. Alguns propõem uma injeção de caos e dinâmica não linear. Alguns acham que a chave está no processamento não-algorítmico. Alguns apelam para futuras descobertas em neurofisiologia. Alguns supõem que a chave para o mistério estará no nível da mecânica quântica. É fácil entender por que todas essas sugestões são apresentadas. Nenhum dos antigos métodos funciona, então a solução deve estar em *algo* novo. Infelizmente, todas essas sugestões sofrem dos mesmos antigos problemas.

O processamento não-algorítmico, por exemplo, é apresentado por Penrose (1989; 1994) por causa do papel que pode desempenhar no processo de *insight*

matemático consciente. Os argumentos sobre a matemática são controversos, mas, mesmo que eles sejam bem-sucedidos e que uma explicação do processamento não-algorítmico no cérebro humano seja apresentada, ela ainda será apenas uma descrição das *funções* envolvidas no raciocínio matemático e outros similares. A questão é deixada sem resposta tanto para um processo não-algorítmico quanto para um processo algorítmico: por que esse processo deve dar origem à experiência? Ao responder a *essa* pergunta, não há nenhum papel especial para o processamento não-algorítmico.

O mesmo vale para a dinâmica não-linear e caótica. Elas podem fornecer uma nova explicação da dinâmica do funcionamento cognitivo, bem diferente da que é dada pelos métodos padrão na ciência cognitiva. Mas da dinâmica só se obtém mais dinâmica. A questão sobre a experiência aqui continua misteriosa como sempre. O ponto é ainda mais claro para novas descobertas na neurofisiologia. Essas novas descobertas podem nos ajudar a progredir significativamente em nossa compreensão do funcionamento cerebral, mas para qualquer processo neural que isolarmos, surgirá sempre a mesma questão. É difícil imaginar o que um proponente da nova neurofisiologia espera que aconteça além da explicação de mais funções cognitivas. Não é como se de repente fôssemos descobrir um brilho fenomênico dentro de um neurônio!

Talvez o “ingrediente extra” mais popular de todos seja a mecânica quântica (e.g., Hameroff, 1994). A atratividade das teorias quânticas da consciência pode ter origem em uma Lei de Minimização do Mistério: a consciência é misteriosa e a mecânica quântica é misteriosa, então talvez os dois mistérios tenham uma fonte comum. No entanto, as teorias quânticas da consciência sofrem das mesmas dificuldades que as teorias neurais ou computacionais. Fenômenos quânticos têm propriedades funcionais notáveis, tais como o não-determinismo e a não-localidade. É natural especular que essas propriedades possam desempenhar algum papel na explicação de funções cognitivas, tais como a escolha aleatória e a integração de informações, e essa hipótese não pode ser descartada *a priori*. Mas quando se trata da explicação da experiência, os processos quânticos estão no mesmo barco que quaisquer outros. A questão de por que esses processos deveriam dar origem à experiência permanece completamente sem resposta.

(Um atrativo especial das teorias quânticas é o fato de que, em algumas interpretações da mecânica quântica, a consciência desempenha um papel ativo no “colapso” da função de onda quântica. Tais interpretações são controversas, mas, de qualquer forma, elas não oferecem nenhuma esperança de *explicar* a consciência em

termos de processos quânticos. Antes, essas teorias *assumem* a existência da consciência e a utilizam na explicação dos processos quânticos. Na melhor das hipóteses, essas teorias nos dizem algo sobre um papel físico que a consciência pode desempenhar. Elas não nos dizem nada sobre como ela surge).

Em última análise, a mesma crítica se aplica a *qualquer* explicação puramente física da consciência. Para qualquer processo físico que especificarmos, haverá uma pergunta sem resposta: Por que esse processo deveria dar origem à experiência? Dado qualquer processo como ele, é conceitualmente coerente que ele poderia ser instanciado na ausência da experiência. Segue-se que uma mera explicação do processo físico não nos dirá por que a experiência surge. A emergência da experiência vai para além do que pode ser derivado de uma teoria física.

Uma explicação puramente física é adequada para explicar *estruturas* físicas, explicando estruturas macroscópicas em termos de constituintes microestruturais detalhados. E ela fornece uma explicação satisfatória do desempenho de *funções*, descrevendo-as em termos dos mecanismos físicos que as desempenham. Isso acontece porque uma descrição física pode *implicar (entail)* os fatos sobre estruturas e funções: uma vez que os detalhes internos da descrição física são dados, as propriedades estruturais e funcionais se seguem como uma consequência automática. Mas a estrutura e a dinâmica dos processos físicos produzem apenas mais estrutura e dinâmica, então não podemos esperar que esses processos expliquem algo mais do que estruturas e funções. Os fatos sobre a experiência não podem ser uma consequência automática de nenhuma descrição física, já que é conceitualmente coerente que qualquer processo possa existir sem experiência. A experiência pode *surgir* do físico, mas não é *implicada* do físico (*is not entailed by the physical*).

A moral de tudo isso é que *você não pode explicar a experiência consciente de uma maneira econômica*. É um fato notável que métodos redutivos (métodos que explicam um fenômeno de alto nível inteiramente em termos de processos físicos mais básicos) funcionam bem em muitos domínios. Em certo sentido, nós *podemos* explicar a maioria dos fenômenos biológicos e cognitivos de uma maneira econômica, na medida em que esses fenômenos são vistos como consequências automáticas de processos mais fundamentais. Seria maravilhoso se métodos redutivos também pudessem explicar a experiência – por um longo tempo eu esperei que eles pudessem. Infelizmente, há razões sistemáticas de por que esses métodos falham. Métodos redutivos são bem-sucedidos na maioria dos domínios porque o que precisa ser explicado neles são

estruturas e funções, e esse é o tipo de coisa que uma descrição física pode implicar. Quando se trata de um problema que está para além da explicação de estruturas e funções, esses métodos são impotentes.

Isso pode parecer remanescente do argumento vitalista de que nenhuma descrição física poderia explicar a vida, mas os casos não são análogos. O que motivou o ceticismo vitalista foi a dúvida de que os mecanismos físicos pudessem desempenhar as muitas funções notáveis associadas à vida, tais como o comportamento adaptativo complexo e a reprodução. O argumento conceitual de que a explicação de funções é o que é preciso foi aceito implicitamente. Mas por carecerem de um conhecimento detalhado sobre os mecanismos bioquímicos, os vitalistas duvidaram que algum processo físico pudesse desempenhar as funções relevantes e apresentaram a hipótese do espírito vital como uma explicação alternativa. Depois que se revelou que os processos físicos podiam desempenhar esse papel, as dúvidas dos vitalistas desapareceram.

Com a experiência, por outro lado, a explicação física das funções não vem ao caso. Em vez disso, a chave aqui é a questão *conceitual* de que a explicação das funções não é suficiente para a explicação da experiência. Essa questão conceitual básica não é algo que uma investigação neurocientífica adicional afetará. De maneira semelhante, a experiência não é análoga ao *élan vital*. O espírito vital foi apresentado como um postulado explicativo para explicar as funções relevantes e, portanto, poderia ser descartado quando essas funções fossem explicadas sem ele. A experiência não é um postulado explicativo, mas sim um *explanandum* por si só, e, portanto, não é um candidato a esse tipo de eliminação.

É tentador notar que vários tipos de fenômenos intrigantes acabaram se mostrando explicáveis em termos físicos. Mas todos eles eram problemas sobre o comportamento observável de objetos físicos, os quais se resumiam a problemas na explicação de estruturas e funções. Por esse motivo, esses fenômenos sempre foram o tipo de coisa que uma descrição física *poderia* explicar, mesmo que em alguns momentos houvesse boas razões para suspeitar que essa explicação não se concretizaria. A indução tentadora a partir desses casos falha no caso da consciência, a qual não é um problema sobre estruturas físicas e funções. O problema da consciência é desconcertante de uma maneira totalmente diferente. Uma análise do problema nos mostra que a experiência consciente não é exatamente o tipo de coisa que uma descrição totalmente redutiva poderia explicar de forma bem-sucedida.

6. Explicação não redutiva

Neste ponto, alguns se sentem tentados a desistir, defendendo que nunca teremos uma teoria da experiência consciente. McGinn (1989), por exemplo, argumenta que o problema é difícil demais para nossas mentes limitadas: somos “cognitivamente fechados” no que diz respeito ao fenômeno. Outros argumentaram que a experiência consciente se encontra totalmente fora do domínio da teoria científica.

Eu acho que esse pessimismo é prematuro. Este não é o momento de desistir, é o momento no qual as coisas ficam interessantes. Quando métodos de explicação simples são descartados, precisamos investigar as alternativas. Uma vez que a explicação redutiva falha, a explicação não redutiva é a escolha natural.

Embora um número considerável de fenômenos tenha se mostrado inteiramente explicável em termos de entidades mais simples que eles próprios, isso não é universal. Na física, por vezes acontece que uma entidade deve ser considerada como *fundamental*. Entidades fundamentais não são explicadas em termos de nada mais simples. Em vez disso, elas são consideradas básicas, e é apresentada uma teoria de como elas se relacionam com tudo o que resta no mundo. Por exemplo, no século XIX, descobriu-se que processos eletromagnéticos não podiam ser totalmente explicados em termos dos processos inteiramente mecânicos aos quais as teorias físicas anteriores apelavam, então Maxwell e outros introduziram a carga eletromagnética e as forças eletromagnéticas como novos componentes fundamentais de uma teoria física. Para explicar o eletromagnetismo, a ontologia da física precisou ser expandida. Novas propriedades e leis básicas foram necessárias para fornecer uma descrição satisfatória dos fenômenos.

Outros elementos que a teoria física considera como fundamentais incluem a massa e o espaço-tempo. Não se fez tentativa nenhuma de explicar esses elementos em termos de algo mais simples. Mas isso não exclui a possibilidade de uma teoria da massa ou do espaço-tempo. Existe uma teoria complexa de como esses elementos se inter-relacionam e das leis básicas nas quais eles se inserem. Esses princípios básicos são usados para explicar muitos fenômenos familiares que concernem a massa, o espaço e o tempo em um nível superior.

Eu sugiro que uma teoria da consciência deva considerar a experiência como fundamental. Nós sabemos que uma teoria da consciência requer a adição de *algo* fundamental à nossa ontologia, uma vez que tudo na teoria física é compatível com a

ausência de consciência. Podemos adicionar um elemento não-físico inteiramente novo, do qual a experiência pode ser derivada, mas é difícil entender como esse elemento seria. É mais provável que consideremos a própria experiência como um elemento fundamental do mundo, ao lado da massa, da carga e do espaço-tempo. Se considerarmos a experiência como fundamental, podemos começar o trabalho de construir uma teoria da experiência.

Onde existe uma propriedade fundamental, existem leis fundamentais. Uma teoria não redutiva da experiência acrescentará novos princípios ao mobiliário das leis básicas da natureza. Em última análise, esses princípios básicos sustentarão a carga explicativa em uma teoria da consciência. Assim como explicamos fenômenos de alto nível familiares que envolvem massa em termos de princípios mais básicos envolvendo massa e outras entidades, também podemos explicar fenômenos familiares que envolvem a experiência em termos de princípios mais básicos que envolvem a experiência e outras entidades.

Em particular, uma teoria não redutiva da experiência especificará princípios básicos que nos dizem como a experiência depende de elementos físicos do mundo. Esses princípios *psicofísicos* não interferem nas leis físicas, uma vez que parece que as leis físicas já formam um sistema fechado. Antes, eles serão um complemento para uma teoria física. Uma teoria física fornece uma teoria de processos físicos, e uma teoria psicofísica nos diz como esses processos dão origem à experiência. Sabemos que a experiência depende de processos físicos, mas também sabemos que essa dependência não pode ser derivada unicamente de leis físicas. Os novos princípios básicos postulados por uma teoria não redutiva nos dão o ingrediente extra do qual precisamos para construir uma ponte explicativa.

É claro que, ao considerar a experiência como fundamental, há um sentido em que essa abordagem não nos diz por que, antes de mais nada, existe experiência. Mas acontece o mesmo com qualquer teoria fundamental. Nada na física nos diz por que, antes de mais nada, existe matéria. Mas não levamos isso em consideração contra teorias da matéria. Certos elementos do mundo precisam ser considerados fundamentais por qualquer teoria científica. Uma teoria da matéria ainda pode explicar vários tipos de fatos sobre a matéria, mostrando como eles são consequências das leis básicas. O mesmo vale para uma teoria da experiência.

Essa posição é qualificada como um tipo de dualismo, já que postula propriedades básicas que estão para além das propriedades evocadas pela física. Mas é

uma versão inocente do dualismo, inteiramente compatível com a visão científica do mundo. Nada nessa abordagem contradiz o que quer que seja na teoria física: precisamos somente adicionar mais princípios *que fazem a ponte (bridging principles)* para explicar como a experiência surge a partir de processos físicos. Não há nada de particularmente espiritual ou místico nessa teoria – sua forma geral é semelhante à de uma teoria física, com algumas poucas entidades fundamentais conectadas por leis fundamentais. Ela expande um pouco a ontologia, é claro, mas Maxwell fez a mesma coisa. De fato, a estrutura geral dessa posição é inteiramente naturalista, defendendo que, em última análise, o universo se reduz a uma rede de entidades básicas que obedecem a leis simples, e defendendo que, em última análise, pode haver uma teoria da consciência construída em termos de tais leis. Se essa posição deve ter um nome, uma boa escolha poderia ser *dualismo naturalista*.

Se essa visão estiver correta, então, de certa forma, uma teoria da consciência terá mais em comum com uma teoria na física do que com uma teoria na biologia. As teorias biológicas não envolvem princípios que sejam fundamentais dessa maneira. Por isso, a teoria biológica tem em si uma certa complexidade e desordem. Mas, na física, na medida em que as teorias lidam com princípios fundamentais, elas aspiram à simplicidade e à elegância. As leis fundamentais da natureza fazem parte do mobiliário básico do mundo, e as teorias físicas nos dizem que esse mobiliário básico é notavelmente simples. Se uma teoria da consciência também envolve princípios fundamentais, então devemos esperar o mesmo. Os princípios de simplicidade, elegância e até beleza que impulsionam a busca dos físicos por uma teoria fundamental também se aplicarão a uma teoria da consciência.

(Uma nota técnica: alguns filósofos argumentam que, embora haja uma lacuna *conceitual* entre processos físicos e a experiência, não é preciso que haja uma lacuna metafísica, de modo que, em certo sentido, a experiência ainda pode ser física (e.g., Hill, 1991, Levine, 1983, Loar, 1990). Geralmente essa linha de argumentação é apoiada por um apelo à noção de necessidade *a posteriori* (Kripke, 1980). Contudo, eu acho que essa posição se apoia em uma má compreensão da necessidade *a posteriori*, ou então requer uma forma inteiramente nova de necessidade na qual não temos motivo algum para crer. Veja Chalmers (1996) (também Jackson, 1994 e Lewis, 1994) para detalhes. De qualquer forma, essa posição ainda admite uma lacuna *explicativa* entre processos físicos e a experiência. Por exemplo, os princípios que conectam o físico e o experiencial não são derivados das leis da física, de forma que tais princípios devem ser

considerados *explicativamente* fundamentais. Dessa forma, mesmo nesse tipo de posicionamento, a estrutura explicativa de uma teoria da consciência será, em muitos aspectos, similar à que eu descrevi).

7. Contornos de uma teoria da consciência

Não é cedo demais para começar a trabalhar em uma teoria. Já estamos em condições de entender certos fatos chave sobre a relação entre processos físicos e a experiência, bem como sobre as regularidades que os conectam. Uma vez que a explicação redutiva é posta de lado, podemos colocar esses fatos na mesa para que eles possam desempenhar seu devido papel de peças iniciais em uma teoria não redutiva da consciência, bem como de restrições das leis básicas que constituem uma teoria final.

Há um problema óbvio que afeta o desenvolvimento de uma teoria da consciência, que é a escassez de dados objetivos. A experiência consciente não é diretamente observável em um contexto experimental, de forma que não podemos gerar dados sobre a relação entre processos físicos e experiência à vontade. Ainda assim, todos nós temos acesso a uma rica fonte de dados em nossos próprios casos. Muitas regularidades importantes entre experiência e processamento podem ser inferidas a partir de considerações sobre a própria experiência de cada um. Também há boas fontes indiretas de dados de casos observáveis, como quando nos baseamos no relato verbal de um sujeito como uma indicação de experiência. Esses métodos têm suas limitações, mas temos dados mais do que suficientes para começar a fazer uma teoria deslançar.

Uma análise filosófica também é útil para fazer um bom negócio com os dados que temos. Esse tipo de análise pode fornecer diversos princípios relacionando a consciência e a cognição, impondo, assim, uma forte restrição à forma de uma teoria final. Como veremos, o método de se utilizar experimentos de pensamento também pode fornecer recompensas significativas. Finalmente, o fato de estarmos procurando por uma teoria *fundamental* significa que podemos apelar a restrições não empíricas tais como simplicidade, homogeneidade e assim por diante no desenvolvimento de uma teoria. Devemos procurar sistematizar as informações que temos, estendê-las o máximo possível através de uma análise cuidadosa e, em seguida, inferir a teoria mais simples possível que explique os dados e continue sendo uma candidata plausível para fazer parte do mobiliário fundamental do mundo.

Tais teorias sempre guardarão um elemento de especulação que não está presente em outras teorias científicas, devido à impossibilidade de testes experimentais intersubjetivos conclusivos. Ainda assim, certamente, podemos construir teorias que sejam compatíveis com os dados que temos, bem como avaliá-las comparando-as umas com as outras. Mesmo na ausência de observação intersubjetiva, existem inúmeros critérios disponíveis para a avaliação de tais teorias: simplicidade, coerência interna, coerência com teorias em outros domínios, capacidade de reproduzir as propriedades da experiência que são familiares do nosso próprio caso e até uma acomodação geral aos ditames do senso comum. Talvez restem indeterminações significativas mesmo quando todas essas restrições forem aplicadas, mas podemos pelo menos desenvolver candidatos plausíveis. Só seremos capazes de avaliar as teorias candidatas, quando elas forem desenvolvidas.

Uma teoria não redutiva da consciência consistirá em diversos *princípios psicofísicos* – princípios conectando as propriedades dos processos físicos às propriedades da experiência. Podemos pensar nesses princípios como encapsulando a maneira pela qual a experiência surge do físico. Em última instância, eles deveriam nos dizer que tipo de sistema físico terá experiências associadas e, para os sistemas que as têm, eles deveriam nos dizer que tipo de propriedades físicas são relevantes para a emergência da experiência, bem como, para qualquer sistema físico, exatamente que tipo de experiência deveríamos esperar que ele produza. Essa é uma tarefa difícil, mas não há razão para não começarmos.

A seguir, apresento meus próprios candidatos aos princípios psicofísicos que poderiam entrar em uma teoria da consciência. Os dois primeiros são *princípios não básicos* – conexões sistemáticas entre processamento e experiência em um nível relativamente alto. Esses princípios podem desempenhar um papel significativo no desenvolvimento e restrição de uma teoria da consciência, mas não são introduzidos em um nível suficientemente fundamental para se qualificarem como leis verdadeiramente básicas. O último princípio é o meu candidato a um *princípio básico* que poderia formar a pedra angular de uma teoria da consciência fundamental. Esse último princípio é particularmente especulativo, mas é o tipo de especulação que é necessária para que um dia tenhamos uma teoria da consciência satisfatória. Só posso apresentar esses princípios aqui de maneira breve. Argumento a favor deles muito mais longamente em Chalmers (1996).

1. *O princípio da coerência estrutural*

Este é um princípio de coerência entre a *estrutura da consciência* e a *estrutura do estado de ciência*. Lembre-se de que “estado de ciência” (*awareness*) foi usado anteriormente para se referir aos vários fenômenos funcionais que estão associados à consciência. Agora estou usando-o para me referir a um processo um pouco mais específico nos fundamentos cognitivos da experiência. Em particular, os conteúdos do estado de ciência devem ser entendidos como aqueles conteúdos de informação que são acessíveis aos sistemas centrais e que são mobilizados de maneira generalizada no controle do comportamento. Em resumo, podemos pensar sobre o estado de ciência como *disponibilidade direta para o controle global*. Para uma primeira aproximação, os conteúdos do estado de ciência são os conteúdos que são diretamente acessíveis e potencialmente relatáveis, pelo menos em um sistema que usa linguagem.

O estado de ciência é uma noção puramente funcional, mas, ainda assim, está intimamente ligada à experiência consciente. Em casos familiares, onde quer que encontremos consciência, encontramos estado de ciência. Onde quer que haja experiência consciente, há alguma informação correspondente no sistema cognitivo que está disponível no controle do comportamento e disponível para um relato verbal. Inversamente, parece que sempre que há uma informação disponível para um relato e para o controle global, há uma experiência consciente correspondente. Assim, há uma correspondência direta entre consciência e estado de ciência.

A correspondência pode ser levada mais longe. É um fato central sobre a experiência que ela tem uma estrutura complexa. O campo visual tem uma geometria complexa, por exemplo. Há também relações de semelhança e de diferença entre experiências, bem como relações em coisas como intensidade relativa. A experiência de todo sujeito pode ser caracterizada e decomposta pelo menos parcialmente em termos dessas propriedades estruturais: relações de similaridade e diferença, localização percebida, intensidade relativa, estrutura geométrica e assim por diante. Também é um fato central que, para cada um desses elementos estruturais, há um elemento correspondente na estrutura de processamento de informação do estado de ciência.

Tomemos as sensações de cor como exemplo. Para cada distinção entre experiências de cor, há uma distinção correspondente no processamento. As diferentes cores fenomênicas que experienciamos formam um espaço tridimensional complexo que varia em matiz, saturação e intensidade. As propriedades desse espaço podem ser

recuperadas a partir de considerações do processamento de informações: o exame dos sistemas visuais mostra que as formas de onda da luz são discriminadas e analisadas em três eixos diferentes, e é essa informação tridimensional que é relevante para o processamento posterior. Portanto, a estrutura tridimensional do espaço fenomênico de cores corresponde diretamente à estrutura tridimensional do estado de ciência visual (*visual awareness*). Isso é exatamente o que esperaríamos. Afinal, toda distinção de cores corresponde a alguma informação relatável e, portanto, a uma distinção que é representada na estrutura do processamento.

De uma maneira mais direta, a estrutura geométrica do campo visual é diretamente refletida em uma estrutura que pode ser recuperada a partir do processamento visual. Todas as relações geométricas correspondem a algo que pode ser relatado e, portanto, são representadas cognitivamente. Se contássemos apenas a história do processamento de informações no sistema visual e cognitivo de um agente, não poderíamos observar *diretamente* as experiências visuais desse agente. Ainda assim, poderíamos inferir as propriedades estruturais dessas experiências.

Em geral, qualquer informação que seja experienciada conscientemente também será representada cognitivamente. A estrutura de granularidade fina (*fine-grained*) do campo visual corresponderá a alguma estrutura de granularidade fina no processamento visual. O mesmo vale para experiências em outras modalidades e até para experiências não sensoriais. Imagens mentais internas têm propriedades geométricas que são representadas no processamento. Até emoções têm propriedades estruturais, tais como intensidade relativa, que correspondem diretamente a uma propriedade estrutural do processamento – onde há maior intensidade, encontramos um efeito maior em processos posteriores. Em geral, justamente porque as propriedades estruturais da experiência são acessíveis e relatáveis, elas serão representadas diretamente na estrutura do estado de ciência (*awareness*).

É esse isomorfismo entre as estruturas da consciência e do estado de ciência que constitui o princípio da coerência estrutural. Esse princípio reflete o fato central de que, embora os processos cognitivos não impliquem conceitualmente fatos sobre a experiência consciente, a consciência e a cognição não flutuam livres uma da outra, mas coexistem de maneira íntima.

Esse princípio tem seus limites. Ele nos permite recuperar propriedades estruturais da experiência a partir das propriedades do processamento de informações, mas nem todas as propriedades da experiência são propriedades estruturais. Existem

propriedades da experiência, tais como a natureza intrínseca de uma sensação de vermelho, que não podem ser completamente capturadas em uma descrição estrutural. A própria inteligibilidade dos cenários de espectro invertido, nos quais as experiências de vermelho e verde são invertidas, mas todas as propriedades estruturais permanecem as mesmas, mostram que as propriedades estruturais restringem a experiência sem esgotá-la. Ainda assim, o próprio fato de que nos sentimos compelidos a deixar as propriedades estruturais inalteradas quando imaginamos experiências invertidas entre sistemas funcionalmente idênticos mostra o quão central é o princípio da coerência estrutural para a nossa concepção de nossas vidas mentais. Ele não é um princípio *logicamente* necessário, já que, afinal, podemos imaginar todo o processamento de informações ocorrendo sem nenhuma experiência. Ainda assim, é uma restrição forte e familiar à conexão psicofísica.

O princípio da coerência estrutural permite um tipo muito útil de explicação indireta da experiência em termos de processos físicos. Por exemplo, podemos usar fatos sobre o processamento neural de informações visuais para explicar indiretamente a estrutura do espaço de cores. Os fatos sobre o processamento neural podem implicar e explicar a estrutura do estado de ciência. Se tomarmos o princípio da coerência como certo, a estrutura da experiência também será explicada. A investigação empírica pode até nos levar a entender melhor a estrutura do estado de ciência de um morcego, esclarecendo indiretamente a incômoda questão de Nagel de como é ser um morcego. Esse princípio fornece uma interpretação natural de grande parte do trabalho existente sobre a explicação da consciência (por exemplo, Clark, 1992 e Hardin, 1992, sobre cores; e Akins, 1993, sobre morcegos), embora com frequência se apele a ele de maneira não explícita. Ele é tão familiar, que é tomado como certo por quase todos, e é uma base central na explicação cognitiva da consciência.

A coerência entre consciência e estado de ciência também permite uma interpretação natural do que tem sido feito na neurociência com o objetivo de isolar o *substrato* (ou *correlato neural*) da consciência. Várias hipóteses específicas já foram apresentadas. Crick e Koch (1990), por exemplo, sugerem que oscilações de 40 Hz podem ser o correlato neural da consciência, enquanto Libet (1993) sugere que a atividade neural estendida temporalmente é central. Se aceitarmos o princípio da coerência, o correlato físico mais *direto* da consciência é o estado de ciência: o processo pelo qual a informação é diretamente disponibilizada para o controle global. As diferentes hipóteses específicas podem ser interpretadas como sugestões empíricas

sobre como se pode alcançar o estado de ciência. Por exemplo, Crick e Koch sugerem que as oscilações de 40 Hz são a porta de acesso (*gateway*) pela qual as informações são integradas na memória de trabalho e, assim, disponibilizadas para processos posteriores. Da mesma forma, é natural supor que a atividade estendida temporalmente de Libet seja relevante precisamente porque somente esse tipo de atividade atinge a disponibilidade global. O mesmo se aplica a outros correlatos sugeridos, como o “espaço de trabalho global” de Baars (1988), as “representações de alta qualidade” de Farah (1994) e as “entradas seletoras (*selector inputs*) para os sistemas de ação” de Shallice (1972). Todos eles podem ser vistos como hipóteses sobre os *mecanismos do estado de ciência*: os mecanismos que desempenham a função de disponibilizar a informação de maneira direta para o controle global.

Dada a coerência entre consciência e estado de ciência, segue-se que um mecanismo do estado de ciência será ele próprio um correlato da experiência consciente. A questão de exatamente *quais* mecanismos no cérebro governam a disponibilidade global é empírica – talvez haja muitos desses mecanismos. Mas se aceitarmos o princípio da coerência, temos motivos para acreditar que os processos que *explicam* o estado de ciência serão ao mesmo tempo parte da *base* da consciência.

2. O princípio da invariância organizacional

Este princípio afirma que quaisquer dois sistemas com a mesma *organização funcional* de baixa granularidade terão experiências qualitativamente idênticas. Se os padrões causais da organização neural fossem duplicados no silício, por exemplo, com um chip de silício para cada neurônio e os mesmos padrões de interação, então as mesmas experiências surgiriam. De acordo com esse princípio, o que importa para a emergência da experiência não é a composição física específica de um sistema, mas o padrão abstrato de interação causal entre seus componentes. Este princípio é controverso, é claro. Na opinião de alguns (e.g., Searle, 1980), a consciência está ligada a uma biologia específica, de modo que um isomorfo de um humano feito de silício não é necessariamente consciente. Contudo, acredito que o princípio possa receber um apoio significativo da análise de experimentos mentais.

Muito brevemente: suponha (a fins de uma *reductio ad absurdum*) que o princípio seja falso e que pudesse haver dois sistemas funcionalmente isomórficos com experiências diferentes. Talvez apenas um dos sistemas seja consciente, ou talvez

ambos o sejam, mas possuem experiências diferentes. Para fins de ilustração, digamos que um sistema é feito de neurônios e o outro de silício, e que um experiencia vermelho, enquanto o outro experiencia azul. Os dois sistemas têm a mesma organização, então podemos imaginar que um se transforme gradualmente no outro, talvez substituindo-se os neurônios, um por um, por chips de silício com a mesma função local. Assim, adquirimos um espectro de casos intermediários, cada um com a mesma organização, mas com uma composição física ligeiramente diferente e com experiências ligeiramente diferentes. Nesse espectro, deve haver dois sistemas, *A* e *B*, entre os quais substituímos menos de um décimo do sistema, mas cujas experiências diferem. Esses dois sistemas são fisicamente idênticos, com a exceção de que um pequeno circuito neural em *A* foi substituído por um circuito de silício em *B*.

O passo chave no experimento mental é identificar o circuito neural relevante em *A* e instalar ao lado dele um circuito de silício causalmente isomórfico, com um interruptor entre os dois. O que acontece quando ligamos o interruptor? Por hipótese, as experiências conscientes do sistema mudarão – digamos, para fins de ilustração, que do vermelho para o azul. Isso decorre do fato de que o sistema após a mudança é essencialmente uma versão de *B*, enquanto antes da mudança é apenas *A*.

Mas considerando as suposições, não há como o sistema *perceber* as mudanças! Sua organização causal permanece constante, de forma que todos os seus estados funcionais e disposições comportamentais permanecem fixos. Até onde ele sabe, nada de anormal aconteceu. Não há espaço para o pensamento: “Hum! Algo estranho acabou de acontecer!” De maneira geral, a estrutura de qualquer pensamento desse tipo deve ser refletida no processamento, mas aqui a estrutura do processamento permanece constante. Se houvesse tal pensamento, ele deveria flutuar totalmente livre do sistema e seria totalmente incapaz de afetar o processamento posterior. (Se ele afetasse o processamento posterior, os sistemas seriam funcionalmente distintos, ao contrário da hipótese). Podemos até ligar e desligar o interruptor várias vezes, de forma que as experiências de vermelho e de azul dançam em alternância diante do “olho interno” do sistema. Segundo a hipótese, o sistema nunca pode perceber esses “*qualia* dançantes”.

Considero que essa é uma *reductio* da suposição original. É um fato central sobre a experiência, muito familiar do nosso próprio caso, que sempre que as experiências mudam significativamente e estamos prestando atenção, podemos perceber a mudança. Se não fosse esse o caso, seríamos levados à possibilidade cética de que nossas experiências estão dançando diante dos nossos olhos o tempo todo. Essa hipótese

tem o mesmo *status* da possibilidade de o mundo ter sido criado há cinco minutos: talvez seja logicamente coerente, mas não é plausível. Dada a suposição extremamente plausível de que mudanças na experiência correspondem a mudanças no processamento, somos levados à conclusão de que a hipótese original é impossível e que quaisquer dois sistemas funcionalmente isomórficos devem ter o mesmo tipo de experiências. Para colocar em termos técnicos, as hipóteses filosóficas dos “*qualia* ausentes” e dos “*qualia* invertidos”, embora logicamente possíveis, são empírica e nomologicamente impossíveis.

(Alguns podem desconfiar que um isomorfo de silício de um sistema neural possa ser impossível por razões técnicas. Essa questão está aberta. O princípio da invariância diz apenas que *se* um isomorfo for possível, então ele terá o mesmo tipo de experiência consciente).

Há mais a ser dito aqui, mas isso já dá um primeiro sabor da ideia. Mais uma vez, esse experimento mental se baseia em fatos familiares sobre a coerência entre a consciência e o processamento cognitivo para chegar a uma conclusão forte sobre a relação entre a estrutura física e a experiência. Se o argumento vingar, sabemos que as únicas propriedades físicas diretamente relevantes para o surgimento da experiência são as propriedades *organizacionais*. Isso atua como mais uma forte restrição a uma teoria da consciência.

3. *A teoria do duplo aspecto da informação*

Os dois princípios precedentes eram princípios *não básicos*. Elas envolvem noções de alto nível, tais como “estado de ciência” e “organização” e, portanto, estão no nível errado para constituir as leis fundamentais em uma teoria da consciência. Ainda assim, eles agem como fortes restrições. O que ainda é preciso são princípios *básicos* que se ajustam a essas restrições e, em última análise, podem explicá-las.

O princípio básico que eu sugiro envolve de maneira central a noção de *informação*. Entendo informação mais ou menos no sentido de Shannon (1948). Onde há informação, há *estados de informação* integrados a um *espaço de informação*. Um espaço de informação possui uma estrutura básica de relações de *diferença* entre seus elementos, caracterizando as formas em que diferentes elementos em um espaço são similares ou diferentes, possivelmente de maneiras complexas. Um espaço de informação é um objeto abstrato, mas seguindo Shannon, podemos ver as informações

como *fisicamente incorporadas* quando há um espaço de estados físicos distintos, cujas diferenças podem ser transmitidas por alguma rota causal. Os próprios estados transmitidos podem ser vistos como constituindo um espaço de informação. Retomando uma frase de Bateson (1972), informação física é uma *diferença que faz a diferença*.

O princípio do duplo aspecto tem origem na observação de que há um isomorfismo direto entre certos espaços de informação fisicamente incorporados e certos espaços de informação *fenomênicos* (ou experienciais). A partir do mesmo tipo de observação que entrou no princípio da coerência estrutural, podemos notar que as diferenças entre estados fenomênicos têm uma estrutura que corresponde diretamente às diferenças integradas a processos físicos, particularmente às diferenças que fazem a diferença em certas rotas causais envolvidas no controle e disponibilidade global. Ou seja, podemos encontrar o *mesmo* espaço de informação abstrato integrado ao processamento físico e à experiência consciente.

Isso leva a uma hipótese natural: que a informação (ou, pelo menos, algumas informações) possui dois aspectos básicos: um aspecto físico e um aspecto fenomênico. Essa hipótese tem o status de um princípio básico que pode subjazer a e explicar a emergência da experiência a partir do físico. A experiência surge em virtude de seu *status* como um aspecto da informação, quando o outro aspecto se encontra incorporado no processamento físico.

Esse princípio é apoiado por várias considerações, as quais eu só posso esboçar brevemente aqui. Primeiro, a consideração do tipo de mudanças físicas que corresponde a mudanças na experiência consciente sugere que tais mudanças são sempre relevantes em virtude de seu papel na constituição de *mudanças informacionais* – diferenças dentro de um espaço abstrato de estados que são divididos precisamente de acordo com suas diferenças causais ao longo de certas rotas causais. Segundo, se o princípio da invariância organizacional é válido, então precisamos encontrar alguma propriedade *organizacional* fundamental à qual a experiência esteja ligada, e a informação é uma propriedade organizacional *par excellence*. Terceiro, esse princípio nos dá esperança de explicar o princípio da coerência estrutural em termos da estrutura presente dentro dos espaços de informação. Quarto, a análise da explicação cognitiva de nossos *juízos* e *afirmações* sobre a experiência consciente (juízos que são funcionalmente explicáveis, mas, ainda assim, profundamente ligados à própria experiência) sugere que a explicação envolve de maneira central os estados de informação integrados ao processamento cognitivo. Segue-se que uma teoria baseada na informação permite uma coerência

profunda entre a explicação da experiência e a explicação de nossos juízos e afirmações sobre ela.

Wheeler (1990) sugeriu que a informação é fundamental para a física do universo. De acordo com essa doutrina “*it from bit*”⁴, as leis da física podem ser construídas em termos de informação, postulando estados diferentes que dão origem a efeitos diferentes sem realmente dizer o que *são* esses estados. É apenas a posição deles em um espaço de informação que conta. Se for assim, então a informação é uma candidata natural a também desempenhar um papel em uma teoria fundamental da consciência. Somos levados a uma concepção do mundo na qual a informação é verdadeiramente fundamental e na qual ela tem dois aspectos básicos, correspondendo às características físicas e fenomênicas do mundo.

É claro, o princípio do duplo aspecto é extremamente especulativo e também é subdeterminado, deixando várias questões-chave sem resposta. Uma questão óbvia é se *toda* informação tem um aspecto fenomênico. Uma possibilidade é que precisemos de uma restrição a mais à teoria fundamental, indicando exatamente que *tipo* de informação tem um aspecto fenomênico. A outra possibilidade é que não haja tal restrição. Se não houver, a experiência é muito mais difundida do que poderíamos ter pensado, já que a informação está em toda parte. A princípio isso é contraintuitivo, mas, ao refletir a respeito, acho que a posição ganha uma certa plausibilidade e elegância. Onde existe processamento simples de informações, existe experiência simples, e onde existe processamento complexo de informações, existe experiência complexa. Um rato tem uma estrutura de processamento de informações mais simples que um humano e possui uma experiência correspondentemente mais simples. Talvez um termostato, uma estrutura de processamento de informações maximamente simples, possa ter uma experiência maximamente simples? De fato, se a experiência é realmente uma propriedade fundamental, seria surpreendente que ela surgisse apenas de vez em quando: a maior parte das propriedades fundamentais são distribuídas de maneira mais uniforme. De qualquer forma, essa é realmente uma questão aberta, mas acredito que a posição não seja tão implausível quanto frequentemente se pensa.

Quando um elo fundamental entre informação e experiência é colocado sobre a mesa, abre-se a porta a uma especulação metafísica mais grandiosa a respeito da

4 Termo de difícil tradução para o português, em geral “*it from bit*” aparece em textos especializados no original, em inglês. Em tradução livre, “coisas a partir de *bits*” (N. do T.).

natureza do mundo. Por exemplo, é comum notar que a física caracteriza suas entidades básicas apenas *extrinsecamente*, em termos de suas relações com outras entidades, que são elas próprias caracterizadas extrinsecamente, e assim por diante. A natureza intrínseca das entidades físicas é deixada de lado. Alguns argumentam que tais propriedades intrínsecas não existem, mas então nos resta um mundo que é puro fluxo causal (um puro fluxo de informações), sem propriedades com as quais a causalidade pode se relacionar. Se admitirmos que existam propriedades intrínsecas, uma especulação natural, dado o que foi dito acima, é que as propriedades intrínsecas do físico (as propriedades com as quais, em última instância, a causalidade se relaciona) são elas próprias propriedades fenomênicas. Podemos dizer que propriedades fenomênicas são o aspecto interno da informação. Isso poderia responder uma preocupação com relação à relevância causal da experiência – uma preocupação natural, dada uma teoria na qual o domínio físico é causalmente fechado e na qual a experiência é suplementar à física. A visão informacional nos permite entender como a experiência pode ter um tipo sutil de relevância causal em virtude de seu *status* como a natureza intrínseca do físico. Provavelmente, é melhor ignorar essa especulação metafísica quando o propósito é de desenvolver uma teoria científica. Mas ao abordar algumas questões filosóficas, ela é bastante sugestiva.

8. Conclusão

A teoria que apresentei é especulativa, mas é uma candidata. Suspeito que os princípios de coerência estrutural e invariância organizacional sejam bases de qualquer teoria satisfatória da consciência. O *status* da teoria do duplo aspecto da informação é menos certo. De fato, agora ela é mais uma ideia do que uma teoria. Para que se tenha esperança de um eventual sucesso explicativo, ela deverá ser especificada de maneira mais completa e desenvolvida em uma forma mais robusta. Ainda assim, a reflexão sobre o que é plausível e implausível, sobre onde ela funciona e onde ela falha, só pode levar a uma teoria melhor.

A maioria das teorias da consciência existentes ou nega o fenômeno, ou explica outra coisa, ou eleva o problema a um mistério eterno. Espero ter demonstrado que é possível progredir no problema, até mesmo quando o levamos a sério. Para avançar ainda mais, precisaremos de mais investigação, teorias mais refinadas e análises mais

cuidadasas. O problema difícil é um problema difícil, mas não há razão para acreditar que ele permanecerá perpetuamente sem solução.

Leitura adicional

Os problemas da consciência foram amplamente discutidos na literatura filosófica recente. Para um esclarecimento conceitual dos vários problemas da consciência, veja Block (1995), Nelkin (1993) e Tye (1995). Os que deram ênfase nas dificuldades de explicar a experiência em termos físicos incluem Hodgson (1988), Jackson (1982), Levine (1983), Lockwood (1989), McGinn (1989), Nagel (1974), Seager (1991), Searle (1991), Strawson (1994) e Velmans (1991), dentre outros. Os que adotam uma abordagem reducionista incluem Churchland (1995), Clark (1992), Dennett (1991), Dretske (1995), Kirk (1994), Rosenthal (1996) e Tye (1995). Não houve muitas tentativas de construir teorias não reducionistas detalhadas na literatura, mas veja Hodgson (1988) e Lockwood (1989) para algumas ideias nessa direção. Duas excelentes coleções de artigos recentes sobre a consciência são Block, Flanagan e Güzeldere (1996) e Metzinger (1995).

Bibliografia

- Akins, K. What is it like to be boring and myopic?. In: B. Dahlbom (Ed.). *Dennett and his Critics*. Oxford: Blackwell, 1993.
- Allport, A. What concept of consciousness? In: A. Marcel & E. Bisiach (Eds.). *Consciousness in contemporary science*. Oxford: Oxford University Press, 1988.
- Baars, B.J. *A cognitive theory of consciousness*. Cambridge: Cambridge University Press, 1988.
- Bateson, G. *Steps to an ecology of mind*. Chandler Publishing, 1972.
- Block, N. On a confusion about the function of consciousness. *Behavioral and Brain Sciences*, 18 (2), p. 227-247, 1995.
- Block, N.; Flanagan, O. & Güzeldere, G. (Eds.). *The nature of consciousness: philosophical and scientific debates*. Cambridge, MA: MIT Press, 1996.
- Chalmers, D. J. *The conscious mind*. New York: Oxford University Press, 1996.
- Churchland, P. M. *The engine of reason, the seat of the soul: a philosophical journey into the brain*. Cambridge, MA: MIT Press, 1995.

- Clark, A. *Sensory qualities*. Oxford: Oxford University Press, 1992.
- Crick, F.; Koch, C. Toward a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2, p. 263-275, 1990.
- Crick, F. *The astonishing hypothesis: the scientific search for the soul*. New York: Scribners, 1994.
- Dennett, D. C. *Consciousness explained*. Boston: Little, Brown, 1991.
- Dretske, F. I. *Naturalizing the mind*. Cambridge, MA: MIT Press, 1995.
- Edelman, G. *The remembered present: a biological theory of consciousness*. New York: Basic Books, 1989.
- Farah, M. J. Visual perception and visual awareness after brain damage: a tutorial overview. In: C. Umiltà & M. Moscovitch (Eds.). *Consciousness and unconscious information processing: attention and performance*, 15. Cambridge, MA: MIT Press, 1994.
- Flohr, H. Qualia and brain processes. In: A. Beckermann, H. Flohr & J. Kim (Eds.). *Emergence or reduction? Prospects for nonreductive physicalism*. Berlin: De Gruyter, 1992.
- Hameroff, S. R. Quantum coherence in microtubules: a neural basis for emergent consciousness? *Journal of Consciousness Studies*, 1, p. 91-118, 1994.
- Hardin, C. L. Physiology, phenomenology, and Spinoza's true colors. In: A. Beckermann, H. Flohr & J. Kim (Eds.), *Emergence or reduction? Prospects for nonreductive physicalism*. Berlin: De Gruyter, 1992.
- Hill, C.S. *Sensations: a defense of type materialism*. Cambridge: Cambridge University Press, 1991.
- Hodgson, D. *The mind matters: consciousness and choice in a quantum world*. Oxford: Oxford University Press, 1988.
- Humphrey, N. *A history of the mind*. New York: Simon and Schuster, 1992.
- Jackendoff, R. *Consciousness and the computational mind*. Cambridge, MA: MIT Press, 1987.
- Jackson, F. Epiphenomenal qualia. *Philosophical Quarterly*, 32, p. 127-36, 1982.
- _____. Finding the mind in the natural world. In: R. Casati, B. Smith & S. White (Eds.). *Philosophy and the cognitive sciences*. Vienna: Hölder-Pichler-Tempsky, 1994.
- Kirk, R. *Raw feeling: a philosophical account of the essence of consciousness*. Oxford: Oxford University Press, 1994.
- Kripke, S. *Naming and necessity*. Cambridge, MA: Harvard University Press, 1980.

- Levine, J. Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, 64, p. 354-61, 1983.
- Lewis, D. Reduction of mind. In: S. Guttenplan (Ed.). *A companion to the philosophy of mind*. Oxford: Blackwell, 1994.
- Libet, B. The neural time factor in conscious and unconscious events. In: G. R. Block & J. Marsh (Eds.). *Experimental and theoretical studies of consciousness* (Ciba Foundation Symposium, 174). Chichester: John Wiley and Sons, 1993.
- Loar, B. Phenomenal states. *Philosophical Perspectives*, 4, p. 81-108, 1990.
- Lockwood, M. *Mind, brain, and the quantum*. Oxford: Blackwell, 1989.
- McGinn, C. Can we solve the mind-body problem? *Mind*, 98, p. 349-66, 1989.
- Metzinger, T. *Conscious experience*. Paderborn: Schöningh, 1995.
- Nagel, T. What is it like to be a bat? *Philosophical Review*, 4, p. 435-50, 1974.
- Nelkin, N. What is consciousness? *Philosophy of Science*, 60, p. 419-34, 1993.
- Newell, A. *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press, 1990.
- Penrose, R. *The emperor's new mind*. Oxford: Oxford University Press, 1989.
- _____. *Shadows of the mind*. Oxford: Oxford University Press, 1994.
- Rosenthal, D.M. A theory of consciousness. In: N. Block, O. Flanagan, and G. Güzeldere (Eds.). *The nature of consciousness*. Cambridge, MA: MIT Press, 1996.
- Seager, W.E. *Metaphysics of consciousness*. London: Routledge, 1991.
- Searle, J.R. Minds, brains and programs. *Behavioral and Brain Sciences*, 3, p. 417-57, 1980.
- _____. *The rediscovery of the mind*. Cambridge, MA: MIT Press, 1992.
- Shallice, T. Dual functions of consciousness. *Psychological Review*, 79, p. 383-93, 1972.
- Shannon, C.E. A mathematical theory of communication. *Bell Systems Technical Journal*, 27, p. 379-423, 1948.
- Strawson, G. *Mental reality*. Cambridge, MA: MIT Press, 1994.
- Tye, M. *Ten problems of consciousness*. Cambridge, MA: MIT Press, 1995.
- Velmans, M. Is human information-processing conscious? *Behavioral and Brain Sciences*, 14, p. 651-69, 1991.
- Wheeler, J.A. Information, physics, quantum: the search for links. In: W. Zurek (Ed.). *Complexity, entropy, and the physics of information*. Redwood City, CA: Addison-Wesley, 1990.

Wilkes, K.V. Yishi, Duh, Um and consciousness. In: A. Marcel & E. Bisiach (Eds.). *consciousness in contemporary science*. Oxford: Oxford University Press, 1988.