

# Enfrentando o Problema da Consciência\*

David John Chalmers

Programa de Filosofia

Escola de Pesquisa de Ciências Sociais

Universidade Nacional da Austrália

Tradução de Ari Raynsford ([www.ariraynsford.com.br](http://www.ariraynsford.com.br))

Revisão de Darcy Brega e Giovanni Barontini

## 1. Introdução

A consciência apresenta os problemas mais desconcertantes na 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. Todos os tipos de fenômenos mentais têm sido submetidos à investigação científica nos últimos anos, mas a consciência tem resistido teimosamente. Muitos tentaram explicá-la, mas as explicações sempre parecem ficar aquém do objetivo. Alguns estudiosos são levados a supor que o problema é intratável e que nenhuma boa explicação pode ser dada.

Para progredir no problema da consciência, precisamos confrontá-lo diretamente. Neste artigo, primeiro isolo a parte verdadeiramente complexa do problema, separando-a das partes mais tratáveis, e apresento um esclarecimento por que ela é tão difícil de ser explicada. Critico alguns trabalhos recentes que utilizam métodos reducionistas para abordar a consciência e argumento que esses métodos inevitavelmente falham ao lidar com a parte mais complexa do problema. Uma vez reconhecida essa falha, a porta para um progresso maior é aberta. Na

---

\* Este artigo foi publicado no *Journal of Consciousness Studies*, 2(3):200-19, 1995. Agradecimentos a Francis Crick, Peggy DesAutels, Matthew Elton, Liane Gabora, Christof Koch, Paul Rhodes, Gregg Rosenberg e Sharon Wahl por seus comentários.

segunda metade do artigo, argumento que, se avançarmos para um novo tipo de explicação não reducionista, podemos dar uma explicação naturalista da consciência. Apresento minha própria proposta para tal explicação: uma teoria não reducionista baseada em princípios de coerência estrutural e invariância organizacional, e de uma visão de duplo aspecto da informação.

## 2. Os Problemas Fáceis e o Problema Difícil

Não existe apenas um problema da consciência. "Consciência" é um termo ambíguo, referindo-se a muitos fenômenos diferentes. Cada um desses fenômenos precisa ser explicado, mas alguns são mais fáceis de explicar do que outros. Inicialmente, é útil dividir os problemas associados à consciência em problemas "fáceis" e "difícil". Os problemas fáceis da consciência são aqueles que parecem ser diretamente suscetíveis aos métodos padrão da ciência cognitiva, pelos quais um fenômeno é explicado em termos de mecanismos computacionais ou neurais. O problema difícil é aquele que parece resistir a esses métodos.

Os problemas fáceis da consciência incluem os seguintes fenômenos:

- a capacidade de discriminar, categorizar e reagir a estímulos ambientais;
- a integração de informações por um sistema cognitivo;
- a relatibilidade de estados mentais;
- a capacidade de um sistema 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 diz-se que um estado mental é consciente quando é verbalmente relatável ou quando é internamente acessível. Às vezes, diz-se que um sistema está consciente de alguma informação quando tem a capacidade de reagir com base nessa informação ou, mais fortemente, quando atende a essa informação, ou quando pode integrá-la e explorá-la no controle de comportamento sofisticado. Outras vezes, dizemos que uma ação é consciente precisamente quando é deliberada. Frequentemente, dizemos que um organismo está consciente como outra forma de dizer que está desperto.

Não há dúvida de que esses fenômenos podem ser explicados cientificamente. Todos eles são diretamente passíveis de explicação em termos de mecanismos computacionais ou neurais. Para explicar o acesso e a capacidade de relatabilidade, por exemplo, precisamos apenas especificar o mecanismo pelo qual as informações sobre estados internos são recuperadas e disponibilizadas para relato verbal. Para explicar a integração de informações, precisamos apenas apresentar os mecanismos pelos quais as informações são reunidas e exploradas por processos posteriores. Para uma explicação do sono e da vigília, um relato neurofisiológico adequado dos processos responsáveis pelo comportamento contrastante dos organismos nesses estados será suficiente. Em cada caso, um modelo cognitivo ou neurofisiológico adequado consegue realizar claramente o trabalho explicativo.

Se esses fenômenos fossem tudo o que a consciência é, então ela não seria um grande problema. Embora ainda não tenhamos algo próximo de uma explicação completa para eles, temos uma ideia clara de como poderíamos explicá-los. É por isso que chamo esses problemas de problemas fáceis. Claro, "fácil" é um termo relativo. Acertar os detalhes provavelmente levará um ou dois séculos de árduo trabalho empírico. Ainda assim, existem todos os motivos para acreditar que os métodos da ciência cognitiva e da neurociência darão certo.

O problema realmente difícil da consciência é o problema da *experiência*. Quando pensamos e percebemos, há um zumbido de processamento de informações, mas também há um aspecto subjetivo. Como Nagel (1974) ressaltou, *há algo que caracteriza* um organismo consciente. Esse aspecto subjetivo é a experiência. Quando vemos, por exemplo, *experienciamos* sensações visuais: a qualidade percebida da vermelhidão, a experiência da escuridão e da luz, a qualidade da profundidade de um campo visual. Outras experiências acompanham a percepção em diferentes modalidades: o som de um clarinete, o cheiro de naftalina. Além disso, há sensações corporais, de dores a orgasmos; imagens mentais que são evocadas internamente; a qualidade vivenciada da emoção e a experiência de um fluxo de pensamento consciente. O que une todos esses estados é que há algo que é como estar neles. Todos são estados de experiência.

É inegável que alguns organismos são sujeitos de experiência. Mas a questão de como esses sistemas são sujeitos de experiência é intrigante. Por que, quando nossos sistemas cognitivos se envolvem no processamento de informações visuais e auditivas, temos uma experiência visual ou auditiva: a qualidade do azul profundo, a sensação do dó central? Como podemos explicar por que existe algo semelhante a entreter uma imagem mental ou vivenciar uma emoção? É amplamente aceito

que a experiência surge de uma base física, mas não temos uma boa explicação por que e como ela surge. Por que o processamento físico deveria dar origem a uma rica vida interior? Parece objetivamente irracional que isso aconteça, e, ainda assim, acontece.

Se há algum problema que se qualifique como o problema da consciência, é esse. Nesse sentido central de "consciência", um organismo é consciente se há algo que o faz ser esse organismo, e um estado mental é consciente se há algo que o faz estar nesse estado. Às vezes, termos como "consciência fenomenal" e "qualia" também são usados, mas acho mais natural falar de "experiência consciente" ou simplesmente de "experiência". Outra maneira útil de evitar confusão (por exemplo, usada por Newell [1990] e Chalmers [1996]) é reservar o termo "consciência" (*consciousness*) para os fenômenos da experiência, usando a expressão menos carregada "percepção consciente" (*awareness*) para os fenômenos mais simples descritos anteriormente. Se tal convenção fosse amplamente adotada, a comunicação seria muito mais fácil; do jeito que as coisas estão, aqueles que falam sobre "consciência" frequentemente estão falando sobre coisas diferentes.

A ambiguidade do termo "consciência" é frequentemente explorada por filósofos e cientistas que escrevem sobre o assunto. É comum ver um artigo sobre consciência começar com uma invocação do mistério da consciência, observando a estranha intangibilidade e inefabilidade da subjetividade, e preocupando-se com o fato de que até o momento não temos uma teoria do fenômeno. Neste caso, o tema 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. Após análise, essa teoria se revela uma teoria de um dos fenômenos mais simples – da relatibilidade, do acesso introspectivo, ou seja lá o que for. No final, o autor declarará que a consciência se mostra tratável, mas o leitor se sentirá vítima de uma propaganda enganosa. 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 dizem respeito à explicação de *habilidades* e *funções* cognitivas. Para explicar uma função cognitiva, precisamos apenas especificar um mecanismo que consiga executar a função. Os métodos da ciência cognitiva são adequados para esse tipo de explicação e, portanto, para os problemas fáceis da consciência. Em contraste, o problema difícil é difícil

precisamente porque não se trata de 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 estrito de algo que um sistema foi projetado para fazer, mas no sentido mais amplo de qualquer papel causal na produção de comportamento que um sistema possa desempenhar.)

Explicar a relatabilidade, por exemplo, é simplesmente explicar como um sistema poderia desempenhar a função de produzir relatórios sobre estados internos. Para explicar o acesso interno, precisamos explicar como um sistema poderia ser apropriadamente afetado por seus estados internos e usar informações sobre esses estados para direcionar processos posteriores. Para explicar integração e controle, precisamos explicar como os processos centrais de um sistema podem reunir conteúdos de informação e utilizá-los para facilitar diversos comportamentos. Todos esses são problemas relacionados à explicação de funções.

Como explicamos o desempenho de uma função? Especificando um *mecanismo* que a executa. Aqui, a modelagem neurofisiológica e cognitiva é perfeita para a tarefa. Se quisermos uma explicação detalhada de nível reduzido, podemos especificar o mecanismo neural que é responsável pela função. Se quisermos uma explicação mais abstrata, podemos especificar um mecanismo em termos computacionais. De qualquer forma, o resultado será uma explicação completa e satisfatória. Uma vez especificado o mecanismo neural ou computacional que executa a função de relato verbal, por exemplo, a maior parte do nosso trabalho de explicar a relatabilidade está concluída.

De certa forma, a questão é trivial. É um fato *conceitual* sobre esses fenômenos que sua explicação envolve apenas a explicação de várias funções, visto que os fenômenos são *funcionalmente definíveis*. Tudo o que *significa* para a relatabilidade ser fundamentada em um sistema é que o sistema tenha a capacidade de reportar verbalmente informações internas. Para um sistema estar desperto, basta que ele esteja adequadamente receptivo às informações do ambiente e seja capaz de usar essas informações para direcionar o comportamento de maneira adequada. Para ver que esse tipo de coisa é um fato conceitual, observe que alguém que diz "você explicou o desempenho da função de relato verbal, mas não explicou a relatabilidade" está cometendo um erro conceitual trivial sobre relatabilidade. Tudo o que *possivelmente* seria necessário para explicar a relatabilidade é uma explicação de como a função relevante é desempenhada; o mesmo vale para os outros fenômenos em questão.

Em todas as ciências de nível superior, a explicação reducionista funciona exatamente dessa maneira. Para explicar o gene, por exemplo, precisávamos especificar o mecanismo que armazena e transmite informações hereditárias de uma geração para a outra. Acontece que o DNA desempenha essa função; uma vez que explicamos como a função é desempenhada, explicamos o gene. Para explicar a vida, precisamos, em última análise, 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 à explicação reducionista. O mesmo se aplica à maioria dos problemas da ciência cognitiva. Para explicar a aprendizagem, precisamos explicar como as capacidades comportamentais de um sistema são modificadas à luz das informações ambientais e como novas informações podem ser utilizadas na adaptação das ações de um sistema ao seu ambiente. Se mostrarmos como um mecanismo neural ou computacional funciona, explicamos a aprendizagem. Podemos dizer o mesmo de outros fenômenos cognitivos, como percepção, memória e linguagem. Às vezes, as funções relevantes precisam ser caracterizadas de forma 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 da experiência consciente, esse tipo de explicação falha. O que torna o problema difícil complexo e quase único é que ele vai *além* dos problemas relacionados ao desempenho de funções. Para entender isso, observe que, mesmo quando explicamos o desempenho de todas as funções cognitivas e comportamentais relacionadas à experiência – discriminação perceptiva, categorização, acesso interno, relato verbal –, ainda resta uma outra pergunta sem resposta: *por que o desempenho dessas funções é acompanhado pela experiência?* Uma explicação simples das funções deixa essa questão em aberto.

Não existe nenhuma outra questão análoga na explicação dos genes, da vida ou da aprendizagem. Se alguém disser: "percebo que você explicou como o DNA armazena e transmite informações hereditárias de uma geração para a outra, mas não explicou como ele é um *gene*", estará cometendo um erro conceitual. Tudo o que significa ser um gene é ser uma entidade que desempenha a relevante função de armazenamento e transmissão. Mas se alguém disser: "percebo que você explicou como a informação é discriminada, integrada e relatada, mas não explicou como ela é *experienciada*", não estará cometendo um erro conceitual. Esta é uma questão adicional não trivial.

Esta outra pergunta é a questão-chave no problema da consciência. Por que todo esse processamento de informações não ocorre "no escuro", livre de qualquer

sensação interna? Por que, quando ondas eletromagnéticas incidem na retina e são discriminadas e categorizadas por um sistema visual, essa discriminação e categorização são vivenciadas como uma sensação de vermelho vivo? Sabemos que a experiência consciente surge quando essas funções são desempenhadas, mas o próprio fato de ela surgir é o mistério central. Existe uma *lacuna explicativa* (uma expressão atribuída a Levine, 1983) entre as funções e a experiência, e precisamos de uma ponte explicativa para atravessá-la. Uma mera descrição das funções permanece de um lado da lacuna, de tal forma que os materiais para a ponte devem ser encontrados em outro lugar.

Isso não quer dizer que a experiência *não tenha* nenhuma função. Talvez ela acabe desempenhando um importante papel cognitivo. Mas, para qualquer papel que possa desempenhar, haverá mais na explicação da experiência do que uma simples explicação da função. Talvez até aconteça que, no decorrer da explicação de uma função, sejamos levados ao insight-chave que permita uma explicação da experiência. Porém, se isso acontecer, a descoberta será uma recompensa explicativa *extra*. Não existe função cognitiva tal que possamos dizer de antemão que a explicação dessa função explicará *automaticamente* a experiência.

Para explicar a experiência, precisamos de uma nova abordagem. Os métodos usuais da ciência cognitiva e da neurociência não são suficientes. Esses métodos foram desenvolvidos precisamente para explicar o desempenho das funções cognitivas e fazem um bom trabalho nisso. Mas, do jeito que estão, esses métodos são adequados *somente* 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, diversos trabalhos abordaram os problemas da consciência no âmbito da ciência cognitiva e da neurociência. Isso pode sugerir que a análise acima seja falha, mas, na verdade, um exame atento dos trabalhos relevantes apenas fornece mais suporte a ela. Quando investigamos a quais aspectos da consciência esses estudos visam e quais aspectos eles acabam explicando, descobrimos que o alvo final da explicação é sempre um dos problemas fáceis. Ilustrarei isso com dois exemplos representativos.

O primeiro é a "teoria neurobiológica da consciência", delineada por Crick e Koch (1990; ver também Crick 1994). Essa teoria centra-se 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 se deve, em parte, ao fato de

que as oscilações parecem estar correlacionadas com a consciência em um número de modalidades diferentes – nos sistemas visual e olfativo, por exemplo – e também porque sugerem um mecanismo pelo qual a vinculação de conteúdos de informação pode ser realizada. A vinculação é o processo pelo qual peças de informações separadamente representadas sobre uma única entidade são reunidas para serem utilizadas em processamento posterior, como quando informações sobre a cor e a forma de um objeto percebido são integradas a partir de caminhos visuais distintos. Seguindo outros pesquisadores (e.g. Eckhorn *et al.*, 1988), Crick e Koch levantam a hipótese de que a vinculação pode ser alcançada pelas oscilações sincronizadas de grupos neuronais que representam os conteúdos relevantes. Quando duas informações são vinculadas, os grupos neurais relevantes oscilarão com a mesma frequência e fase.

Os detalhes de como essa vinculação pode ser alcançada ainda são pouco compreendidos, mas suponhamos que possam ser decifrados. O que a teoria resultante conseguiria explicar? Claramente, ela poderia explicar a vinculação de conteúdos de informação e, talvez, conseguisse produzir uma explicação mais geral da integração de informações no cérebro. Crick e Koch também sugerem que essas oscilações ativam os mecanismos da memória de curto prazo, de modo que poderia haver uma explicação desta e talvez de outras formas de memória de longo prazo. A teoria poderia finalmente levar a uma explicação geral de como a informação percebida é vinculada e armazenada na memória, para uso em processamento posterior.

Tal 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. Esta afirmação é discutível – a vinculação também não ocorre no processamento de informações inconscientes? –, mas mesmo que 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 desempenham na vinculação e no armazenamento, mas a questão por que a vinculação e o armazenamento devem ser acompanhados pela experiência nunca é abordada. Se não sabemos por que a vinculação e o armazenamento devam 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 dão origem à experiência, os detalhes neurofisiológicos seriam apenas a cereja do bolo. A teoria de Crick e Koch se sustenta *pressupondo* uma conexão entre vinculação e experiência e, portanto, não pode fazer nada para explicar essa ligação.

Em última análise, não creio que Crick e Koch afirmem estar abordando o problema difícil, embora alguns os tenham interpretado de outra forma. Uma entrevista com Koch apresenta uma declaração clara das limitações das ambições da teoria:

Bem, vamos primeiro esquecer os aspectos realmente difíceis, como os sentimentos subjetivos, pois eles podem não ter uma solução científica. O estado subjetivo da ação, da dor, do prazer, de ver o azul, de cheirar uma rosa — parece haver um enorme salto entre o nível materialista de explicar moléculas e neurônios e o nível subjetivo. Vamos nos concentrar em coisas que são mais fáceis de estudar — como a percepção visual. Agora, você está falando comigo, mas não está olhando para mim, está olhando para o cappuccino e, portanto, está ciente dele. Você pode dizer: "É uma xícara e há um líquido nela". Se eu lha oferecer, você moverá o braço e a pegará — você responderá de maneira expressiva. É isso que eu chamo de percepção consciente. (What is Consciousness? *Discover*, nov. 1992, p. 96.)

O segundo exemplo é uma abordagem no nível da psicologia cognitiva. Trata-se da teoria do espaço de trabalho global da consciência de Bernard Baars, apresentada em seu livro *A Cognitive Theory of Consciousness*. Segundo essa teoria, o conteúdo da consciência está contido em um espaço de trabalho global, um processador central usado para mediar a comunicação entre uma série de processadores especializados não conscientes. Quando esses processadores especializados precisam transmitir informações para o restante do sistema, eles o fazem enviando essas informações para o espaço de trabalho, que atua como uma espécie de quadro-negro comum para o restante do sistema, acessível a todos os outros processadores.

Baars utiliza esse modelo para abordar diversos aspectos da cognição humana e para explicar uma série de contrastes entre o funcionamento cognitivo consciente e inconsciente. Em última análise, porém, trata-se de uma teoria de *acessibilidade cognitiva*, que explica como certos conteúdos informacionais são amplamente acessíveis dentro de um sistema, bem como uma teoria de integração e relatabilidade informacionais. A teoria se mostra promissora como uma teoria da percepção consciente, o correlato funcional da experiência consciente, mas não oferece uma explicação para a experiência em si.

De acordo com essa teoria, pode-se supor que o conteúdo da experiência seja precisamente o conteúdo do espaço de trabalho. Mas, mesmo que assim seja, nada interno à teoria explica por que a informação dentro do espaço de trabalho global é experienciada. O máximo que a teoria pode fazer é dizer que a informação é experienciada porque é *globalmente acessível*. 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 ligação permanece sem resposta.

Quase todos os trabalhos dos últimos anos que adotam uma abordagem cognitiva ou neurocientífica da consciência poderiam ser submetidos a uma crítica semelhante. O modelo de "Darwinismo Neural" de Edelman (1989), por exemplo, aborda questões sobre a consciência perceptiva e o autoconceito, mas não diz nada sobre por que também deveria haver experiência. O modelo de "rascunhos múltiplos" de Dennett (1991) é amplamente direcionado a explicar a capacidade de relatabilidade de certos conteúdos mentais. A teoria do "nível intermediário" de Jackendoff (1987) fornece uma explicação de alguns processos computacionais subjacentes à consciência, mas Jackendoff enfatiza que a questão de como esses processos se "projetam" na experiência consciente permanece um mistério.

Pesquisadores que utilizam esses métodos em geral não são explícitos sobre suas atitudes em relação ao problema da experiência consciente, embora às vezes assumam uma posição clara. Mesmo entre aqueles que são claros sobre ela, as atitudes diferem amplamente. Ao situar esse tipo de trabalho em relação ao problema da experiência, diversas estratégias diferentes estão disponíveis. Seria útil se essas escolhas estratégicas fossem explicitadas com mais frequência.

A primeira estratégia é simplesmente *explicar algo diferente*. Alguns pesquisadores são explícitos ao afirmar que o problema da experiência é muito difícil por enquanto e, talvez, até mesmo completamente fora do domínio da ciência. Em vez disso, eles optam por abordar um dos problemas mais tratáveis, como a relatabilidade ou o autoconceito. Embora eu tenha chamado esses problemas de problemas "fáceis", eles estão entre os problemas não resolvidos mais interessantes da ciência cognitiva e, portanto, esse trabalho certamente vale o esforço. O pior que se pode dizer dessa escolha é que, no contexto da pesquisa sobre a consciência, ela é relativamente pouco ambiciosa, e o trabalho pode, às vezes, ser mal interpretado.

A segunda opção é adotar uma linha mais dura e *negar o fenômeno*. (Variações dessa abordagem são adotadas por Allport 1988; Dennett 1991; Wilkes 1988.) De acordo com essa linha, uma vez explicadas funções como acessibilidade, relatabilidade e similares, não há nenhum outro fenômeno chamado "experiência" a ser explicado. Alguns negam explicitamente o fenômeno, sustentando, por exemplo, que o que não é externamente verificável não pode ser real. Outros alcançam o mesmo efeito permitindo que a experiência exista, mas somente se equiparmos "experiência" a algo como a capacidade de discriminar e relatar. Essas abordagens levam a uma teoria mais simples, mas, em última análise, são

insatisfatórias. A experiência é o aspecto mais central e manifesto de nossas vidas mentais e, de fato, talvez seja o *explanandum* fundamental na ciência da mente. Devido a esse status como *explanandum*, a experiência não pode ser descartada como o espírito vital quando surge uma nova teoria. Em vez disso, é o fato central que qualquer teoria da consciência deve explicar. Uma teoria que nega o fenômeno "resolve" o problema ao se esquivar da questão.

Em uma terceira opção, alguns pesquisadores *afirmam explicar* a experiência em seu sentido pleno. Esses pesquisadores (ao contrário dos citados acima) desejam levar a experiência muito a sério; eles expõem seu modelo ou teoria funcional e afirmam que ele explica toda a qualidade subjetiva da experiência (por exemplo, Flohr 1992; Humphrey 1992). Entretanto, a etapa relevante na explicação geralmente é ignorada rapidamente e acaba parecendo mágica. Após alguns detalhes sobre o processamento da informação serem fornecidos, a experiência entra repentinamente em cena, mas permanece obscuro como esses processos poderiam dar origem à experiência. Talvez seja simplesmente dado como certo que sim, mas então temos uma explicação incompleta e uma versão da quinta estratégia abaixo.

Uma quarta abordagem, mais promissora, recorre a esses métodos para *explicar a estrutura da experiência*. Por exemplo, pode-se argumentar que uma explicação das discriminações feitas pelo sistema visual consegue explicar as relações estruturais entre diferentes experiências de cores, bem como a estrutura geométrica do campo visual (ver, por exemplo, Clark, 1992 e Hardin, 1992). Em geral, certos fatos sobre estruturas encontrados no processamento corresponderão e, possivelmente, explicarão fatos sobre a estrutura da experiência. Essa estratégia é plausível, mas limitada. Na melhor das hipóteses, toma a existência da experiência como garantida e considera alguns fatos sobre sua estrutura, provendo uma espécie de explicação não reducionista dos aspectos estruturais da experiência (falarei mais sobre isso posteriormente). Isso é útil para muitos propósitos, mas não nos diz nada sobre por que deveria existir experiência em primeiro lugar.

Uma quinta e razoável estratégia é *isolar o substrato da experiência*. Afinal, quase todos admitem que a experiência *surge* de uma forma ou de outra a partir de processos cerebrais, e faz sentido identificar o tipo de processo do qual ela surge. Crick e Koch propõem seu trabalho o isolamento do correlato neural da consciência, por exemplo, e Edelman (1989) e Jackendoff (1987) fazem afirmações semelhantes. A justificação dessas afirmações requer uma análise teórica cuidadosa, especialmente porque a experiência não é diretamente observável em contextos experimentais; porém, quando aplicada criteriosamente, essa estratégia pode

lançar luz indireta sobre o problema da experiência. No entanto, a estratégia é claramente incompleta. Para uma teoria satisfatória, precisamos saber mais do que apenas quais processos dão origem à experiência; precisamos de uma explicação do porquê e do como. Uma teoria completa da consciência deve construir uma ponte explicativa.

## 5. O Ingrediente Extra

Vimos que existem razões sistemáticas pelas quais os métodos usuais da ciência cognitiva e da neurociência falham em explicar a experiência consciente. Esses são simplesmente os tipos de método errados: nada do que eles nos fornecem pode gerar uma explicação. Para levar em conta a experiência consciente, precisamos de um *ingrediente extra* na explicação. Isso representa um desafio para aqueles que levam a sério o problema difícil da consciência: qual é o seu ingrediente extra e por que *ele* deveria explicar a experiência consciente?

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

O processamento não algorítmico, por exemplo, é proposto por Penrose (1989, 1994) devido ao papel que pode desempenhar no processo de insight matemático consciente. Os argumentos sobre matemática são controversos, mas mesmo que sejam bem-sucedidos e uma explicação do processamento não algorítmico no cérebro humano seja apresentada, ainda assim será apenas uma explicação das *funções* envolvidas no raciocínio matemático e afins. Tanto para um processo não algorítmico quanto para um processo algorítmico, a questão permanece sem resposta: por que esse processo deveria dar origem à experiência? Ao responder a *essa* pergunta, não se encontra um papel especial para o processamento não algorítmico.

O mesmo se aplica às dinâmicas não lineares e caóticas. Essas podem fornecer uma nova explicação da dinâmica do funcionamento cognitivo, bastante diferente daquela fornecida pelos métodos padrão da ciência cognitiva. Mas, a partir da dinâmica, obtém-se apenas mais dinâmica. A questão sobre a experiência aqui é tão misteriosa como sempre. O ponto fica ainda mais claro para novas

descobertas em neurofisiologia. Essas novas descobertas podem nos ajudar a fazer progressos significativos na compreensão da função cerebral, mas para qualquer processo neural que isolarmos, a mesma questão sempre surgirá. É difícil imaginar o que um defensor da nova neurofisiologia espera que aconteça, além da explicação de funções cognitivas adicionais. Não é como se, de repente, descobríssemos um brilho fenomenal 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 advir de 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. Os fenômenos quânticos possuem algumas propriedades funcionais notáveis, como o não determinismo e a não localidade. É natural especular que essas propriedades possam desempenhar algum papel na explicação de funções cognitivas, como a escolha aleatória e a integração de informações, e essa hipótese não pode ser descartada *a priori*. Porém, quando se trata da explicação da experiência, os processos quânticos estão no mesmo barco que quaisquer outros. A questão por que esses processos deveriam dar origem à experiência permanece totalmente 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 modo, não oferecem nenhuma esperança de *explicar* a consciência em termos de processos quânticos. Em vez disso, essas teorias *assumem* a existência da consciência e a usam na explicação dos processos quânticos. Na melhor das hipóteses, essas teorias nos dizem algo sobre o 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 especifiquemos, haverá uma pergunta sem resposta: por que esse processo deveria dar origem à experiência? Dado qualquer processo desse tipo, é conceitualmente coerente que ele possa ser explicado na ausência de experiência. Conclui-se que nenhuma mera explicação do processo físico nos dirá por que a experiência surge. O surgimento da experiência vai além do que pode ser derivado da teoria física.

A explanação puramente física é adequada para a explicação de *estruturas* físicas, expondo estruturas macroscópicas em termos de constituintes microestruturais detalhados; e fornece uma explicação satisfatória do desempenho

de *funções*, explicando-as em termos dos mecanismos físicos que as executam. Isso ocorre porque uma explicação física pode *vincular* os fatos sobre estruturas e funções: uma vez fornecidos os detalhes internos da explicação física, as propriedades estruturais e funcionais surgem como consequência automática. Mas a estrutura e a dinâmica dos processos físicos produzem apenas mais estrutura e dinâmica, de modo que estruturas e funções são tudo o que podemos esperar que esses processos expliquem. Os fatos sobre a experiência não podem ser uma consequência automática de qualquer explicação física, pois é conceitualmente coerente que qualquer processo possa existir sem experiência. A experiência pode surgir do físico, mas não é *ocasionada* pelo físico.

A moral de tudo isso é que *não se pode explicar a experiência consciente de forma elementar*. É um fato notável que métodos reducionistas – métodos que explicam um fenômeno de nível elevado inteiramente em termos de processos físicos mais básicos – funcionem bem em muitos domínios. Em certo sentido, pode-se explicar a maioria dos fenômenos biológicos e cognitivos de forma simples, uma vez que esses fenômenos são vistos como consequências automáticas de processos mais fundamentais. Seria maravilhoso se os métodos reducionistas também pudessem explicar a experiência; por muito tempo, esperei que pudessem. Infelizmente, existem razões sistemáticas por que esses métodos devem falhar. Métodos reducionistas são bem-sucedidos na maioria dos domínios porque o que precisa ser explicado nesses domínios são estruturas e funções, e esse é o tipo de coisa que uma explicação física pode ocasionar. Quando se trata de um problema que vai além da explicação de estruturas e funções, esses métodos são ineficazes.

Isso pode parecer uma reminiscência da afirmação vitalista de que nenhuma consideração física poderia explicar a vida, mas os casos são desiguais. O que motivou o ceticismo vitalista foi a dúvida sobre se os mecanismos físicos poderiam desempenhar as muitas funções notáveis associadas à vida, como o comportamento adaptativo complexo e a reprodução. A afirmação conceitual de que a explicação de funções é o que é necessário era implicitamente aceita, mas, por falta de conhecimento detalhado dos mecanismos bioquímicos, os vitalistas duvidaram que algum processo físico pudesse realizar a tarefa e propuseram a hipótese do espírito vital como uma explicação alternativa. Assim que se constatou que os processos físicos poderiam desempenhar as funções relevantes, as dúvidas vitalistas se dissiparam.

Com a experiência, por outro lado, a explicação física das funções não está em questão. A chave, em vez disso, é o ponto *conceitual* de que a explicação das funções não é suficiente para a explicação da experiência. Esse ponto conceitual

básico não é algo que investigações neurocientíficas futuras afetarão. De forma semelhante, a experiência não é análoga ao *élan vital*. O espírito vital foi apresentado como um postulado explicativo, a fim de elucidar as funções relevantes e, portanto, poderia, ser descartado quando essas funções fossem explicadas sem ele. A experiência não é um postulado explicativo, mas um *explanandum* por si só e, portanto, não é candidata a esse tipo de eliminação.

É tentador observar que todos os tipos de fenômenos intrigantes acabaram se revelando explicáveis em termos físicos. Mas cada um deles era um problema relacionado ao comportamento observável de objetos físicos, resumindo-se a problemas de explicação de estruturas e funções. Por isso, esses fenômenos sempre foram o tipo de coisa que uma explicação física *poderia* elucidar, mesmo que em alguns momentos tenha havido boas razões para suspeitar que tal explicação não seria encontrada.

## 6. Explicação Não Reducionista

Nesse ponto, alguns se sentem tentados a desistir, sustentando que jamais teremos uma teoria da experiência consciente. McGinn (1989), por exemplo, argumenta que o problema é complexo demais para nossas mentes limitadas; estamos "cognitivamente fechados" em relação ao fenômeno. Outros argumentam que a experiência consciente está completamente fora do domínio da teoria científica.

Creio que esse pessimismo é prematuro. Este não é o momento para desistir; é o momento em que as coisas ficam interessantes. Quando métodos simples de explicação são descartados, precisamos investigar as alternativas. Dado que a explicação reducionista falha, a explicação não reducionista é a escolha natural.

Embora um número notável de fenômenos tenha se mostrado totalmente explicável em termos de entidades mais simples do que eles, isso não é universal. Na física, ocasionalmente, acontece que uma entidade tenha que ser considerada *fundamental*. Entidades fundamentais não são explicadas em termos de algo mais simples. Em vez disso, são consideradas básicas e proporcionam uma teoria de como se relacionam com tudo o mais do mundo. Por exemplo, no século XIX, descobriu-se que os processos eletromagnéticos não podiam ser explicados em termos dos processos totalmente mecânicos aos quais as teorias físicas anteriores recorriam; então, Maxwell e outros apresentaram a carga eletromagnética e as forças eletromagnéticas como novos componentes fundamentais de uma teoria física. Para explicar o eletromagnetismo, a ontologia da física teve de ser expandida.

Novas propriedades e leis básicas foram necessárias para dar uma explicação satisfatória dos fenômenos.

Outras características que a teoria física considera fundamentais incluem a massa e o espaço-tempo. Não há nenhuma tentativa de explicar essas características 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 sobre como essas características se inter-relacionam e sobre as leis básicas em que se inserem. Esses princípios básicos são usados para explicar muitos fenômenos familiares relativos à massa, ao espaço e ao tempo em um nível mais elevado.

Sugiro que uma teoria da consciência deva assumir a experiência como fundamental. Sabemos que uma teoria da consciência requer a adição de *algo* fundamental à nossa ontologia, pois tudo na teoria física é compatível com a ausência de consciência. Poderíamos adicionar alguma característica não física inteiramente nova, da qual a experiência pudesse ser derivada, mas é difícil imaginar como seria tal característica. Mais provavelmente, consideraremos a própria experiência como uma característica fundamental do mundo, juntamente com a massa, a carga e o espaço-tempo. Se considerarmos a experiência como fundamental, então poderemos nos dedicar à construção de uma teoria da experiência.

Onde há uma propriedade fundamental, há leis fundamentais. Uma teoria não reducionista da experiência adicionará novos princípios ao conjunto de leis básicas da natureza. Esses princípios básicos, em última análise, carregarão o fardo explicativo em uma teoria da consciência. Assim como explicamos fenômenos familiares de alto nível que envolvem massa em termos de princípios mais básicos envolvendo massa e outras entidades, poderíamos explicar fenômenos familiares envolvendo experiência em termos de princípios mais básicos envolvendo experiência e outras entidades.

Em particular, uma teoria não reducionista da experiência especificará princípios básicos que nos dizem como a experiência depende das características físicas do mundo. Esses princípios *psicofísicos* não interferirão nas leis físicas, visto que estas parecem já formar um sistema fechado. Em vez disso, serão um suplemento a 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 apenas de leis físicas. Os novos princípios básicos postulados por uma teoria não reducionista nos fornecem o ingrediente extra necessário para construir uma ponte explicativa.

É claro que, ao assumir a experiência como fundamental, há uma percepção de que essa abordagem não nos diz por que existe experiência em primeiro lugar. Mas isso vale para qualquer teoria fundamental. Nada na física nos diz por que existe matéria em primeiro lugar, mas não consideramos isso como indo de encontro a teorias da matéria. Certas características do mundo precisam ser consideradas fundamentais por qualquer teoria científica. Uma teoria da matéria ainda pode explicar todos os tipos de fatos sobre a matéria, mostrando como eles são consequências das leis básicas. O mesmo vale para uma teoria da experiência.

Essa posição se qualifica como uma variedade de dualismo, pois postula propriedades básicas que vão além das propriedades invocadas pela física. Mas é uma versão inocente de dualismo, inteiramente compatível com a visão científica do mundo. Nada nessa abordagem contradiz qualquer coisa da teoria física; precisamos apenas adicionar mais princípios *vinculatórios* para explicar como a experiência surge a partir de processos físicos. Não há nada particularmente espiritual ou místico nessa teoria – sua forma geral é semelhante à de uma teoria física, com algumas entidades fundamentais conectadas por leis fundamentais. Certamente, ela expande ligeiramente a ontologia, mas Maxwell fez a mesma coisa. Na verdade, a estrutura geral dessa posição é inteiramente naturalista, admitindo que o universo se reduz a uma rede de entidades básicas que obedecem a leis simples, e admitindo que possa haver, em última análise, uma teoria da consciência formulada em termos dessas leis. Se essa posição precisar de um nome, uma boa escolha poderia ser *dualismo naturalista*.

Se essa visão estiver correta, então, de certo modo, uma teoria da consciência terá mais em comum com uma teoria da física do que com uma teoria da biologia. Teorias biológicas não envolvem princípios fundamentais dessa forma; portanto, a teoria biológica apresenta certa complexidade e desordem; mas teorias da física, na medida em que lidam com princípios fundamentais, aspiram à simplicidade e à elegância. As leis fundamentais da natureza fazem parte da estrutura básica do mundo, e as teorias físicas nos dizem que essa estrutura básica é notavelmente simples. Se uma teoria da consciência também envolve princípios fundamentais, então devemos esperar o mesmo. Os princípios de simplicidade, elegância e até mesmo 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 os processos físicos e a experiência, não precisa haver uma lacuna metafísica, de modo que a experiência pode, em certo sentido, ainda ser física [e.g. Hill 1991, Levine 1983, Loar 1990]. Normalmente, essa linha de

argumentação é apoiada por um apelo à noção de necessidade *a posteriori* [Kripke 1980]. Contudo, penso que essa posição se baseia em um mal-entendido da necessidade *a posteriori*, ou requer um tipo inteiramente novo de necessidade, o qual não temos razão para acreditar; ver Chalmers 1996 [também Jackson 1994 e Lewis 1994] para mais detalhes. De qualquer forma, essa posição ainda admite uma lacuna explicativa entre os processos físicos e a experiência. Por exemplo, os princípios que conectam o físico e o experiencial não serão deriváveis das leis da física e, portanto, tais princípios devem ser assumidos como explicativamente fundamentais. Desse modo, mesmo neste tipo de perspectiva, a estrutura explicativa de uma teoria da consciência será muito semelhante à que descrevi.)

## 7. Esboço de uma Teoria da Consciência

Não é cedo demais para começarmos a trabalhar em uma teoria. Já estamos em condições de compreender certos fatos-chave sobre a relação entre processos físicos e experiência, e sobre as regularidades que os conectam. Uma vez deixada de lado a explicação reducionista, podemos colocar esses fatos sobre a mesa para que possam desempenhar seu papel adequado como peças iniciais de uma teoria não reducionista da consciência e como condicionantes a leis básicas que constituem uma teoria definitiva.

Há um problema óbvio que dificulta o desenvolvimento de uma teoria da consciência: a escassez de dados objetivos. A experiência consciente não é diretamente observável em um contexto experimental, portanto, não podemos gerar dados à vontade sobre a relação entre processos físicos e experiência. No entanto, todos temos acesso a uma rica fonte de dados em nosso próprio caso. Muitas regularidades importantes entre experiência e processamento podem ser inferidas a partir de considerações sobre a própria experiência. Existem também boas fontes indiretas de dados de casos observáveis, como quando se confia no relato verbal de um sujeito como indicação de experiência. Esses métodos têm suas limitações, mas temos dados mais do que suficientes para lançar uma teoria.

A análise filosófica também é útil para extrair o melhor custo-benefício dos dados que possuímos. Esse tipo de análise pode gerar uma série de princípios que relacionam consciência e cognição, restringindo assim fortemente a forma de uma teoria definitiva. O método de experimentos mentais também pode gerar recompensas significativas, como veremos. Por fim, o fato de estarmos buscando uma teoria *fundamental* significa que podemos recorrer a condicionantes não empíricos como simplicidade, homogeneidade e afins, ao desenvolvê-la. Devemos

buscar sistematizar as informações que possuímos, estendê-las o máximo possível por meio de uma análise cuidadosa e, então, fazer a inferência para a teoria mais simples possível que explique os dados, permanecendo, ao mesmo tempo, uma candidata plausível a fazer parte dos apetrechos fundamentais do mundo.

Tais teorias sempre reterão um elemento de especulação que não está presente em outras teorias científicas, devido à impossibilidade de testes experimentais intersubjetivos conclusivos. Ainda assim, podemos certamente construir teorias que sejam compatíveis com os dados que possuímos e avaliá-las em comparação 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 de outros domínios, a capacidade de reproduzir as propriedades da experiência que nos são familiares em nosso próprio caso e até mesmo uma adequação geral aos ditames do senso comum. Talvez ainda existam indeterminações significativas mesmo quando todos esses condicionantes forem aplicados, mas podemos pelo menos desenvolver candidatas plausíveis. Somente quando as teorias candidatas forem desenvolvidas seremos capazes de avaliá-las.

Uma teoria da consciência não reducionista consistirá em uma série de *princípios psicofísicos*, princípios que conectam as propriedades dos processos físicos às propriedades da experiência. Podemos pensar nesses princípios como encapsulando a maneira como a experiência surge do físico. Em última análise, esses princípios devem nos dizer que tipo de sistemas físicos terão experiências associadas e, para os sistemas que as têm, devem nos dizer que tipo de propriedades físicas são relevantes para o surgimento da experiência e que tipo de experiência devemos esperar que qualquer sistema físico produza. Essa é uma tarefa árdua, mas não há razão para não a começarmos.

A seguir, apresento meus próprios candidatos para os princípios psicofísicos que poderiam compor uma teoria da consciência. Os dois primeiros são *princípios não básicos* – conexões sistemáticas entre processamento e experiência em um nível relativamente elevado. Esses princípios podem desempenhar um papel significativo no desenvolvimento e na limitação de uma teoria da consciência, mas não são formulados em um nível suficientemente fundamental para se qualificarem como leis verdadeiramente básicas. O último princípio é meu candidato para um *princípio básico* que pode formar a pedra angular de uma teoria fundamental da consciência. Esse último princípio é particularmente especulativo, mas é o tipo de especulação necessária se quisermos ter uma teoria da consciência satisfatória.

Consigo apresentar aqui esses princípios apenas brevemente; defendendo-os com muito mais detalhes em Chalmers (1996).

### 7.1 O princípio da coerência estrutural

Esse é um princípio de coerência entre a *estrutura da consciência* e a *estrutura da percepção consciente*. Lembre-se de que a expressão "percepção consciente" foi usada anteriormente para se referir aos diversos fenômenos funcionais 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 da percepção consciente devem ser entendidos como aqueles conteúdos de informação que são acessíveis aos sistemas centrais e utilizados de forma generalizada no controle do comportamento. Em resumo, podemos pensar a percepção consciente como *disponibilidade direta para o controle global*. Em uma primeira aproximação, os conteúdos de percepção consciente são aqueles que são diretamente acessíveis e potencialmente relatáveis, pelo menos em um sistema que utilize linguagem.

A percepção consciente é uma noção puramente funcional, mas, ainda assim, está intimamente ligada à experiência consciente. Em casos comuns, onde quer que encontremos consciência, encontramos percepção consciente. Onde quer que haja experiência consciente, há alguma informação correspondente no sistema cognitivo que está disponível para controle do comportamento e disponível para relato verbal. Por outro lado, parece que sempre que há informação disponível para relato e para controle global, há uma experiência consciente correspondente. Assim, há uma correspondência direta entre consciência e percepção consciente.

A correspondência pode ser aprofundada. É um fato central sobre a experiência que ela tenha uma estrutura complexa. O campo visual, por exemplo, possui uma geometria complexa. Existem também relações de similaridade e diferença entre experiências, e relações em aspectos como intensidade relativa. A experiência de cada sujeito pode ser, pelo menos parcialmente, caracterizada e decomposta 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 uma dessas características estruturais, haja uma característica correspondente na estrutura de processamento de informações da percepção consciente.

Tomemos como exemplo as sensações de cor. Para cada distinção entre experiências de cores, há uma distinção correspondente no processamento. As

diferentes cores fenomenais que experienciamos formam um espaço tridimensional complexo, variando em matiz, saturação e intensidade. As propriedades desse espaço podem ser recuperadas a partir de considerações de processamento de informações: o exame dos sistemas visuais mostra que as formas de onda da luz são discriminadas e analisadas ao longo de três eixos diferentes, e é essa informação tridimensional que é relevante para o processamento posterior. Portanto, a estrutura tridimensional do espaço fenomenal de cores corresponde diretamente à estrutura tridimensional da percepção consciente visual. É exatamente isso que esperaríamos. Afinal, toda distinção de cores corresponde a alguma informação reportável e, desse modo, a uma distinção que é representada na estrutura do processamento.

De forma mais direta, a estrutura geométrica do campo visual é refletida inteiramente em uma estrutura que pode ser recuperada a partir do processamento visual. Toda relação geométrica corresponde a algo que pode ser relatado e, portanto, cognitivamente representado. Se nos fosse dada apenas a história do processamento de informações no sistema visual e cognitivo de um agente, não poderíamos observar *diretamente* as experiências visuais desse agente, mas, ainda assim, poderíamos inferir as propriedades estruturais dessas experiências.

Em geral, qualquer informação conscientemente experienciada também será cognitivamente representada. A estrutura granular do campo visual corresponderá a alguma estrutura granular no processamento visual. O mesmo se aplica a experiências em outras modalidades e até mesmo a experiências não sensoriais. Imagens mentais internas possuem propriedades geométricas que são representadas no processamento. Até mesmo emoções possuem propriedades estruturais, como intensidade relativa, que correspondem diretamente a uma propriedade estrutural do processamento; onde há maior intensidade, encontramos um efeito maior nos processos posteriores. Em geral, precisamente porque as propriedades estruturais da experiência são acessíveis e relatáveis, essas propriedades serão diretamente representadas na estrutura da percepção consciente.

É esse isomorfismo entre as estruturas da consciência e da percepção que constitui o *princípio da coerência estrutural*. Este 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 sim aderem de forma íntima.

Esse princípio tem seus limites. Ele nos permite recuperar propriedades estruturais da experiência a partir de propriedades de processamento de informação, mas nem todas as propriedades da experiência são propriedades estruturais. Há propriedades da experiência, tais como a natureza intrínseca de uma sensação de vermelho, que não podem ser totalmente capturadas em uma descrição estrutural. A própria inteligibilidade dos cenários de espectro invertido, onde as experiências de vermelho e verde são invertidas, mas todas as propriedades estruturais permanecem as mesmas, mostra que as propriedades estruturais restringem a experiência sem esgotá-la. No entanto, o próprio fato de nos sentirmos compelidos a deixar as propriedades estruturais inalteradas, quando imaginamos experiências invertidas entre sistemas funcionalmente idênticos, demonstra quão central é o princípio da coerência estrutural para a concepção de nossas vidas mentais. Não é um princípio *logicamente* necessário, pois, afinal, podemos imaginar todo o processamento de informações ocorrendo sem qualquer experiência, mas, ainda assim, é um condicionante 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 da informação visual para explicar indiretamente a estrutura do espaço de cores. Os fatos sobre o processamento neural podem implicar e explicar a estrutura da consciência; se tomarmos o princípio da coerência como certo, a estrutura da experiência também será explicada. A investigação empírica pode até nos levar a compreender melhor a estrutura da percepção consciente de um morcego, lançando luz indireta sobre a incômoda questão de Nagel sobre como é ser um morcego. Esse princípio fornece uma interpretação natural de muitos trabalhos existentes sobre a explicação da consciência (e.g. Clark 1992; Hardin 1992 sobre cores; e Akins 1993 sobre morcegos), embora seja frequentemente invocado inexplicitamente. É tão familiar que é considerado óbvio por quase todos e é um pilar central na explicação cognitiva da consciência.

A coerência entre consciência e percepção consciente também permite uma interpretação natural do trabalho em neurociência voltado para o isolamento do *substrato* (ou *correlato neural*) da consciência. Diversas hipóteses específicas foram propostas. Por exemplo, Crick e Koch (1990) sugerem que oscilações de 40 Hz podem ser o correlato neural da consciência, enquanto Libet (1993) sugere que a atividade neural temporalmente estendida é central. Se aceitarmos o princípio da coerência, o correlato físico mais *direto* da consciência é a percepção consciente: o processo pelo qual a informação é disponibilizada diretamente para controle global.

As diferentes hipóteses específicas podem ser interpretadas como sugestões empíricas sobre como a percepção consciente pode ser alcançada. Por exemplo, Crick e Koch sugerem que oscilações de 40 Hz são a porta de entrada pela qual a informação é integrada à memória de trabalho e, assim, disponibilizada para processos posteriores. Da mesma forma, é natural supor que a atividade temporalmente estendida de Libet seja relevante precisamente porque somente esse tipo de atividade alcança 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 os "inputs seletores para sistemas de ação" de Shallice (1972). Todos esses podem ser vistos como hipóteses sobre os *mecanismos de percepção consciente*: os mecanismos que desempenham a função de tornar a informação diretamente disponível para controle global.

Dada a coerência entre consciência e percepção consciente, segue-se que um mecanismo de percepção será, ele próprio, um correlato da experiência consciente. A questão de *quais* mecanismos cerebrais regem a disponibilidade global é empírica; talvez existam muitos desses mecanismos. Mas, se aceitarmos o princípio da coerência, temos motivos para acreditar que os processos que *explicam* a percepção consciente farão, ao mesmo tempo, parte da base da consciência.

## 7.2 O princípio da invariância organizacional

Esse princípio afirma que quaisquer dois sistemas com a mesma *organização funcional* de granularidade fina terão experiências qualitativamente idênticas. Se os padrões causais da organização neural fossem duplicados em silício, por exemplo, com um chip de silício para cada neurônio e os mesmos padrões de interação, então surgiriam as mesmas experiências. De acordo com esse princípio, o que importa para o surgimento da experiência não é a composição física específica de um sistema, mas o padrão abstrato de interação causal entre seus componentes. Esse princípio é controverso, claro. Alguns pesquisadores (e.g. Searle, 1980) acreditam que a consciência está vinculada a uma biologia específica, de modo que um isomorfo de silício de um humano não precisa ser consciente. No entanto, acredito que o princípio pode ser significativamente corroborado pela análise de experimentos mentais.

Muito resumidamente: suponhamos (para fins de *reductio ad absurdum*) que o princípio seja falso e que poderia haver dois sistemas funcionalmente isomórficos com experiências diferentes. Talvez apenas um dos sistemas seja consciente, ou talvez ambos sejam conscientes, mas têm experiências diferentes. Para fins de

ilustração, digamos que um sistema é feito de neurônios e o outro de silício, e que um experiencia o vermelho enquanto o outro experiencia o azul. Os dois sistemas têm a mesma organização; assim, podemos imaginar a transformação gradual de um no outro, talvez substituindo neurônios, um de cada vez, por chips de silício com a mesma função local. Desse modo, obtemos um espectro de casos intermediários, cada um com a mesma organização, mas com constituição física e experiências ligeiramente diferentes. Ao longo desse 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, exceto que um pequeno circuito neural em *A* foi substituído por um circuito de silício em *B*.

O passo fundamental no experimento mental é pegar 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 acionamos o interruptor? Por hipótese, as experiências conscientes do sistema mudarão; de vermelho para azul, digamos, para fins ilustrativos. 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, dadas as premissas, não há como o sistema *notar* as mudanças! Sua organização causal permanece constante, de modo que todos os seus estados funcionais e disposições comportamentais permanecem fixos. Para o sistema, nada de anormal aconteceu. Não há espaço para o pensamento: "Hum! Algo estranho acabou de acontecer!". Em geral, a estrutura de qualquer pensamento desse tipo deve ser refletida no processamento, mas a estrutura do processamento permanece constante aqui. Se houvesse tal pensamento, ele deveria flutuar completamente livre do sistema e seria totalmente impotente para afetar o processamento posterior. (Se afetasse o processamento posterior, os sistemas seriam funcionalmente distintos, ao contrário da hipótese.) Poderíamos até mesmo acionar o interruptor várias vezes, de modo que as experiências de vermelho e azul dançassem para frente e para trás diante do "olho interior" do sistema. De acordo com a hipótese, o sistema jamais conseguiria perceber essas "qualia dançantes".

Considero que isso seja uma *reductio* da suposição original. É um fato central sobre a experiência, muito familiar em nosso próprio caso, que sempre que as experiências mudam significativamente e estamos atentos, podemos perceber a mudança; se não fosse esse o caso, seríamos levados à possibilidade cética de que nossas experiências estão dançando diante de nossos olhos o tempo todo. Essa hipótese tem o mesmo status que a possibilidade de que o mundo tenha 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. Em termos técnicos, as hipóteses filosóficas de "qualia ausente" e "qualia invertida", embora logicamente possíveis, são empírica e nomologicamente impossíveis.

(Algumas pessoas podem se preocupar que um isomorfo de silício de um sistema neural possa ser impossível por razões técnicas. Essa questão está em aberto. 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 proporciona o sabor básico. Mais uma vez, esse experimento mental baseia-se em fatos familiares sobre a coerência entre consciência e processamento cognitivo, para chegar a uma conclusão sólida sobre a relação entre estrutura física e experiência. Se o argumento for válido, sabemos que as únicas propriedades físicas diretamente relevantes para o surgimento da experiência são as propriedades *organizacionais*. Isso atua como mais um forte condicionante a uma teoria da consciência.

### 7.3 A teoria do duplo aspecto da informação

Os dois princípios precedentes são princípios *não básicos*. Envolvem noções de nível elevado, como "percepção consciente" e "organização", e, portanto, situam-se no nível errado para constituir as leis fundamentais de uma teoria da consciência. No entanto, atuam como fortes condicionantes. O que é ainda necessário são princípios básicos que se ajustem a esses condicionantes e que possam, em última análise, explicá-los.

O princípio básico que sugiro envolve, centralmente, a noção de *informação*. Entendo informação mais ou menos no sentido de Shannon (1948). Onde há informação, há *estados de informação* incorporados em um *espaço de informação*. Um espaço de informação possui uma estrutura básica de relações de *diferença* entre seus elementos, caracterizando os modos pelos quais diferentes elementos em um espaço são semelhantes ou diferentes, possivelmente de maneiras complexas. Um espaço de informação é um objeto abstrato, mas, seguindo Shannon, podemos ver a informação como *fisicamente incorporada* quando há um espaço de estados físicos distintos, cujas diferenças entre eles possam ser transmitidas por algum caminho causal. Os estados que são transmitidos podem ser

vistos como constituindo, eles próprios, um espaço de informação. Usando uma frase de Bateson (1972), a informação física é uma *diferença que faz uma diferença*.

O princípio do duplo aspecto decorre da observação de que existe um isomorfismo direto entre certos espaços de informação fisicamente incorporados e certos espaços de informação *fenomenais* (ou experienciais). A partir do mesmo tipo de observação que levou ao princípio da coerência estrutural, podemos notar que as diferenças entre estados fenomenais têm uma estrutura que corresponde diretamente às diferenças embutidas em processos físicos; em particular, àquelas diferenças que fazem uma diferença em certos caminhos causais implicados na disponibilidade e no controle globais. Ou seja, podemos encontrar o *mesmo* espaço abstrato de informação embutido no processamento físico e na experiência consciente.

Isto leva a uma hipótese natural: a de que a informação (ou pelo menos alguma informação) possui dois aspectos básicos: um aspecto físico e um aspecto fenomenal. Este tem o status de um princípio básico que pode fundamentar e explicar o surgimento da experiência a partir do físico. A experiência surge em virtude de seu status como um aspecto da informação, enquanto o outro aspecto se encontra incorporado no processamento físico.

Esse princípio é sustentado por uma série de considerações, que só consigo delinear brevemente aqui. Primeiro, a consideração do tipo de mudanças físicas que correspondem a mudanças na experiência consciente sugere que tais mudanças são sempre relevantes, em virtude de seu papel na constituição de *mudanças informacionais* – diferenças no âmbito de um espaço abstrato de estados, que são divididos precisamente de acordo com suas diferenças causais ao longo de certos caminhos causais. Em segundo lugar, para que o princípio da invariância organizacional se mantenha, precisamos encontrar alguma propriedade *organizacional* fundamental à qual a experiência possa ser vinculada, e a informação é uma propriedade organizacional *por excelência*. Em terceiro lugar, esse princípio oferece alguma esperança de explicar o princípio da coerência estrutural em termos da estrutura presente nos espaços de informação. Em quarto lugar, a análise da explicação cognitiva de nossos *julgamentos* e *afirmações* sobre a experiência consciente – julgamentos que são funcionalmente explicáveis, mas, ainda assim, profundamente ligados à própria experiência – sugere que a explicação envolve centralmente os estados de informação incorporados ao processamento cognitivo. Conclui-se que uma teoria baseada em informação permite uma coerência profunda entre a explicação da experiência e a explicação de nossos julgamentos e afirmações sobre ela.

Wheeler (1990) sugeriu que a informação é fundamental para a física do universo. De acordo com essa doutrina do "it from bit" (isto é, todas as coisas físicas têm origem na teoria da informação), as leis da física podem ser formuladas em termos de informação, postulando diferentes estados que dão origem a diferentes efeitos, sem, de fato, dizer quais *são* esses estados. É apenas a posição deles em um espaço informacional que importa. Se assim for, então a informação é uma candidata natural a também desempenhar um papel em uma teoria fundamental da consciência. Somos levados a uma concepção de mundo na qual a informação é verdadeiramente essencial e na qual ela possui dois aspectos básicos, correspondentes às características físicas e fenomenais do mundo.

É claro que o princípio do duplo aspecto da informação é extremamente especulativo e também subdeterminado, deixando uma série de questões-chave sem resposta. Uma questão óbvia é saber se *toda* informação tem um aspecto fenomenal. Uma possibilidade é que precisemos de um condicionante adicional à teoria fundamental, indicando exatamente que *tipo* de informação tem um aspecto fenomenal. A outra possibilidade é que não exista tal condicionante. Se não existe, então a experiência seria muito mais difundida do que poderíamos acreditar, visto que a informação estaria em toda parte. A princípio, isso é contraintuitivo, mas, refletindo, acho que a posição ganha certa plausibilidade e elegância. Onde há processamento simples de informações, há experiência simples, e onde há processamento complexo de informações, há experiência complexa. Um rato tem uma estrutura de processamento de informações mais simples do que um humano e tem 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 maioria das propriedades fundamentais é distribuída de forma mais uniforme. De qualquer modo, esta é uma questão em aberto, mas acredito que a posição não seja tão implausível quanto frequentemente se pensa.

Uma vez que um elo fundamental entre informação e experiência seja estabelecido, abre-se a porta para especulações metafísicas mais amplas sobre a natureza do mundo. Por exemplo, observa-se frequentemente 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 resta um mundo que é puro fluxo causal (um puro fluxo de informação) sem propriedades para a causalidade relacionar-se. Se admitirmos que

existem propriedades intrínsecas, dado o exposto acima, uma especulação natural é que as propriedades intrínsecas do físico – as propriedades que a causalidade, em última análise, relaciona – são, elas próprias, propriedades fenomenais. Poderíamos dizer que as propriedades fenomenais são o aspecto interno da informação. Isso poderia responder a uma preocupação sobre a relevância causal da experiência — uma preocupação natural, dada uma imagem em que o domínio físico é causalmente fechado e em que a experiência é suplementar ao físico. 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, seja melhor ignorar essa especulação metafísica para fins de desenvolvimento de uma teoria científica, mas, ao abordar algumas questões filosóficas, ela é bastante sugestiva.

## **8. Conclusão**

A teoria que apresentei é especulativa, mas é uma teoria passível de ser analisada. Suspeito que os princípios da coerência estrutural e invariância organizacional serão pilares em qualquer teoria satisfatória da consciência; o status da teoria do duplo aspecto da informação é menos certo. Na verdade, no momento, ela é mais uma ideia do que uma teoria. Para ter alguma esperança de eventual sucesso explicativo, ela terá de ser especificada de modo mais completo e desenvolvido para atingir uma forma mais poderosa. Ainda assim, a reflexão sobre o que é plausível e implausível nela, sobre onde funciona e onde falha, só pode levar a uma teoria melhor.

A maioria das teorias existentes sobre a consciência nega o fenômeno, explica algo diferente ou eleva o problema a um mistério eterno. Espero ter demonstrado que é possível progredir no problema, levando-o a sério. Para progredir ainda mais, precisaremos de mais investigação, teorias mais refinadas e análises mais cuidadosas. O problema difícil é um problema difícil, mas não há razão para se acreditar que ele permanecerá permanentemente sem solução.

## **Leitura Adicional**

Os problemas da consciência têm sido amplamente discutidos na literatura filosófica recente. Para esclarecimentos conceituais dos vários problemas da consciência, ver Block 1995, Nelkin 1993 e Tye 1995. Aqueles que enfatizaram as dificuldades de explicar a experiência em termos físicos incluem Hodgson 1988,

Jackson 1982, Levine 1983, Lockwood 1989, McGinn 1989, Nagel 1974, Seager 1991, Searle 1992, Strawson 1994 e Velmans 1991, entre outros. Aqueles que adotam uma abordagem reducionista incluem Churchland (1995), Clark (1992), Dennett (1991), Dretske (1995), Kirk (1994), Rosenthal (1996) e Tye (1995). Não há muitas tentativas de construir teorias não reducionistas detalhadas na literatura, mas veja Hodgson (1988) e Lockwood (1989) para algumas reflexões nessa direção. Duas excelentes coletâneas de artigos recentes sobre consciência são Block, Flanagan & Güzeldere (1996) e Metzinger (1995).

## Bibliografia

Akins, K. 1993. What is it like to be boring and myopic? Em (B. Dahlbom, ed.) *Dennett and his Critics*. Oxford: Blackwell.

Allport, A. 1988. What concept of consciousness? Em (A. Marcel & E. Bisiach, eds.) *Consciousness in Contemporary Science*. Oxford: Oxford University Press.

Baars, B.J. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Bateson, G. 1972. *Steps to an Ecology of Mind*. Chandler Publishing.

Block, N. 1995. On a confusion about the function of consciousness. *Behavioral and Brain Sciences*.

Block, N, O Flanagan, and G Güzeldere, (eds.) 1996. *The Nature of Consciousness: Philosophical and Scientific Debates*. Cambridge, MA: MIT Press.

Chalmers, D.J. 1996. *The Conscious Mind*. New York: Oxford University Press.

Churchland, P.M. 1995. *The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain*. Cambridge, MA: MIT Press.

Clark, A. 1992. *Sensory Qualities*. Oxford: Oxford University Press.

Crick, F. and C Koch, 1990. Toward a neurobiological theory of consciousness. *Seminars in the Neurosciences* 2:263-275.

Crick, F. 1994. *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Scribners.

Dennett, D.C. 1991. *Consciousness Explained*. Boston: Little, Brown.

Dretske, F.I. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Edelman, G. 1989. *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.

Farah, M.J. 1994. Visual perception and visual awareness after brain damage: A tutorial overview. Em (C. Umiltà and M. Moscovitch, eds.) *Consciousness and Unconscious Information Processing: Attention and Performance 15*. Cambridge, MA: MIT Press.

Flohr, H. 1992. Qualia and brain processes. Em (A. Beckermann, H. Flohr, and J. Kim, eds.) *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. Berlin: De Gruyter.

Hameroff, S.R. 1994. Quantum coherence in microtubules: A neural basis for emergent consciousness? *Journal of Consciousness Studies* 1:91-118.

Hardin, C.L. 1992. Physiology, phenomenology, and Spinoza's true colors. Em (A. Beckermann, H. Flohr & J. Kim, eds.) *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. Berlin: De Gruyter.

Hill, C.S. 1991. *Sensations: A Defense of Type Materialism*. Cambridge: Cambridge University Press.

Hodgson, D. 1988. *The Mind Matters: Consciousness and Choice in a Quantum World*. Oxford: Oxford University Press.

Humphrey, N. 1992. *A History of the Mind*. New York: Simon and Schuster.

Jackendoff, R. 1987. *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.

Jackson, F. 1982. Epiphenomenal qualia. *Philosophical Quarterly* 32: 127-36.

Jackson, F. 1994. Finding the mind in the natural world. In (R. Casati, B. Smith, & S. White, eds.) *Philosophy and the Cognitive Sciences*. Vienna: Hölder-Pichler-Tempsky.

Kirk, R. 1994. *Raw Feeling: A Philosophical Account of the Essence of Consciousness*. Oxford: Oxford University Press.

Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Levine, J. 1983. Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64:354-61.

Lewis, D. 1994. Reduction of mind. In (S. Guttenplan, ed.) *A Companion to the Philosophy of Mind*. Oxford: Blackwell.

Libet, B. 1993. The neural time factor in conscious and unconscious events. Em (G.R. Block & J. Marsh, eds.) *Experimental and Theoretical Studies of Consciousness* (Ciba Foundation Symposium 174). Chichester: John Wiley and Sons.

Loar, B. 1990. Phenomenal states. *Philosophical Perspectives* 4:81-108.

Lockwood, M. 1989. *Mind, Brain, and the Quantum*. Oxford: Blackwell.

McGinn, C. 1989. Can we solve the mind-body problem? *Mind* 98:349-66.

Metzinger, T. 1995. *Conscious Experience*. Paderborn: Schöningh.

Nagel, T. 1974. What is it like to be a bat? *Philosophical Review* 4:435-50.

Nelkin, N. 1993. What is consciousness? *Philosophy of Science* 60:419-34.

Newell, A. 1990. *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Penrose, R. 1989. *The Emperor's New Mind*. Oxford: Oxford University Press.

Penrose, R. 1994. *Shadows of the Mind*. Oxford: Oxford University Press.

Rosenthal, D.M. 1996. A theory of consciousness. Em (N. Block, O. Flanagan, & G. Güzeldere, eds.) *The Nature of Consciousness*. Cambridge, MA: MIT Press.

Seager, W.E. 1991. *Metaphysics of Consciousness*. London: Routledge.

Searle, J.R. 1980. Minds, brains and programs. *Behavioral and Brain Sciences* 3:417-57.

Searle, J.R. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Shallice, T. 1972. Dual functions of consciousness. *Psychological Review* 79:383-93.

Shannon, C.E. 1948. A mathematical theory of communication. *Bell Systems Technical Journal* 27: 379-423.

Strawson, G. 1994. *Mental Reality*. Cambridge, MA: MIT Press.

Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

Velmans, M. 1991. Is human information-processing conscious? *Behavioral and Brain Sciences* 14:651-69.

Wheeler, J.A. 1990. Information, physics, quantum: The search for links. Em (W. Zurek, ed.) *Complexity, Entropy, and the Physics of Information*. Redwood City, CA: Addison-Wesley.

Wilkes, K.V. 1988. Yishi, Duh, Um and consciousness. Em (A. Marcel & E. Bisiach, eds.) *Consciousness in Contemporary Science*. Oxford: Oxford University Press.

David John Chalmers é um filósofo australiano, notabilizado por seus estudos em Filosofia da Mente. Chalmers é professor de filosofia e diretor do Centro de Estudos da Consciência da Universidade Nacional da Austrália, e publica artigos, ensaios e livros com regularidade. Ele cunhou a expressão "o problema difícil" da consciência.