

o enigma da experiência consciente

Estamos finalmente sondando um dos maiores mistérios da existência. Mas o conhecimento apenas do cérebro pode não ser suficiente para nos levar ao cerne do problema.

Por David J. Chalmers

Original: CHALMERS, DAVID JOHN (1995). “The puzzle of conscious experience”. *Scientific American* 273(6): 80-86 (dezembro). Reimpressão, com nova diagramação e pintura do Magritte, disponível online em: <http://consc.net/papers/puzzle.pdf>.

O presente texto é uma reimpressão, com pequenas modificações, da tradução feita em 2004 por Luís M.S. Augusto (Inst. de Filosofia, Universidade do Porto), em: <http://criticanarede.com/docs/chalmers.pdf>. Outra versão em português: “O enigma da consciência”. *Scientific American Brasil Especial 4, Segredos da Mente*, junho 2004, p. 40-49. Há também uma versão abreviada em: BonJour, L. & Baker, A. (orgs.) (2010). *Filosofia: textos fundamentais comentados*. Trad. M.C. Santos Rocha et al. Porto Alegre: Artmed, pp. 260-5.

Reimpressão preparada por Osvaldo Pessoa Jr. para a disciplina TCFC III – Filosofia das Ciências Neuras, FFLCH-USP, São Paulo, 2018. A paginação da versão original está indicada no texto, entre colchetes.



René Magritte, *Le double secret* (1927)

Consciência, a experiência subjetiva do eu interior, constitui um dos maiores desafios da neurociência. Mesmo um conhecimento detalhado do funcionamento do cérebro e dos correlatos neurais da consciência pode não conseguir explicar como ou por que os seres humanos têm mentes conscientes de si e do mundo.

[80]

A experiência consciente é ao mesmo tempo a coisa que melhor conhecemos no mundo e a mais misteriosa. Não há nada que conheçamos de forma mais direta do que a consciência, mas é incrivelmente difícil conciliá-la com o resto dos nossos conhecimentos. Por que é que existe? O que faz? Como é possível que nasça de processos neuronais no cérebro? Estas são algumas das questões mais intrigantes de toda a ciência.

De um ponto de vista objetivo, o cérebro é compreendido de forma razoável. Quando o leitor olha para esta página, um turbilhão de processamentos tem lugar: fótons atingem a sua retina, sinais elétricos são transmitidos para o seu nervo óptico e entre diferentes áreas do seu cérebro. Você pode até responder com um sorriso, com um semblante de perplexidade ou com um comentário qualquer. Mas há também um aspecto subjetivo. Quando o leitor olha para a página, está consciente dela, vivenciando [*experiencing*] diretamente as imagens e as palavras como parte da sua vida mental privada. Tem impressões vívidas das cores e das formas das imagens. Ao mesmo tempo, pode estar sentindo algumas emoções e formando alguns pensamentos. Todas juntas essas experiências constituem a consciência: a vida interior, subjetiva, da mente.

Durante muitos anos, os investigadores do cérebro e da mente evitaram deliberadamente a consciência. A visão predominante era a de que a ciência, dependente da objetividade, não podia incluir algo tão subjetivo como a consciência. O movimento behaviorista na psicologia, dominante em boa parte do século XX, concentrou-se no comportamento exterior e excluiu qualquer menção a processos mentais internos. Mais tarde, o desenvolvimento das ciências cognitivas focou a atenção nos processos que ocorrem no interior da cabeça. Contudo, a consciência continuava fora de alcance, prestando-se somente a ser um assunto de discussões pela noite afora por entre copos.

Nos últimos anos, contudo, um número cada vez maior de neurocientistas, psicólogos e filósofos tem rejeitado a ideia de que a consciência não pode ser estudada e está agora tentando explorar os seus segredos. Como é de esperar de uma área tão recente, deparamo-nos com um emaranhado de teorias diferentes e antagônicas, as quais por vezes se servem dos mesmos conceitos básicos de formas incompatíveis. Para ajudar a desemaranhar este problema é fundamental o raciocínio filosófico.

A multiplicidade de perspectivas dentro deste campo estende-se desde as teorias reducionistas, segundo as quais a consciência pode ser explicada pelos métodos convencionais da neurociência e da psicologia, até as posições chamadas “misterianas” [*mysterians*], que afirmam que jamais chegaremos a compreender a consciência. Penso que uma análise aturada pode mostrar que ambas as posições estão erradas e que a verdade se encontra em algum lugar a meio caminho entre ambas.

Contra o reducionismo argumentarei que os instrumentos das neurociências não podem fornecer uma descrição completa da experiência consciente, apesar do muito que têm para oferecer. Contra os misterianos defenderei que a consciência pode ser explicada por um novo tipo de teoria. Os pormenores de uma tal teoria ainda não estão disponíveis, mas um raciocínio cuidadoso e inferências bem feitas podem revelar algo sobre a sua natureza geral. Por exemplo, uma tal teoria envolverá provavelmente novas leis fundamentais e o conceito de informação pode vir a desempenhar um papel central.

Esta luz no fim do túnel sugere que a teoria da consciência pode vir a ter consequências espantosas no que diz respeito à nossa visão do universo e de nós próprios.

O Problema Difícil

Os pesquisadores usam a palavra “consciência” de muitas maneiras diferentes. De modo a esclarecer a questão, temos que primeiramente separar os problemas que se encontram geralmente associados sob o manto deste substantivo. Para tal penso que é útil distinguirmos entre os “problemas fáceis” e o “problema difícil” [*hard problem*] da consciência. [81] Os problemas fáceis não são de maneira alguma triviais – pelo contrário, representam um desafio tão grande quanto a maioria dos problemas da psicologia e da biologia – mas é no problema difícil que se encontra o principal mistério.

Entre os problemas fáceis da consciência incluem-se os seguintes: Como pode um sujeito humano discriminar estímulos sensoriais e reagir a eles de forma apropriada? Como é que o cérebro integra informação de muitas fontes diferentes e usa essa informação para controlar o comportamento? Como é que os indivíduos conseguem verbalizar os seus estados internos? Apesar de todas estas questões estarem relacionadas com a consciência, todas elas dizem respeito aos mecanismos objetivos do sistema cognitivo. Assim sendo, temos todas as razões para esperar que a investigação em curso em psicologia cognitiva e nas neurociências acabe por responder a estas questões.

O problema difícil, pelo contrário, é a questão de como é que processos físicos no cérebro dão origem à experiência subjetiva. Este mistério envolve o aspecto interno do pensamento e da percepção: o modo como as coisas são sentidas pelo sujeito. Quando vemos, por exemplo, vivenciamos sensações visuais tais como a de um azul vivo. Ou pensemos no inefável som de um oboé distante, no sofrimento causado por uma dor intensa, num instante de felicidade ou na qualidade meditativa de um momento perdido em pensamento. Tudo isto faz parte daquilo que chamo consciência. São estes fenômenos que constituem o verdadeiro mistério da mente.

De modo a ilustrar a distinção, recorramos a um experimento mental elaborado pelo filósofo australiano Frank Jackson. Suponhamos que Mary, uma neurocientista do século XXIII, é a especialista mundialmente mais reputada nos processos cerebrais responsáveis pela visão da cor. Mas Mary viveu toda a sua vida num quarto preto e branco [e cinza], e nunca viu quaisquer outras cores [Fig. 1]. [82] Ela sabe tudo o que há para saber sobre os processos físicos cerebrais – a sua biologia, estrutura e função. Este conhecimento permite-lhe compreender tudo o que há para saber acerca dos problemas fáceis: como é que o cérebro discrimina estímulos, integra informação e produz relatos verbais. A partir do seu conhecimento sobre a visão das cores, ela sabe o modo como os nomes das diferentes cores correspondem aos comprimentos de onda do espectro da luz. Mas há algo crucial acerca da visão da cor que Mary ainda não sabe: como é [*what it is like to*] vivenciar uma cor como o vermelho. Conclui-se que há fatos sobre a experiência consciente que não são dedutíveis dos fatos físicos acerca do funcionamento do cérebro.

A verdade é que ninguém sabe por que é que estes processos físicos são acompanhados por uma experiência consciente. Por que é que quando os nossos cérebros processam a luz de um certo comprimento de onda, temos a experiência de roxo profundo? Por que é que temos qualquer experiência? Não poderia um autômato inconsciente realizar as mesmas operações tão bem quanto nós? Estas são perguntas às quais gostaríamos que a teoria da consciência pudesse responder.



[Figura 1] *Uma neurocientista isolada em uma sala preta e branca sabe tudo sobre como o cérebro processa as cores, mas não sabe como é vê-las. Essa situação sugere que o conhecimento do cérebro não fornece um conhecimento completo da experiência consciente.*

Não nego que a consciência tenha origem no cérebro. Sabemos, por exemplo, que a experiência subjetiva da visão está intimamente ligada a processos no córtex visual [Fig. 2]. Mas é exatamente esta ligação que nos deixa perplexos. Notavelmente, a experiência subjetiva parece emergir de um processo físico. Mas não fazemos a mínima ideia do como ou do porquê disto ser assim.

A Neurociência é Suficiente?

Se olharmos para a profusão de trabalhos recentes sobre a consciência levados a cabo na neurociência e na psicologia, podemos pensar que o mistério está prestes a ser desvendado. Contudo, uma análise mais atenta mostra-nos que quase todo o trabalho em curso se centra somente nos problemas fáceis da consciência. A confiança da perspectiva reducionista advém dos progressos alcançados no que diz respeito aos problemas fáceis, mas nada disto faz qualquer diferença no que diz respeito ao problema difícil.

Considere-se a hipótese avançada pelos neurobiólogos Francis Crick, do Salt Institute for Biological Studies em San Diego, e Christof Koch, do California Institute of Technology. Eles sugerem que a consciência pode ter origem em certas oscilações no córtex cerebral que se sincronizam quando os neurônios disparam 40 vezes por segundo. Crick e Koch acreditam que este fenômeno pode explicar como é que diferentes propriedades de um mesmo objeto percebido (a sua cor e a sua forma, por exemplo), processadas em diferentes partes do cérebro, se fundem num todo coerente. De acordo com esta teoria, dois blocos de informação tornam-se integrados [*bound*] precisamente quando são representados por disparos neuronais sincronizados.

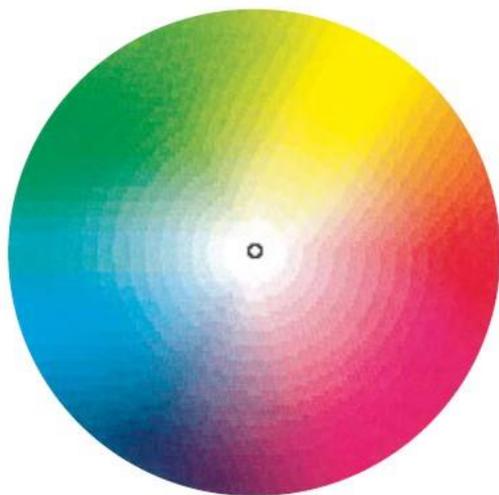
Esta hipótese poderia esclarecer de forma plausível um dos problemas fáceis relativos ao modo como a informação é integrada no cérebro. Mas por que é que oscilações sincronizadas deveriam dar origem a uma experiência visual, qualquer que

seja a quantidade de integração que ocorre? Esta questão envolve o problema difícil, para o qual a teoria em causa nada tem para oferecer. De fato, Crick e Koch são agnósticos em relação à questão de se o problema difícil pode ser resolvido pela ciência.

O mesmo tipo de crítica pode aplicar-se a quase todo o trabalho recente sobre a consciência. Nos seu livro de 1991, *Consciousness Explained* [A Consciência Explicada, não traduzido], o filósofo Daniel C. Dennett expôs uma teoria sofisticada sobre o modo como numerosos processos independentes no cérebro se combinariam entre si para produzir uma resposta coerente a um evento percebido. A teoria pode ser útil para explicar como é que produzimos relatos verbais acerca dos nossos estados internos, mas nos diz muito pouco acerca da questão de por que deveria haver uma experiência subjetiva por trás destes relatos. Como outras teorias reducionistas, a teoria de Dennett é uma teoria dos problemas fáceis.

A principal característica que estes problemas fáceis têm em comum é a de que todos eles tratam do modo como se desempenha uma função cognitiva ou comportamental. Em última instância, todos estes problemas se reduzem a questões acerca do modo como o cérebro desempenha determinadas tarefas – como discrimina estímulos, integra informação, produz relatos e por aí fora. Assim que a neurobiologia especificar os mecanismos neuronais apropriados, mostrando o modo como as funções se desempenham, os problemas fáceis estarão resolvidos.

O problema difícil da consciência, pelo contrário, está para além dos problemas relativos ao modo como as funções se desempenham. Mesmo que cada uma das funções cognitivas e comportamentais relacionadas com a consciência fosse explicada, restaria ainda um mistério adicional por explicar: por que é que o desempenho destas funções é acompanhado por uma experiência consciente? É este enigma suplementar que faz com que o problema difícil seja difícil.



[Figura 2] O disco de cores dispõe os matizes de tal forma que aqueles vivenciados como semelhantes estão mais próximos. As cores próximas correspondem também a representações perceptivas semelhantes no cérebro.

A Lacuna Explicativa

Há quem tenha sugerido que para resolvermos o problema difícil temos que introduzir novos instrumentos de explicação física: dinâmica não-linear, por exemplo, ou as novas descobertas das neurociências, ou a mecânica quântica. Mas estas ideias sofrem exatamente da mesma dificuldade. Tomemos uma proposta de Stuart R. Hameroff, da Universidade do Arizona, e de Roger Penrose, da Universidade de Oxford. Eles defendem que a consciência tem origem em processos quânticos que

ocorrem em microtúbulos, estruturas de proteína no interior de neurônios. É possível (senão mesmo provável) que uma tal hipótese possa levar a uma explicação do modo como o cérebro toma decisões ou até do modo como ele prova teoremas matemáticos, como sugerem Hameroff e Penrose. Mas mesmo que consiga alcançar este feito, a teoria nada tem a dizer sobre o modo como estes processos podem dar origem à experiência consciente. De fato, o mesmo problema surge em qualquer teoria da consciência que se baseie exclusivamente em processamento físico.

O problema é que as teorias físicas servem para explicar por que os sistemas têm uma determinada estrutura física e como eles desempenham as suas várias funções. A maior parte dos problemas científicos tem esta forma; para explicar a vida, por exemplo, precisamos descrever o modo como um sistema físico pode se reproduzir, adaptar e fazer o seu metabolismo. Mas a consciência é um tipo de problema completamente diferente, na medida em que está para além da explicação científica da estrutura e da função.

Certamente que a neurociência não é irrelevante para o estudo da consciência. Quanto mais não seja, podem ser capazes de revelar a natureza do correlato neural da consciência – os processos cerebrais mais diretamente associados à experiência consciente [Fig. 3]. Pode mesmo chegar a fornecer uma correspondência pormenorizada entre processos específicos no cérebro e componentes da experiência com eles relacionados. Mas antes de sabermos porque é que esses processos dão origem à experiência consciente, não teremos ultrapassado aquilo que o filósofo Joseph Levine denominou [83] a lacuna explicativa [*explanatory gap*] entre os processos físicos e a consciência. Para dar esse salto é necessário um novo tipo de teoria.

Uma Verdadeira Teoria de Tudo

Ao procurarmos uma alternativa, uma observação chave é que nem todas as entidades da ciência são explicáveis em termos de entidades mais básicas. Na física, por exemplo, o espaço-tempo, a massa e a corrente elétrica (entre outras coisas) são considerados características [*features*] fundamentais do mundo, dado que não são redutíveis a nada mais simples. Apesar desta irreduzibilidade, há teorias pormenorizadas e úteis que relacionam estas entidades umas com as outras em termos de leis fundamentais. Em conjunto, estas características e estas leis explicam uma grande variedade de fenômenos complexos e sutis.

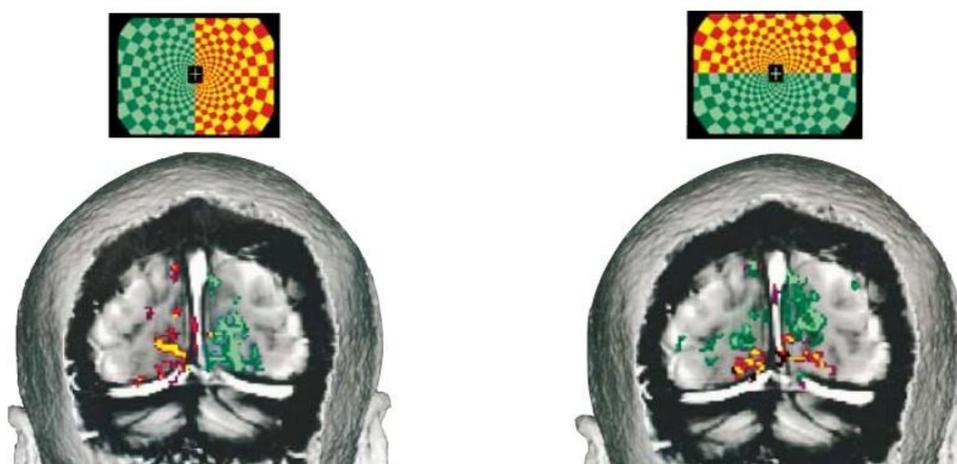
Muita gente partilha da crença de que a física fornece um catálogo completo das características e das leis fundamentais do universo. Como o físico Steven Weinberg diz na sua obra de 1992, *Dreams of a Final Theory* [*Sonhos de uma Teoria Final*, Rocco, 1996], o objetivo da física é uma “teoria de tudo” a partir da qual tudo o que se sabe acerca do universo pode ser derivado. Contudo, Weinberg reconhece que a consciência apresenta problemas especiais. Apesar do poder explicativo da teoria física, a existência da consciência não parece ser derivável das leis físicas. Ele defende a física ao argumentar que ela acabará por explicar aquilo que ele chama os correlatos objetivos da consciência (ou seja, os correlatos neurais), mas é claro que isto não equivale a explicar a própria consciência. Se a existência da consciência não é derivável das leis da física, então uma teoria da física não é uma verdadeira teoria de tudo. Logo, uma teoria final tem que conter um componente fundamental adicional.

De modo a atingirmos esse objetivo, proponho que se considere a experiência consciente como uma característica fundamental, irreduzível a algo mais básico. Esta ideia pode parecer inicialmente estranha, mas é exigida por questões de consistência. Descobriu-se no século XIX que os fenômenos eletromagnéticos não podiam ser

explicados em termos de princípios já conhecidos. A consequência disto foi que os cientistas introduziram a carga eletromagnética [elétrica] como uma nova entidade fundamental e passaram a estudar as leis fundamentais que lhe estavam associadas. Devíamos aplicar o mesmo tipo de raciocínio no que diz respeito à consciência. Se as teorias existentes fundamentais não conseguem englobá-la, então precisamos de algo novo.

Onde há propriedades fundamentais, há leis fundamentais. Neste caso, as leis devem relacionar a experiência com elementos da teoria física. É quase certo que estas leis não irão interferir com as leis do mundo físico; parece que estas últimas formam um sistema fechado de pleno direito. Pelo contrário, as leis irão servir como uma ponte, especificando o modo como a experiência depende de processos físicos subjacentes. É esta a ponte que irá cobrir a lacuna explicativa.

A ser assim, uma teoria completa terá dois componentes: leis físicas que exprimem o comportamento dos sistemas físicos do nível infinitesimal ao nível cosmológico, e aquilo que poderemos chamar leis psicofísicas, as quais nos dirão como é que alguns desses sistemas estão associados à experiência consciente. São estes dois componentes que irão constituir uma verdadeira teoria de tudo.



[Figura 3] As variações de fluxo sanguíneo no córtex visual demonstram como o cérebro de uma pessoa reage ao padrão que está sendo visualizado. Nesta imagem, as cores representam a atividade cortical correspondente à visão que a pessoa tem da metade vertical ou da metade horizontal da ilustração. O experimento pode iluminar um correlato neural da experiência visual (Fonte: Universidade de Pittsburgh).

À Procura de uma Teoria

Supondo, por enquanto, que elas existem, como é que podemos descobrir essas leis psicofísicas? O maior obstáculo para a consecução deste objetivo é a falta de dados. Do modo como a descrevi, a consciência é subjetiva, de modo que não temos nenhum meio de a controlar nos outros. Contudo, isto não representa um beco sem saída, mas tão somente um obstáculo. Para começar, cada um de nós tem acesso às suas próprias experiências, o que representa um manancial para a formulação de teorias. Podemos também confiar de forma plausível em informação indireta como, por exemplo, as descrições que os sujeitos fazem das suas experiências. Argumentos filosóficos e experimentos mentais também têm um papel a desempenhar. É certo que estes métodos

apresentam limitações, mas oferecem-nos mais do que o suficiente para deitarmos mãos à obra.

Estas teorias não serão testáveis de forma conclusiva, [84] pelo que serão inevitavelmente mais conjecturais do que as teorias das disciplinas científicas mais convencionais. Contudo, não há nenhuma razão para que elas não possam ser fortemente restringidas, de modo a poderem dar conta de modo preciso quer das nossas experiências na primeira pessoa, quer dos testemunhos dos relatos por parte de sujeitos. Caso encontremos uma teoria que explique melhor os dados do que uma outra teoria qualquer de igual simplicidade, então temos boas razões para a adotarmos. Neste preciso momento nem sequer temos uma teoria que se encaixe aos dados que possuímos, pelo que quaisquer preocupações com a testabilidade são prematuras.

Podemos começar por procurar leis de ponte [*bridging laws*] de nível superior, que ligariam processos físicos à experiência em nível cotidiano. Os contornos básicos de uma tal lei podem ser esboçados a partir da observação do fato de que quando estamos conscientes de algo, somos geralmente capazes de agir de acordo com isso ou de falar sobre isso – e estas são funções físicas, objetivas. Por outro lado, quando há informação diretamente disponível para a ação e para a linguagem, essa informação é geralmente consciente. Assim, a consciência correlaciona-se bem com aquilo que podemos chamar de “estar ciente” ou “ter ciência” [*awareness*]: o processo pelo qual informação no cérebro se torna globalmente disponível para os processos motores tais como a linguagem e a ação corporal.¹

A noção pode parecer trivial. Mas tal como a defino aqui, terciência é objetiva e física, o mesmo não se passando com a consciência. Precisamos refinar a definição de terciência de modo a podermos estender o conceito aos animais e aos bebês, ambos incapazes de usar a linguagem. Mas podemos, pelo menos nos casos que conhecemos melhor, ver os contornos de uma lei psicofísica: sempre que há terciência, há consciência, e vice-versa.

De modo a avançarmos nesta linha de pensamento, debruçemo-nos sobre a estrutura presente na experiência consciente. A experiência de um campo de visão, por exemplo, é um mosaico de cores, formas e padrões em constante mutação e, como tal, possui uma estrutura geométrica detalhada. O fato de podermos descrever esta estrutura, reconhecer muitos dos seus constituintes e realizar outras ações que dependam dessa estrutura sugere que ela corresponde diretamente à estrutura da informação tornada disponível no cérebro através dos processos neurais da terciência.

De modo semelhante, as nossas experiências da cor têm uma estrutura tridimensional intrínseca que se espelha na estrutura dos processos de informação no córtex visual. Esta estrutura é representada pelos círculos e tabelas de cores usadas pelos pintores. As cores ocupam uma posição de acordo com um padrão sistemático: do vermelho ao verde num eixo, do azul ao amarelo no outro, e do preto ao branco num terceiro eixo. As cores próximas umas das outras num círculo de cores são vivenciadas como semelhantes (ver Fig. 2). É mais que provável que elas também correspondam a representações perceptivas semelhantes no cérebro, como parte de um sistema de complexa codificação tridimensional entre neurônios, um sistema que ainda não

¹ O *Dicionário Inglês-Português* editado por Houaiss traduz “*aware*” por “ciente”. Em inglês, “*awareness*” é sinônimo de “*state of being aware of something*”, que se traduziria por “estado de estar ciente de algo” ou, simplificando, “estar ciente”, que é sinônimo de “ter ciência” (que é mais ambíguo, pois pode também ser entendido como a posse de conhecimento científico, que não é o sentido pretendido aqui). Para formar uma palavra única em português, teríamos “estar-ciente” ou “ter-ciência”. Destes, o neologismo *terciência* será adotado, rimando com “consciência” e “senciência”. Notamos que Chalmers toma emprestado o termo *awareness* para um uso incomum, para o correlato físico da consciência, então mais uma razão para adotarmos este neologismo.

compreendemos completamente. Podemos reformular o conceito subjacente como um princípio de coerência estrutural: a estrutura da experiência consciente espelha-se na estrutura da informação na terciência e vice-versa.

Um outro candidato a uma lei psicofísica é um princípio de invariância organizacional. Segundo este princípio, sistemas físicos com a mesma organização abstrata darão origem ao mesmo tipo de experiência consciente, independentemente daquilo de que são feitos. Por exemplo, [85] se as interações precisas entre os nossos neurônios pudessem ser duplicadas com chips de silício, o mesmo tipo de consciência surgiria. Esta ideia é algo controversa, mas penso que ela é fortemente apoiada por experimentos mentais que descrevem a substituição gradual de neurônios por *chips* de silício. A implicação notável é que a consciência pode um dia vir a ser obtida em máquinas (ver o Apêndice 1).

Informação: Física e Vivencial

O objetivo último de uma teoria da consciência é alcançar um conjunto simples e elegante de leis fundamentais, análogas às leis fundamentais da física. Contudo, os princípios acima descritos provavelmente não são fundamentais. Parecem ser, pelo contrário, leis psicofísicas de nível superior análogas aos princípios macroscópicos na física, como os princípios da termodinâmica e da cinemática. Quais poderão ser as leis fundamentais subjacentes? Ninguém sabe, mas eu não me importo em especular.

Proponho que as leis psicofísicas primárias poderão envolver de modo central o conceito de informação. A noção abstrata de informação, tal como foi proposta nos anos 1940 por Claude E. Shannon, do Massachusetts Institute of Technology, é a de um conjunto de estados separados com uma estrutura básica de similaridades e diferenças entre si. Podemos conceber um código binário de 10 bits como um estado informacional, por exemplo. Esses estados informacionais podem ser concretizados no mundo físico. Isto acontece sempre que eles correspondem a estados físicos (voltagens, por exemplo); as diferenças entre eles podem ser transmitidas ao longo de um percurso como as linhas de telefone.

Podemos também encontrar informação concretizada na experiência consciente. O padrão de manchas de cor num campo visual, por exemplo, pode ser visto como sendo análogo ao padrão dos *pixels* que cobrem uma tela. De modo intrigante, encontramos os mesmos estados informacionais presentes na experiência consciente e nos processos físicos do cérebro. A codificação tridimensional dos espaços de cor, por exemplo, sugere que o estado informacional numa experiência de cor corresponde diretamente a um estado informacional no cérebro. Podemos mesmo encarar os dois estados como aspectos distintos de um único estado informacional que é simultaneamente concretizado no processamento físico e na experiência consciente.

Segue-se daí uma hipótese natural. Talvez a informação, ou pelo menos alguma informação, apresente dois aspectos básicos: um aspecto físico e um aspecto vivencial [*experiential*]. Esta hipótese tem o estatuto de um princípio fundamental que pode subjazer à relação entre processos físicos e experiência. Onde quer que seja que encontremos experiência consciente, ela existe como um aspecto de um estado informacional, sendo que o seu outro aspecto está presente num processo físico no cérebro. Esta proposta precisa ser desenvolvida para gerar uma teoria satisfatória, Mas ela ajusta-se bem aos princípios mencionados acima – sistemas com a mesma organização concretizarão a mesma informação, por exemplo – e pode explicar muitas das características da nossa experiência consciente.

Esta teoria é pelo menos compatível com muitas outras, por exemplo, com a hipótese do físico John A. Wheeler de que a informação é fundamental para a física do universo. As leis da física poderão [86] vir a ser enunciadas em termos informacionais, o que permitiria uma congruência satisfatória entre as construtos tanto das leis da física quanto das leis psicofísicas. Pode até mesmo vir a dar-se o caso de uma teoria da física e uma teoria da consciência acabarem por se consolidarem numa única teoria da informação mais abrangente.

A ubiquidade da informação pode representar um potencial problema. Até mesmo um termostato substancia alguma informação, por exemplo, mas estará ele consciente? Há pelo menos duas respostas possíveis para esta questão. Em primeiro lugar, podíamos restringir as leis fundamentais de modo a que somente alguma informação tenha um aspecto vivencial, talvez dependente do modo como é processada fisicamente. Em segundo lugar, podemos agarrar o touro pelos cornos [*bite the bullet*] e deixar que toda a informação tenha um aspecto vivencial – onde ocorrer processamento de informação complexa há experiência complexa, e onde ocorrer processamento de informação simples há experiência simples. A ser assim, até mesmo um termostato poderia ter experiências, mesmo sendo bastante mais simples do que até mesmo a mais básica experiência de cor, e certamente não seriam acompanhadas de emoção ou pensamento. Isto parece estranho à primeira vista, mas se a experiência [subjéctiva] é verdadeiramente fundamental, então pode bem dar-se o caso que ela seja amplamente difundida. Seja como for, a escolha entre estas alternativas deveria depender da questão de qual delas pode ser integrada na teoria melhor de que dispomos.

É claro que tais teorias podem estar completamente erradas. Por outro lado, podem muito bem vir a desenvolver-se até formarem uma teoria mais forte, capaz de prever a estrutura precisa da nossa experiência consciente a partir de processos físicos nos nossos cérebros. No caso de este projeto ser bem sucedido, teremos boas razões para aceitar a teoria. Caso contrário, se ela falhar, outros caminhos serão percorridos e outras teorias fundamentais alternativas serão formuladas. Deste modo pode bem ser que um dia possamos chegar a resolver o maior mistério da mente.

Apêndice 1: Qualia Bailando num Cérebro Sintético

Uma questão que muita gente acha verdadeiramente fascinante é a da possibilidade de a consciência poder surgir de um sistema sintético complexo. Embora estejamos a décadas ou até mesmo a séculos de distância da construção de um tal sistema, temos um experimento mental simples que oferece bons indícios de que um cérebro artificial, se organizado de modo apropriado, teria o mesmo tipo de experiências conscientes de um ser humano.

Imaginemos um sistema baseado em silício no qual os chips estejam organizados e funcionam do mesmo modo que os neurônios em seu cérebro. Ou seja, cada chip no sistema de silício faz exatamente aquilo que o seu análogo natural faz, e está interligado com os elementos circundantes precisamente do mesmo modo. Deste modo, o comportamento do sistema artificial será exatamente o mesmo que o nosso. A pergunta fulcral é: Estará ele consciente do mesmo modo que nós?

Suponhamos, para argumentar, que não. (Usamos aqui um método de raciocínio conhecido como *reductio ad absurdum*, no qual se assume a hipótese contrária e se demonstra que ela leva a uma conclusão insustentável.) Ou seja, ou esse sistema tem experiências diferentes – por exemplo, uma experiência da cor azul enquanto nós vemos a cor vermelha – ou pura e simplesmente não tem qualquer experiência. Consideraremos o primeiro caso; o raciocínio é semelhante em ambos os casos.

Visto que os chips e os neurônios têm a mesma função, eles são intercambiáveis, usando interfaces apropriadas. Assim sendo, os chips podem substituir os neurônios, produzindo uma sequência contínua de casos nos quais uma proporção sucessivamente maior de neurônios é substituída por chips. Ao longo desta sequência, a experiência consciente do sistema irá também se modificar. Por exemplo, poderíamos substituir todos os neurônios do nosso córtex visual por uma versão feita de silício, com idêntica organização. O cérebro daí resultante, possuindo um córtex visual artificial, terá uma experiência consciente diferente da do cérebro original: onde você anteriormente via vermelho, agora vivenciará o roxo (ou talvez um cor de rosa pálido, no caso de o sistema integralmente feito de silício não ter qualquer experiência).

Ambos os córtices visuais são então ligados aos nossos cérebros por meio de um interruptor de duas posições. Com o interruptor numa posição usamos o nosso córtex visual natural e com o interruptor na outra posição ativamos o córtex visual artificial. Quando se muda a posição do interruptor, a nossa experiência passa de vermelho a roxo ou vice-versa. Se mudarmos as posições do interruptor repetidamente, as nossas experiências “bailarão” entre os dois estados conscientes diferentes (vermelho e roxo), conhecidos como *qualia*.

Contudo, visto que a organização do nosso cérebro não se alterou, não poderia haver uma alteração comportamental quando se muda a posição do interruptor. Logo, quando lhe perguntarem o que está vendo, você dirá que nada mudou. Você dirá que está vendo vermelho e que não viu nada além do vermelho – ainda que as duas cores bailem em frente aos seus olhos. Esta conclusão é tão irrazoável que será melhor que seja concebida como um *reductio ad absurdum* da hipótese original – que um sistema artificial com organização e funcionamento idênticos tem uma experiência consciente diferente daquela de um cérebro neural. Voltarmos atrás na nossa hipótese estabelece o contrário: que sistemas com a mesma organização têm a mesma experiência consciente.

[Figura 4] *Em um experimento mental, uma maçã pode se alternar do vermelho para o azul.*

