

A TEORIA DA CONSCIÊNCIA DE DAVID CHALMERS

João de Fernandes Teixeira

*Departamento de Filosofia, Universidade Federal de S. Carlos.
Grupo de Ciência Cognitiva, Instituto de Estudos Avançados da USP*

*O artigo tem por objetivo apresentar e discutir a teoria da consciência elaborada pelo filósofo David Chalmers no seu livro *The Conscious Mind*, publicado em 1996. O artigo é dividido em duas partes. A primeira expõe os principais delineamentos da teoria de Chalmers; a segunda discute seus principais conceitos, abordando a plausibilidade metafísica da existência dos “zumbis” e a idéia de superveniência.*

Descritores: Consciência. Inteligência artificial. Cartesianismo. Cognição.

Num artigo publicado em 1978, o filósofo Daniel Dennett observou que a questão da natureza da consciência constitui o problema mais difícil a ser enfrentado pela Filosofia da Mente, a parte da ciência da mente que mais tem resistido ao estudo, “the last bastion of occult properties, epiphenomena, immeasurable subjective states - in short, the one area of mind best left to the philosophers, who are welcome to it.” (Dennett, 1978, p.149). Não existe nada mais imediato do que a experiência consciente mas ao mesmo tempo não existe nada tão difícil a ser explicado.

No panorama da Filosofia da Mente e da Ciência Cognitiva a questão da natureza da consciência começa a ocupar lugar central nas pesquisas a partir do final da última década, após um longo e deliberado silêncio sobre esta questão por parte dos filósofos da mente e dos estudiosos de Inteligência Artificial. Marcos do reaparecimento de uma preocupação crescente com a questão da natureza da consciência são os estudos de Jackendoff (1987), Calvin (1990), Dennett (1991) e Flanagan

(1992). Estes trabalhos procuraram desmistificar a noção de consciência e situá-la, seja no âmbito de teorias cognitivistas, seja no âmbito das neurociências. Tentava-se mostrar que este fenômeno é suscetível de ser tratado pelos métodos tradicionais da Ciência Cognitiva, através de teorias computacionais ou através do estudo de mecanismos neurais. Sentia-se a necessidade de formular uma *teoria* da consciência e não apenas de curvar-se diante da perplexidade dos problemas envolvidos no estudo da natureza dos estados conscientes.

É neste contexto que se insere o livro de D. J. Chalmers, “The Conscious Mind”, talvez a tentativa mais recente de se formular uma teoria abrangente da natureza da consciência. Sua teoria é ousada e corre na direção oposta a tudo o que os cientistas cognitivos e neurocientistas desejam: reduzir estados conscientes a uma base neurofisiológica ou física.

Chalmers toma como ponto de partida aquilo que para muitos (aí incluídos até alguns neurocientistas) constitui o horizonte intransponível de qualquer teoria científica da natureza da consciência: reconhecer que não é possível formular uma teoria que explique plenamente como um sinal cerebral pode dar origem a um estado consciente.¹ Em outras palavras, a consciência deve ser o ponto de partida, e não o ponto de chegada de qualquer teoria da mente; uma perspectiva que converge com as teorias físicas contemporâneas nas quais o psiquismo ou a mente do observador emerge como um elemento necessário para explicar o comportamento da natureza.² Neste sentido, Chalmers sugere que uma teoria da

1 Este é o chamado “problema da geração”, aparentemente formulado pela primeira vez por John Tyndall que afirmava que “The passage from the physics of the brain to the corresponding facts of consciousness is unthinkable. Granted that a definite thought and a definite molecular action in the brain occur simultaneously, we do not possess the intellectual organ, nor apparently any rudiment of the organ which would enable us to pass, by a process of reasoning, from one to the other.” (citado por James, 1890, p.147). O mesmo ponto de vista parece ser sustentado por alguns filósofos contemporâneos, como, por exemplo, McGinn (1989), que argumenta que este é um problema excessivamente complexo para nossas mentes, na medida em que a experiência consciente está fora do escopo de qualquer teoria científica.

2 Num artigo publicado em 1980 H. J. Morowitz observava que “first, the human mind, including consciousness and reflective thought, can be explained by activities

consciência deve tomar a noção de *experiência consciente* como sendo um primitivo. Uma teoria da consciência requer a adição de algo fundamental à nossa ontologia, na medida em que tudo em teoria física é compatível com a ausência de consciência. A experiência consciente deve ser considerada como sendo uma característica fundamental do mundo, do mesmo jeito que massa, carga eletromagnética e espaço-tempo.

Muitos fenômenos são explicáveis em termos de entidades mais simples do que eles, mas isto não é universal. As vezes certas entidades precisam ser tomadas como primitivas ou *fundamentais*. Entidades fundamentais não podem ser explicadas em termos de algo mais simples. Por exemplo, no século XIX ficou claro que processos eletromagnéticos não poderiam ser explicados em termos de processos mecânicos. Diante disto, Maxwell introduziu as noções de carga e força eletromagnética como componentes fundamentais de sua teoria física. Ou seja, para explicar o eletromagnetismo a ontologia da física teve de ser expandida. Outras características que a teoria física assume como fundamentais são as noções de massa e de espaço-tempo. Nunca se procurou explicar estas noções em termos de algo mais simples, o que entretanto não descarta a possibilidade de se construir uma teoria a partir dos conceitos de massa ou de espaço-tempo.

Esta posição é uma variedade de dualismo, na medida em que ela postula propriedades básicas além daquelas estipuladas pela física. Mas trata-se de uma variedade inocente de dualismo, inteiramente compatível com uma visão científica do mundo. Como assevera Chalmers, não há nada místico ou espiritual nesta teoria. É uma teoria inteiramente naturalista, na medida que, segundo ela, o universo não é nada mais do que uma

of the central nervous system, which, in turn, can be reduced to the biological structure and function of that physiological system. Second, biological phenomena at all levels can be totally understood in terms of atomic physics, that is, through the action and interaction of the component atoms of carbon, nitrogen, oxygen and so forth. Third and last, atomic physics, which is now understood most fully by means of quantum mechanics, must be formulated with the mind as a primitive component of the system.” (p.39).

rede de entidades básicas que obedecem um conjunto de leis e a consciência pode ser explicada a partir destas. Trata-se de um *dualismo naturalista*.

O dualismo naturalista permite desenvolver uma teoria não-reducionista da consciência que consistirá de um conjunto de *princípios psicofísicos* ou seja, princípios que conectam propriedades de processos físicos com propriedades da experiência. Podemos pensar nestes princípios como englobando a maneira pela qual a experiência consciente emerge da estrutura física. Em última análise, esses princípios devem nos dizer que tipo de sistemas físicos podem gerar experiências e, no caso de sistemas que o fazem, eles devem nos dizer que tipo de propriedades físicas são relevantes para a emergência da experiência consciente.

A defesa deste ponto de vista orienta o modo pelo qual Chalmers estrutura seu livro: num primeiro momento, é preciso reconhecer a verdadeira dimensão do problema da consciência, desvinculando-o de um conjunto de problemas subsidiários que podem ocultar ou escamotear a sua identificação adequada. O segundo momento, consiste em atacar as explicações funcionais e reducionistas da consciência, e mostrar em que sentido estas podem ser necessárias mas não suficientes para dar conta da natureza específica da experiência consciente. A terceira parte do livro esboça uma teoria geral da consciência com base num conjunto de princípios psicofísicos.

O reconhecimento do problema da consciência significa sustentar que este não é um pseudo-problema e que o filósofo da mente não pode fugir da tarefa de ter de enfrentá-lo seriamente. Esta tentação pode surgir pelo fato de estarmos enfrentando um problema extremamente árduo. Para começar, a Filosofia da Mente não reconhece a existência de apenas um problema da consciência. “Consciência” é um termo polissêmico e por vezes ambíguo, que se refere a vários tipos de fenômenos, como por exemplo:

- a habilidade para discriminar, categorizar e reagir a estímulos ambientais,
- a integração da informação através de um sistema cognitivo,
- a capacidade de relatar a ocorrência de estados mentais,

- a habilidade de um sistema para acessar seus próprios estados internos,
- o foco da atenção,
- o controle deliberado do comportamento,
- a diferença entre sono e vigília.

Todos estes fenômenos estão associados com a noção de consciência. Por exemplo, diz-se que um estado mental é consciente quando ele é passível de ser relatado verbalmente ou quando ele é internamente acessível. As vezes, diz-se que um sistema está consciente de uma informação quando ele tem a habilidade de reagir com base nela ou quando ele a integra e a elabora para produzir determinados comportamentos. Dizemos freqüentemente que uma ação é consciente porque ela é deliberada. Outras vezes, referimo-nos a um organismo como estando consciente quando este está em vigília.

No entender de Chalmers nenhum destes fenômenos - nem tampouco seu conjunto - caracteriza o verdadeiro problema da consciência: eles constituem apenas os aspectos *funcionais* da experiência consciente. Isto significa dizer que, em última análise, estes fenômenos *podem vir a ser* explicados cientificamente. Em outras palavras, nada impede que algum dia eles possam vir a ser explicados seja através de um modelo computacional seja através da descoberta de mecanismos neurais. Por exemplo, para explicar o acesso e a capacidade de relatar a ocorrência de estados mentais, basta especificar o mecanismo através do qual a informação acerca de estados mentais é recuperada e tornada disponível para relato verbal. Para explicar a integração da informação precisamos apenas conceber mecanismos através dos quais esta seja combinada e em seguida utilizada em outros processos. Para explicar a distinção entre sono e vigília uma explicação em termos neurofisiológicos que dê conta da diferença de comportamento do organismo nestes dois estados é mais do que suficiente.

Se explicar a consciência se resumisse a explicar estes fenômenos, então não haveria um *problema filosófico* da consciência. Embora estes sejam problemas empíricos de difícil solução, eles ainda não caracteri-

zam os verdadeiros problemas colocados pela consciência. Estes são, em última análise, os “easy problems”

A grande dificuldade é o chamado problema da *experiência* (“hard problem”). Quando pensamos e percebemos o mundo existe um tipo de processamento de informação mas também um aspecto subjetivo nele envolvido. Como Nagel (1974) coloca, existe “something it is like to be a conscious organism.” Este aspecto subjetivo é a *experiência consciente*. Como caracterizar a experiência consciente? O que significa ter uma imagem mental neste momento ou experimentar uma sensação corporal qualquer? O que unifica tudo isto? A experiência emerge de uma base física mas não sabemos como isto é possível. Como algo físico pode dar lugar a experiências internas ou estados internos?

O reconhecimento da existência de um “hard problem” tem como consequência uma desqualificação das tentativas de explicação funcional da natureza da consciência entendida como *experiência consciente*. Explicações funcionais podem ser necessárias, mas certamente não serão suficientes para explicar a natureza da experiência consciente. Pois, como explicamos o desempenho de uma função? Especificando *o mecanismo que desempenha a função*. A aplicação de conhecimentos oriundos da neurofisiologia e das ciências cognitivas pode resolver vários problemas neste sentido. Se mostrarmos como um mecanismo neuronal ou computacional pode desempenhar uma determinada tarefa, teremos explicado o fenômeno em questão.

Mas no caso da *experiência consciente* este tipo de explicação falha. O problema da experiência consciente requer algo mais do que explicar o desempenho de funções. Em outras palavras, o “hard problem” persiste mesmo quando o desempenho de todas as funções relevantes é explicado. A questão que se coloca é a seguinte: *Por que o desempenho destas funções é acompanhado por experiências?* Ou seja, pode-se explicar como a informação é discriminada, integrada e relatada, mas isto não significa explicar como ela é *experienciada*. Esta é a questão chave no problema da consciência - explicar como e porque surge a experiência no decorrer do processamento de informação. Não existe nenhuma função cognitiva cuja explicação leve automaticamente à uma explicação da

experiência consciente. A experiência consciente *supervem* a sua base física, ou seja, nenhum fato do mundo, mesmo a nível microfísico, implica necessariamente na produção de estados conscientes.³

O conceito de *superveniência*, cuidadosamente analisado por Chalmers em seu livro sustenta este ponto de vista. Uma propriedade *B* de um determinado indivíduo é chamada de superveniente se é produzida por um conjunto de propriedades *A* desse mesmo indivíduo. Por exemplo, um conjunto de propriedades físicas pode determinar um conjunto de propriedades biológicas na medida em que fenômenos vitais dependem de uma base física. Estes fenômenos vitais são então *supervenientes* em relação a sua base física; se as propriedades físicas variarem, as propriedades biológicas também variarão. A determinação de propriedades supervenientes pode ser *lógica* (conceitual) ou *natural* (empírica ou nômica). No caso da superveniência lógica as propriedades *B* são consequência automática da existência das propriedades *A*, ou seja, não seria possível conceber *A* sem conceber *B*. Já no caso da superveniência natural é possível conceber *A* sem conceber *B*, mas existe uma conexão empírica, *de fato*, entre *A* e *B*.

Ora, o esforço de Chalmers será mostrar que estados conscientes não são logicamente supervenientes em relação a estados físicos: é perfeitamente concebível a existência de duas criaturas fisicamente idênticas sendo que uma desenvolve experiências conscientes e outra não. O exemplo paradigmático invocado por Chalmers é a plausibilidade de concebermos criaturas como zumbis. Neste experimento mental⁴, um zumbi é uma criatura fisicamente idêntica a mim, molécula por molécula. Ele é também funcionalmente equivalente a mim, no sentido de que ele

3 “that is, that all the microphysical facts in the world do not entail the facts about consciousness.” (Chalmers, 1996, p.93).

4 Um experimento mental, figura frequentemente utilizada na literatura da filosofia da mente, consiste em imaginar uma situação hipotética, algo que teoricamente pode vir a ser realizado mas que não contraria possibilidades físicas e lógicas. A importância dos experimentos mentais consiste no fato de que destas situações hipotéticas podemos extrair imediatamente consequências conceituais importantes.

pode fazer tudo o que eu faço. Contudo, posso perfeitamente conceber que este zumbi não tenha experiências conscientes. Este zumbi pode ser até uma réplica de mim mesmo, mas replicar minhas características físicas e funcionais não implica, automaticamente, em replicar minha possibilidade de ter estados conscientes. O mesmo poderia ser dito de um robô que replicasse totalmente minhas possibilidades funcionais, um robô humanóide como é o caso do COG.⁵ Assim sendo, nada indica que estados conscientes sejam logicamente supervenientes em relação a estados físicos e nem mesmo a determinadas arquiteturas funcionais. Estados conscientes são, no máximo, natural ou empiricamente supervenientes em relação a estados físicos, ou seja, não há conexão lógica entre base física ou arquitetura funcional e consciência. A consciência é contingente em relação a sua base física; ela é um *fator suplementar*.⁶

A crítica às possibilidades das explicações funcionais é seguida, no texto de Chalmers, por um ataque às explicações reducionistas, conservando a mesma linha de raciocínio. As explicações redutivistas, quase sempre no âmbito da ciência cognitiva ou da neurociência, escamoteiam a verdadeira natureza do problema da consciência e o identificam com os “easy problems”. Dentre os vários modelos de explicação reducionista analisados por Chalmers chamam a atenção os de Crick e Koch (1990), de Baars (1988) e de Dennett (1991).

⁵ Cog é o nome de um robot cujo projeto está atualmente sendo desenvolvido no laboratório de inteligência artificial do MIT. A idéia é construir um robot humanóide, uma máquina geral que possa fazer tudo o que um ser humano faz.

⁶ Veja-se por exemplo, a passagem onde Chalmers afirma: “... consciousness is a surprising feature of the universe. Our grounds for belief in consciousness derive solely from our own experience of it. Even if we knew every last detail about the physics of the universe - the configuration, causation and evolution among all the fields and particles in the spatiotemporal manifold that information would not lead us to postulate the existence of conscious experience. My knowledge of consciousness, in the first instance, comes from my own case, not from any external observation. It is my first-person experience of consciousness that forces the problem on me.” (1996, p.101).

Crick e Koch desenvolveram a chamada “teoria neurobiológica da consciência” Esta teoria baseia-se na descoberta de uma constância em certas oscilações neuronais que se situam entre 35-75 hertz no córtex cerebral. Crick e Koch desenvolvem a hipótese de que estas oscilações são responsáveis pela produção da consciência, na medida em que elas estão relacionadas com o estado de vigília num número grande de modalidades - visual e olfatória - bem como com a integração de informação. Os autores sugerem que no processo de integração de diferentes segmentos de informação, grupos neuronais oscilam na mesma frequência e fase numa sincronização perfeita. A integração de informação (*binding*), por sua vez, possibilita a identificação perceptual de objetos fora de nós, o que seria um primeiro passo para a explicação da natureza da consciência.

A objeção de Chalmers consiste em sustentar que este tipo de teoria é muito sugestivo, mas ela não nos diz nada acerca de como e porque alguns conteúdos mentais tornam-se experiências conscientes. A descoberta das oscilações por Crick e Koch sugere que estas seriam os *correlatos neurais* da experiência. Mas o “hard problem” permanece intocado: por que as oscilações geram experiências conscientes? Qual é a conexão entre estes dois fenômenos?

O segundo modelo explicativo criticado por Chalmers é oriundo da psicologia cognitiva. É a teoria do espaço global da consciência (*global workspace*), desenvolvida por Baars (1988). De acordo com esta teoria, os conteúdos conscientes estão contidos num *espaço global*: uma espécie de processador central usado para mediar a comunicação com um conjunto de processadores especializados não-conscientes. Quando estes processadores especializados precisam transmitir informação para o resto do sistema, eles o fazem mandando informação para o espaço global que atua como uma espécie de quadro comunitário, acessível a todos os outros processadores.

Baars utiliza-se deste modelo para se referir a muitos aspectos da cognição humana e para explicar uma série de contrastes entre funcionamento cognitivo consciente e inconsciente. Em última análise, estamos diante de uma teoria da *acessibilidade cognitiva* que explica como certos conteúdos informacionais tornam-se acessíveis dentro de um sistema. É

também uma teoria da integração informacional da mente e da possibilidade de auto-relatar conteúdos mentais. Contudo, ela não oferece uma teoria da experiência.

Poder-se-ia supor que, de acordo com esta teoria, os conteúdos da experiência são os conteúdos do espaço global. Mas nada explica porque a informação no interior do espaço global é experienciada. Esta teoria pode, no máximo, asseverar que a informação é experienciada porque ela é *globalmente acessível*. Mas por que a acessibilidade global teria de dar, necessariamente, origem à experiência consciente? Não seria possível ocorrer a acessibilidade global através do “workspace” sem ocorrer experiência consciente?

O modelo das “múltiplas camadas” (multiple drafts) desenvolvido por Dennett (1991) também é criticado por Chalmers. A idéia de Dennett baseia-se num modelo chamado “pandemonium”, uma série de pequenos agentes que disputam a primazia pelo foco da atenção. Tudo se passa como se o agente que “gritar mais alto” no meio desta disputa possa então “subir ao palco” e orientar o processamento subsequente - este agente corresponde, metaforicamente, a um estado mental que se torna (momentaneamente) consciente.⁷ Não existe “supervisor” neste modelo, sua dinâmica é ditada por princípios de auto-organização que Dennett toma emprestados de teorias conexionistas. Este modelo, no entender de Chalmers, pode no máximo explicar a possibilidade de se relatar o conteúdo de certos estados mentais. Ele escorrega para o lado dos “easy problems” na medida em que estipula *como* um estado mental torna-se consciente mas não o que seja a própria experiência consciente.

A revisão e a crítica de teorias contemporâneas da consciência prossegue na segunda parte do livro de Chalmers concentrando-se nos vários tipos de estratégias teóricas utilizadas pelos pesquisadores. A primeira

⁷ Dennett afirmou, certa vez, que a melhor maneira de entender seu modelo é pensar que a consciência é como a fama. Todos querem ser famosos e disputam um lugar no palco, mas o são apenas por alguns minutos e logo em seguida são substituídos por outros. O mesmo ocorre com estados mentais: quando se tornam “famosos” são conscientes por alguns segundos.

delas consiste em dizer que a experiência é um *fator complementar* na explicação dos mecanismos da consciência. Este tipo de abordagem deixa de lado as tentativas de explicar a natureza da experiência e concentra-se na explicação dos mecanismos cognitivos subjacentes à consciência, ou seja concentra-se nos “aspectos simples” (easy problem) do problema.

O segundo tipo de estratégia consiste em *negar a especificidade do fenômeno*. Esta linha é desenvolvida por pesquisadores como Allport (1988), e Wilkes (1988). De acordo com esta estratégia, se funções como acessibilidade, capacidade de relatar estados internos e outras são explicadas, não há necessidade de explicar o que chamamos de “experiência”. Alguns partidários desta estratégia procuram negar o fenômeno experiência dizendo que ele não é externamente verificável e portanto não é algo real. Esta estratégia tem como resultado a formulação de teorias bastante simples, mas insatisfatórias. Na realidade, eles escamoteiam o problema.

Na terceira estratégia, alguns pesquisadores *afirmam ter explicado a experiência*. Eles abordam este aspecto do problema seriamente, e dizem que sua teoria funcional explica as qualidades subjetivas da experiência (Flohr, 1992; Humphrey, 1992). Eles explicam como o processamento de informação ocorre e, subitamente a idéia de experiência é introduzida. Contudo, não explicam como a consciência emerge desses processos.

Uma quarta estratégia apela para a idéia de *explicar a estrutura da experiência*. Argumenta-se por exemplo, que uma explicação de como o sistema visual opera discriminações pode explicar as relações entre diferentes experiências de cor (ver Clark, 1992 e Hardin, 1992). Fatos acerca dessas estruturas no processamento corresponderiam a fatos na estrutura da experiência. O problema desta estratégia é que ela toma a própria existência da experiência como ponto de partida - e isto significa, de certa maneira, escamotear uma explicação de como e porquê a experiência se forma nestes fenômenos.

Uma quinta estratégia consiste em *isolar o substrato da experiência*. Toma-se como ponto de partida o fato de que a experiência emerge de processos cerebrais. É preciso então identificar os processos que le-

vam ao aparecimento deste tipo de fenômeno específico. Esta é a linha adotada por Crick e Koch, ao tentar isolar o correlato neuronal da consciência. O mesmo tipo de linha é adotada por Edelman (1989) e Jackendoff (1987). Contudo, esta estratégia é ainda insatisfatória. Uma teoria satisfatória tem de fornecer mais do que simplesmente isolar os processos que dão lugar ao aparecimento da experiência.

Todas estas estratégias falham na medida em que não fornecem um bom método para explicar o *ingrediente suplementar* (extraness) necessário para se obter uma explicação da natureza da consciência. Mas o que poderia ser este ingrediente suplementar e como ele poderia explicar a natureza da *experiência consciente*?

A análise de Chalmers recobre as tentativas de alguns teóricos que propuseram que este ingrediente suplementar deve ser procurado na teoria do caos ou na dinâmica não-linear. Outros sugerem que a chave para isto está no processamento não-algorítmico. Outros apelam para futuras descobertas da neurofisiologia e outros ainda, para a mecânica quântica.

O processamento não-algorítmico é sugerido por Penrose (1989, 1994) por causa do papel da consciência na intuição matemática. Mas este tipo de explicação - na concepção de Chalmers - ainda seria apenas uma explicação de *funções* envolvidas no raciocínio matemático. Pois mesmo que falemos de processamento não-algorítmico podemos ainda questionar porque este último daria origem à experiência. Assim sendo, a teoria de processamento não-algorítmico não teria nenhuma vantagem aparente.

O mesmo é afirmado por Chalmers acerca de processamento não-linear e da dinâmica do caos. Uma aplicação destas teorias pode fornecer uma explicação da dinâmica de funcionamento cognitivo, mas a questão da experiência ainda permanece inexplicada. Podemos sustentar a mesma afirmação acerca de possíveis descobertas neurofisiológicas.

Uma concepção de ingrediente suplementar que tem ganhado terreno ultimamente origina-se da mecânica quântica (Hameroff, 1994). A inspiração desta proposta baseia-se na idéia de que fenômenos quânticos têm características funcionais extremamente interessantes, como, por exemplo, o indeterminismo e a não-localidade. Poder-se-ia então especu-

lar que estas propriedades seriam responsáveis por certos processos cognitivos como, por exemplo, escolha randômica ou integração de informação. Mas, novamente, a crítica de Chalmers recai no fato de que estas teorias nada nos dizem acerca da natureza da experiência consciente.

A mesma crítica é por ele estendida a qualquer tentativa de explicar a consciência em termos puramente físicos. Pois qualquer teoria que siga esta linha, enfrentará no final o mesmo tipo de questão: por que tal e tal processo dá origem à experiência? Qualquer processo funcional pode ser instanciado sem a participação da experiência o que mostra que a experiência ultrapassa o que pode ser derivado de qualquer teoria física.

Explicações físicas são boas enquanto explicação do desempenho de *funções*, explicando estas últimas em termos de mecanismos físicos que as desempenham. Mas fatos acerca da experiência não podem ser consequência automática de nenhuma explicação física - eles podem existir sem experiências. A experiência *pode* emergir de uma estrutura física, mas não é consequência desta.

Chegamos assim à proposta de uma teoria não-reducionista da experiência consciente. O esboço desta teoria ocupa a terceira parte do livro de Chalmers, a parte que ele chama de “construtiva” na medida em que oferece uma alternativa a todas as teorias anteriormente criticadas. Esta teoria deve ser compatível com a proposta não-reducionista e com o dualismo naturalista, ou seja, ela não deve conflitar com os resultados da ciência. Em outras palavras, este dualismo brando deve especificar um conjunto de princípios básicos que nos mostrem como a experiência consciente *supervem* às características físicas do mundo. Estes *princípios psicofísicos* não interferem com as leis físicas na medida em que estas últimas formam um sistema fechado. Na realidade, elas suplementam a teoria física.

Chalmers identifica três princípios psicofísicos na sua teoria: o princípio de coerência estrutural, o princípio de invariância organizacional e o princípio do duplo aspecto da teoria da informação. O primeiro princípio estabelece uma relação coerente entre a “*structure of consciousness*” e a “*structure of awareness*” ou seja, toda experiência consciente é cognitivamente representada, ou seja, assume a forma de um pro-

cesso cognitivo, embora nem tudo o que seja cognitivamente representável seja necessariamente consciente. Existe uma relação íntima entre cognição e consciência que torna os estados conscientes passíveis de relato verbal, acessíveis aos sistemas centrais que controlam o comportamento e tudo o mais que compõe a “*structure of awareness*” Este quase-isomorfismo entre *structure of consciousness* e *structure of awareness* permite que teorias cognitivas e neurofisiológicas sirvam de ponto de partida para uma teoria da experiência consciente: estas teorias devem explicar a base física ou os correlatos neurofisiológicos sobre os quais a experiência consciente supervém.

O princípio da invariância organizacional estipula que dois sistemas com a mesma organização funcional terão experiências qualitativamente idênticas. Isto significa dizer que se construirmos uma réplica do cérebro humano em silicone preservando os mesmos padrões causais de organização neuronal, este cérebro replicado poderá ter as mesmas experiências que o cérebro humano. O que conta na emergência de experiências não é o tipo de substrato físico de um sistema mas seu princípio arquetônico ou a organização de seus componentes.

O terceiro princípio, do duplo aspecto da informação é o princípio básico e fundamental da teoria da consciência de Chalmers. Ele toma como ponto de partida a noção de informação tal como é definida por Shannon (1948) e sustenta que esta tem um duplo aspecto: um físico e outro fenomênico. É o aspecto fenomênico que dá origem à experiência consciente e este princípio é, sem dúvida, o mais controverso na teoria de Chalmers: afinal, quais são as peculiaridades da informação que podem dar origem a estados conscientes? Será a consciência privilégio apenas de cérebros humanos ou poderá ela ser estendida a outros processadores de informação como cérebros de animais ou até mesmo máquinas?

É notável o quanto este aspecto permanece obscuro na teoria de Chalmers e o situa ao lado do grupo de filósofos contemporâneos como McGinn que foram chamados de “New Mysterians” por suporem que há

algo de misterioso na explicação da consciência.⁸ Em várias passagens de seu livro nota-se um constante flerte com posições dualistas que são, em seguida, abrandadas pela idéia de um “dualismo naturalista”⁹ Afinal, ao reconhecer que a “experiência consciente” é uma dimensão qualitativa do universo ou um “primitivo” da mobília do mundo estaremos tão distantes assim da idéia cartesiana da pluralidade das substâncias? Pouco podemos dizer do “aspecto dual da informação” da mesma maneira que pouco se pode dizer das características da “substância pensante” cartesiana. A irredutibilidade da dimensão subjetiva da experiência consciente parece originar-se do fato desta apresentar-se como um dado imediato - mas será este o único ponto de partida plausível para iniciarmos uma teoria da consciência? Por que teríamos de necessariamente iniciar nossa reflexão assumindo uma posição solitária? Quando olhamos para uma lagosta sendo jogada na água quente, contorcendo-se com a dor, não estamos intuitivamente atribuindo algum tipo de experiência consciente a esse organismo?

O flerte de Chalmers com o cartesianismo torna-se igualmente evidente na sua teoria da superveniência dos estados conscientes. A crítica a explicações reducionistas e puramente funcionais da natureza da consciência encontra-se, de maneira embrionária, nos escritos de Descartes sobre os autômatas. Descartes sustentava que a duplicação de característi-

⁸ Numa entrevista concedida a Robert Wright, da revista Time de abril de 1996, McGinn afirma “For human beings to try to grasp how subjective experience arises from matter is like slugs trying to do Freudian psychoanalysis. They just don’t have the conceptual equipment.” (p.45).

⁹ Veja-se por exemplo uma das passagens finais do seu livro onde ele diz “I have advocated some counterintuitive views in this work. I resisted mind-body dualism for a long time, but I have now come to the point where I accept it, not just as the only tenable view but as a satisfying view in its own right. It is always possible that I am confused, or that there is a new and radical possibility that I have overlooked, but I can comfortably say that I think dualism is very likely true. I have also raised the possibility of a kind of panpsychism. Like mind-body dualism, this is initially counterintuitive, but the counterintuitiveness disappears with time. I am unsure whether the view is true or false, but it is at least intellectually appealing, and on reflection it is not too crazy to be acceptable.” (p.357).

cas materiais e funcionais de um ser humano poderia ser condição necessária mas não suficiente para se replicar a vida mental humana.¹⁰ Um autômato bem construído pode vir a fazer tudo o que um ser humano faz, mas nunca se igualaria a este: seria, no máximo, uma proeza de engenharia, algo que, contudo, não teria *alma* (e não poderíamos substituir esta palavra por “experiência consciente”?) Neste sentido, o autômato de Descartes não é muito diferente do zumbi de Chalmers.

A diferença entre a posição de Chalmers e a posição cartesiana consiste no fato de Descartes ter afirmado, categoricamente, que a vida mental *não pode supervir* no autômata. Chalmers deixa aberta esta possibilidade, ao defender a Inteligência Artificial no sentido forte, nos últimos capítulos de seu livro. Mas a pressuposição de Chalmers de que a similaridade funcional não é suficiente e não *implica* na produção de estados conscientes é inteiramente metafísica. Afinal, se mantivermos o primado da primeira pessoa para fundar nossa teoria da consciência, o que pode nos garantir que um robô que faça tudo o que um ser humano pode fazer não tem experiências conscientes?

Esta última questão faz-nos refletir sobre outros problemas que surgem a partir da teoria de Chalmers - problemas tão interessantes quanto complexos. Em primeiro lugar, destaca-se o chamado *problema da repredicação*. Suponhamos que por um certo período de tempo tenhamos convivido com um robô de forma humanóide, uma réplica cuja aparência externa fosse exatamente igual à de um ser humano. Este robô poderia ser, por exemplo, o COG, o robô humanóide que no momento está sendo desenvolvido no MIT. O COG estaria convivendo conosco e seu comportamento seria indistinguível daquele exibido por um ser humano qualquer. Ocorre que *não sabemos* que estávamos lidando com um robô e não um ser humano. Isto significa que por muito tempo estaríamos atribuindo ao COG os mesmos predicados mentais que normalmente atribuímos a um ser humano, incluindo a capacidade de desenvol-

¹⁰ A este respeito poderíamos citar várias passagens do *Discurso do Método*. Mais ilustrativa, contudo, é a carta de Descartes ao Marquês de Newcastle, de 23 de novembro de 1646, onde estas posições são sustentadas de maneira mais explícita.

ver comportamentos e experiências conscientes. Um dia, o COG (que não sabíamos ser um robô) escorrega, cai e bate a cabeça na banheira. Seu crânio se rompe e, em vez de encontrarmos dentro dele a massa encefálica de um ser humano, encontramos fios e *chips* de computador. Teria cabimento *retirar* todos os predicados mentais que vínhamos atribuindo a ele até então - predicados mentais que o equiparavam a um ser humano normal? Teria cabimento afirmar: “bem, agora que eu descobri que você é na verdade um robô, então você não tinha estados mentais nem tampouco experiências conscientes?”

A segunda questão surge no mesmo esteio da primeira: COG seria, no máximo, um zumbi. Mas será possível supor a existência de zumbis, mesmo enquanto possibilidade metafísica? A suposição fundamental subjacente à concepção de zumbi defendida por Chalmers é que estas seriam criaturas que agem, conversam, sentem dores etc, ou seja, poderiam passar no Teste de Turing¹¹ de maneira eficiente. A única - e grande diferença - estaria no fato de que eles não poderiam ter experiências conscientes. Mas, se um zumbi é, do ponto de vista comportamental, indistinguível de um ser humano, o que poderia nos impedir de atribuir a ele a propriedade de ter consciência? O que ocorreria se, durante o teste de Turing o interrogador formulasse a questão: “Você tem experiências conscientes?” ou “Você tem consciência daquilo que acabaram de perguntar a você?” Haveria duas possibilidades de resposta, uma afirmativa outra negativa. Mas, em ambos os casos, a noção de experiência consciente já se encontra pressuposta na resposta que o zumbi pode dar, seja ela afirmativa ou negativa, esteja ele mentindo ou não. Alternativamente, ele poderia ser incapaz de fornecer qualquer tipo de resposta, mas, neste caso, ele não passaria no Teste de Turing e sua suposta existência como ser que faz

11 O Teste de Turing, criado pelo matemático inglês homônimo, consiste em comparar os comportamentos manifestos de um organismo humano com aqueles produzidos por um robô ou computador criado para desenvolver tarefas humanas. Se da comparação resultar que as características dos comportamentos do organismo são indistinguíveis daquelas dos *outputs* produzidos pela máquina, podemos, de acordo com Turing, atribuir a esta estados mentais.

tudo que um ser humano pode fazer - exceto ter estados conscientes - tornar-se-ia uma impossibilidade, ou melhor, uma contradição em termos...

Uma terceira série de questões surge ao refletirmos sobre a noção de superveniência introduzida por Chalmers. Terá sentido, afinal de contas, afirmar que a consciência constitui um ingrediente suplementar que supervém à organização mental e funcional de um organismo ou sistema? Não estaríamos aqui diante de uma confusão conceitual? Até que ponto é sustentável a independência da experiência consciente em relação à organização funcional ou à estrutura física de um organismo? Tomemos os predicados *ser consciente* e *ter saúde*. Em ambos os casos, a atribuição destes predicados não dependeria da possibilidade de explicar o funcionamento de uma estrutura física específica de um organismo, isto é, em ambos os casos, a atribuição destes predicados fundamenta-se na observação de uma característica global do organismo. Contudo, aqui corremos o risco de deslizar da idéia de *característica global* para a idéia de *característica adicional*. Não teria cabimento supor que - mesmo por um ato de imaginação filosófica - poderíamos remover a saúde de um organismo ao mesmo tempo que mantemos a totalidade de seus órgãos e suas interações em perfeito estado, ou, inversamente, que poderíamos remover alguns desses órgãos e, mesmo assim, achar que preservamos a saúde do organismo, isto é, que ela poderia permanecer intacta. Ora, por que não poderíamos afirmar o mesmo em relação à consciência?

TEIXEIRA, J.F. On Chalmer's Theory of Consciousness. *Psicologia USP, São Paulo*, v.8, n.2, p.109-128, 1997.

Abstract: The paper focuses on Chalmer's theory of consciousness as it is presented in his most recent book, *The Conscious Mind*, published in 1996. The first part is devoted to a presentation of the main outlines of Chalmer's theory. The second part discusses such a theory by focusing on the metaphysical plausibility of the existence of zombies as well as on the notion of supervenience.

Index terms: Consciousness. Artificial intelligence. Cartesianism. Cognition.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALLPORT, A. What concept of consciousness? In: MARCEL, A.J.; BISIACH, E., eds. *Consciousness in contemporary science*. Oxford, Oxford University Press / Clarendon Press, 1988. p.159-82.
- BAARS, B.J. *A cognitive theory of consciousness*. Cambridge, MA, Cambridge University Press, 1988.
- CALVIN, W.H. *The cerebral symphony*. New York, Bantam Books, 1990.
- CHALMERS, D.J. *The conscious mind*. New York, Oxford University Press, 1996.
- CLARK, A. *Sensory qualities*. Oxford, Clarendon Press, 1992.
- CRICK, F.; KOCH, C. Toward a neurobiological theory of consciousness. *Seminars in the Neurosciences*, v.2, p.263-75, 1990.
- DENNETT, D. *Consciousness explained*. Boston, Little, Brown, 1991.
- DENNETT, D. *Toward a cognitive theory of consciousness*. In: *Brainstorms*. Cambridge, MA, MIT Press, 1978. p.149-73.
- DESCARTES, R. *Oeuvres et lettres*. Paris, Gallimard, 1953. (Bibliothèque de la Pléiade)
- EDELMAN, G. *The remembered present: a biological theory of consciousness*. New York, Basic Books, 1989.
- FLANAGAN, O. *Consciousness reconsidered*. Cambridge, MA, MIT Press, 1992.
- FLOHR, H. Qualia and brain processes. In: BECKERMANN, A.; FLOHR, H.; KIM, J., eds. *Emergence or reduction? Prospects for nonreductive physicalism*. Berlin, DeGruyter, 1992.
- HAMEROFF, S.R. Quantum coherence in microtubules: a neural basis for emergent consciousness? *Journal of Consciousness Studies*, v.1, p.91-118, 1994.
- HARDIN, C.L. Physiology, phenomenology and Spinoza's true colors. In: BECKERMANN, A.; FLOHR, H.; KIM, J., eds. *Emergence or reduction? Prospects for nonreductive physicalism*. Berlin, DeGruyter, 1992.
- HUMPHREY, N. *A history of the mind*. New York, Simon and Schuster, 1992.
- JACKENDOFF, R. *Consciousness and the computational mind*. Cambridge, MA, MIT Press, 1987.
- JAMES, W. *The principles of psychology*. New York, Henry Holt, 1890. v.1. Reprinted by Dover Books, 1950.
- McGINN, C. Can we solve the mind-body problem? *Mind*, v.98, p.349-66, 1989.

- MOROWITZ, H.J. Rediscovering the mind. *Psychology Today*, v.14, n.3, p.12-8, 1980. Reimpresso em: DENNETT, D.; HOFSTADTER, D., eds. *The Mind's I*. Cambridge, MA, MIT Press, 1981.
- NAGEL, T. What is it to be like a bat? *Philosophical Review*, v.4, p.435-50, 1974.
- PENROSE, R. *The emperor's new mind*. Oxford, Oxford University Press, 1989.
- PENROSE, R. *Shadows of the mind*. Oxford, Oxford University Press, 1994.
- SHANNON, C.E. A mathematical theory of communication. *Bell Systems Technical Journal*, v.27, p.379-423, 1948.
- TYNDALL, J. *Fragments of science: a series of detached essays, addresses and reviews*. London, Longman, 1879.
- WILKES, K.V. Yishi, duh, um and consciousness. In: MARCEL, A.J.; BISIACH, E., eds. *Consciousness in contemporary science*. Oxford, Oxford University Press / Clarendon Press, 1988. p.16-41.