

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
PROGRAMA DE MESTRADO E DOUTORADO EM HISTÓRIA DAS  
CIÊNCIAS, DAS TÉCNICAS E EPISTEMOLOGIA  
ÁREA DE CONCENTRAÇÃO: EPISTEMOLOGIA, LÓGICAS E  
TEORIAS DA MENTE

DISSERTAÇÃO DE MESTRADO

CONSCIÊNCIA: RELAÇÕES ENTRE IRREDUTIBILIDADE E  
COMPUTABILIDADE

ROGERIO MANDELLI

RIO DE JANEIRO  
2012

*TÍTULO: “CONSCIÊNCIA: RELAÇÕES ENTRE IRREDUTIBILIDADE E COMPUTABILIDADE”*

Esta dissertação de mestrado foi julgada adequada como requisito para obtenção do título de mestre em História das Ciências das Técnicas e Epistemologia (Áreas de concentração: *Epistemologia, Lógicas e Teorias da Mente*) pelo programa de Mestrado e Doutorado em História das Ciências e das Técnicas e Epistemologia (HCTE) da Universidade Federal do Rio de Janeiro. A dissertação foi aprovada em sua forma final em sessão pública de defesa, no dia 1º de março de 2012, pela banca examinadora composta pelos seguintes professores:

---

*Prof. Dr. Luiz Pinguelli Rosa - UFRJ (Presidente);*

---

*Prof. Dr. Ricardo Silva Kubrusly – UFRJ (Examinador);*

---

*Prof. Dr. Luiz Alberto Oliveira – CPBF (Examinador);*

---

*Prof. Dr. Sergio Exel – Escola Politécnica UFRJ (Examinador).*

*RIO DE JANEIRO – RJ, BRASIL*

*PROF. DR. RICARDO SILVA KUBRUSLY  
COORDENADOR DO PROGRAMA DE MESTRADO E DOUTORADO EM HCTE  
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO*

ROGERIO MANDELLI

CONSCIÊNCIA: RELAÇÕES ENTRE IRREDUTIBILIDADE E  
COMPUTABILIDADE

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em História das Ciências e das Técnicas e Epistemologia (HCTE), Universidade Federal do Rio de Janeiro.

Orientador: Prof. Luiz Pinguelli Rosa

RIO DE JANEIRO

2012

Dados Internacionais de Catalogação na Publicação (CIP)  
Biblioteca da Universidade Federal do Rio de Janeiro – Rio de Janeiro – RJ

M271 Mandelli, Rogerio.

Consciência: relações entre a irreducibilidade e computabilidade –  
Rio de Janeiro: UFRJ, 2012.

123 f: il.

Dissertação (Mestrado em História das Ciências e das Técnicas e Epistemologia) – Universidade Federal do Rio de Janeiro, Centro de Ciências Matemáticas e da Natureza, Programa de Pós Graduação em História das Ciências e das Técnicas e Epistemologia, 2012.

Orientador: Prof. Dr. Luiz Pinguelli Rosa.

1. Consciência – Teses. 2. Filosofia da mente – Teses. 3. Inteligência artificial – Teses. 4. Funções computáveis – Teses. I. Rosa, Luiz, Luiz Pinguelli (Orient.). II Universidade Federal do Rio de Janeiro, Centro de Ciências Matemáticas e da Natureza, Programa de Pós Graduação em História das Ciências e das Técnicas e Epistemologia. III. Título.

CDD 006.3

## Dedicatória

*Aos Meus Pais e à Natureza, responsáveis por eu aqui estar.*

## Agradecimentos

Ao admirável e incansável professor Luiz Pinguelli Rosa, pela generosidade e por demonstrar sempre o profundo amor pelo que faz: ensinar e orientar.

Ao meu querido professor Ricardo Kubrusly pela confiança, pela capacidade de surpreender-nos a cada aula e por demonstrar, a cada dia, as infinitas possibilidades que poesia e ciência, juntas, podem gerar.

A minha família. Meus pais, irmãos, sobrinhos e cunhadas, por apostar e acreditar que sempre há recomeços e que, principalmente, sempre valem a pena.

Ao meu amigo e irmão de jornada, Alexandre Valença, pela sua amizade e lealdade incondicionais, nos momentos alegres e, principalmente, nas etapas difíceis da caminhada.

Aos meus amigos, os presentes e os que, de alguma forma, já não compartilham os mesmos caminhos.

Ao grande professor Luiz Alberto Oliveira, que, do CBPF, pôde contribuir com seu vasto e luminoso conhecimento de ciência e filosofia.

A todos os professores do HCTE, que, de forma direta ou indireta, contribuíram para a realização desta jornada.

Ao amigo Nelson Job, por ter-me apresentado ao HCTE. Agradeço também à querida Elaine de Paiva Andrade pelas aulas, pelos diálogos e pelo apoio constante.

Às queridíssimas Fátima Alexandre e Daniela Carnevali, secretárias do Professor Pinguelli, pela amizade e carinho, e por diversas vezes terem conseguido um jeito de estarmos reunidos com o professor, sempre acompanhados dos intermináveis 'papo-cabeça'.

Às competentes Mariah Martins e Gabriela Evangelista e a todos os funcionários do HCTE e da UFRJ com os quais tive o privilégio de conviver.

A todos os professores que passaram pela minha vida e também a todas as pessoas que, de algum modo, também desempenharam esse papel.

À Natureza Una e seus insondáveis atributos do Tempo, do Espaço e da Causalidade... por proporcionar essa sensacional experiência que é Viver!

*Para chegares ao que não sabes  
Deves seguir por um caminho que é o caminho da ignorância.*

*Para possuíres o que não possuis  
Deves seguir pelo caminho do despojamento.*

*Para chegares ao que não és  
Deves cruzar pelo caminho em que não és.*

*E o que não sabes é apenas o que sabes*

*E o que possuis é o que não possuis*

*E onde estás é onde não estás.*

*In order to arrive at what you do not know  
You must go by a way which is the way of ignorance.*

*In order to possess what you do not possess*

*You must go by the way of dispossession.*

*In order to arrive at what you are not*

*You must go through the way in which you are not.*

*And what you do not know is the only thing you know*

*And what you own is what you do not own*

*And where you are is where you are not.*

QUATRO QUARTETOS (1943)

T.S. ELIOT



## Resumo

O objetivo desta dissertação é a discussão das ideias contemporâneas sobre a consciência, a partir da filosofia da mente e suas repercussões no cenário da inteligência artificial. São abordadas algumas questões relacionadas à problemática mente-corpo, presentes na obra do filósofo David Chalmers e as consequências e dificuldades na construção de uma teoria consistente da consciência. A dissertação busca mostrar a possibilidade da redutibilidade da consciência a fenômenos físicos, baseada na superveniência lógica dos estados qualitativos das experiências da realidade objetiva. Assim pode-se conjecturar se os fatos físicos implicam os fatos mentais e estes, então, possam ser considerados como apenas mais uma parcela do mundo concreto. Como contraponto ao antifisicalismo, será descrito o modelo quântico da mente de Roger Penrose e Stuart Hameroff, que, apesar de carente ainda de provas definitivas no campo experimental, ecoa como uma coerente possibilidade fisicalista. Tal modelo propõe que o fluxo de consciência seja um fenômeno discreto e não contínuo e que é gerado a partir de microestruturas presentes nos neurônios cerebrais. Sob a ótica dos problemas e modelos apresentados, o ponto conclusivo deste trabalho procura demonstrar e propor discussões sobre a computabilidade da mente humana e seus limites e, em lados opostos: a inteligência artificial forte e a inteligência artificial fraca e suas nuances. Busca-se, ao final, lançar um pouco mais de luz sobre as antigas questões que permeiam toda a dissertação e que fazem parte da curiosidade humana: como nos identificamos e nos relacionamos com o mundo a nossa volta.

**Palavras-chave:** Consciência, mente, irredutibilidade, computabilidade.

### **Abstract**

The objective of this thesis is discussion of contemporary ideas about consciousness, from the philosophy of mind and its repercussions in the artificial intelligence field. It covers some issues related to the mind-body problem, present in the work of the philosopher David Chalmers, and the consequences and difficulties in building of a consistent theory of consciousness. The dissertation seeks to show the possibility of reducibility of consciousness to physical phenomena, based on logic supervenience of qualitative experience states from objective reality. Thus, we may conjecture if the mental facts arise from physical facts, then, they can be regarded as just a portion of the concrete world. As a counterpoint to antiphysicalism, the quantum mind model of Roger Penrose and Stuart Hameroff will be described, which, although still lacking definite evidence in the experimental field, resonates as a coherent possibility in physicalist approach. This model proposes that the stream of consciousness is a discrete phenomenon and not continuous and it is generated from microstructures present in brain neurons. From the perspective of the problems and models presented, the conclusive point of this paper seeks to demonstrate and propose discussions on the computability of the human mind and its limits and, on opposite sides: the strong artificial intelligence and weak artificial intelligence and its nuances. At the end looking up throw a little more light on old issues that permeate the entire dissertation and that are part of human curiosity: how we identify and relate to the world around us.

**Keywords:** Consciousness, mind, irreducibility, computability.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>13</b>
<b>2</b>	<b>QUESTÕES SOBRE A IRREDUTIBILIDADE DA CONSCIÊNCIA</b> .....	<b>21</b>
<b>2.1.</b>	<b>Que é consciência?</b> .....	<b>21</b>
<b>2.2.</b>	<b>Os Problemas da Consciência</b> .....	<b>28</b>
<b>2.3.</b>	<b>Resistências ao Reduccionismo</b> .....	<b>32</b>
<b>2.4.</b>	<b>A Questão da Superveniência da Consciência e o Mundo Físico</b> .....	<b>35</b>
<b>2.5.</b>	<b>A Visão Materialista</b> .....	<b>44</b>
<b>2.6.</b>	<b>Direções para uma Teoria Não Reduccionista</b> .....	<b>48</b>
<b>3</b>	<b>A FÍSICA DA MENTE?</b> .....	<b>53</b>
<b>3.1.</b>	<b>Emergência</b> .....	<b>54</b>
<b>3.2.</b>	<b>A Teoria Quântica e o Cérebro</b> .....	<b>57</b>
<b>3.3.</b>	<b>A Natureza do Mundo Quântico e o Problema Fundamental</b> .....	<b>58</b>
<b>3.4.</b>	<b>A Proposta de Penrose</b> .....	<b>62</b>
<b>3.5.</b>	<b>Críticas ao Modelo de Penrose-Hameroff</b> .....	<b>79</b>
<b>4</b>	<b>COMPUTABILIDADE E INTELIGÊNCIA ARTIFICIAL</b> .....	<b>82</b>
<b>4.1.</b>	<b>Gödel e a Máquina de Turing</b> .....	<b>82</b>
<b>4.2.</b>	<b>Consciência e Computação</b> .....	<b>89</b>
<b>4.3.</b>	<b>Computadores Conscientes</b> .....	<b>98</b>
<b>4.4.</b>	<b>Inteligência Artificial Forte?</b> .....	<b>105</b>
<b>4.5.</b>	<b>Inteligência Artificial e Platonismo</b> .....	<b>108</b>
<b>5</b>	<b>CONCLUSÃO</b> .....	<b>110</b>
<b>6</b>	<b>REFERÊNCIAS</b> .....	<b>117</b>

## Lista de ilustrações

<b>FIGURA 1.....</b>	<b>66</b>
<b>FIGURA 2.....</b>	<b>72</b>
<b>FIGURA 3.....</b>	<b>75</b>
<b>FIGURA 4.....</b>	<b>97</b>

# 1 INTRODUÇÃO

A consciência ocupa um lugar de destaque ao longo da história do pensamento ocidental, talvez por estar intimamente ligada na forma como vemos, ouvimos e interpretamos cada instante e como representamos e construímos a nossa realidade. Os domínios de tal problema são vastos. Começam nas nossas primeiras interações com o mundo, no modo como nos relacionamos com ele e estendem-se aos mais altos domínios do conhecimento humano, da ciência e aos ainda misteriosos campos da subjetividade.

Tal problema pode ser também nomeado como a questão mente-corpo, cujos estudos se debruçam sobre as relações entre a mente, também atualmente chamada de processos mentais e os estados físicos e seus respectivos processos. O trabalho dos cientistas e filósofos é buscar determinar a natureza do fenômeno conhecido como mente e seus processos correlatos e como ou se estes afetam ou são afetados pelos processos físicos e biológicos.

Mesmo com todo o poder das delimitações inerentes ao pensamento analítico, a questão da consciência ou problema mente-corpo ainda talvez esteja longe de ser solucionado, mas caminhos promissores para tal entendimento há muito estão sendo desenvolvidos. Esse cenário nebuloso já havia sido vislumbrado há milênios pelos pensadores gregos, conforme é demonstrado nas palavras de Heráclito, em seu fragmento número 45<sup>1</sup>: “Jamais seria possível descobrir os limites da alma, ainda que todos os caminhos fossem percorridos; tão profunda é a sua medida.”

Na parte introdutória do livro de Thomas Robinson, o professor Gabrielle Cornelli, da Universidade de Brasília, destaca que a complexidade do pensamento que os primeiros filósofos desenvolveram ecoa de forma inegável nos termos que definem, desde então, os âmbitos semânticos da filosofia com tal: *arché*<sup>2</sup>, *phýsis*<sup>3</sup>,

---

<sup>1</sup> Robinson (2010).

<sup>2</sup> Para os pré-socráticos, *arché* (ἀρχή) seria um princípio que deveria estar presente em todos os momentos da existência de todas as coisas; no início, no desenvolvimento e no fim de tudo. Princípio pelo qual tudo vem a ser.

<sup>3</sup> *Phýsis* pode ser traduzida por natureza, mas seu significado é mais amplo. Refere-se também à realidade, não aquela pronta e acabada, mas a que se encontra em movimento e

*logos*<sup>4</sup>, *psyché*<sup>5</sup>. Especificamente, a palavra *psyché* é bastante utilizada no contexto contemporâneo da psicologia e das ciências cognitivas. Muitos dos termos utilizados nessas áreas derivam diretamente da referida palavra, não sendo esta apenas uma palavra grega, pois é encontrada já nos primeiros registros da língua, ou seja, nos poemas de Homero. Sua tradução mais usual é ‘alma’, como na tradução do fragmento de Homero descrito anteriormente, mas muitos comentadores contemporâneos já preferem traduzir o termo por ‘mente’.

Não importando seu significado sintático, mas sim o semântico, autores posteriores a Homero e Heráclito, passando pelas tragédias gregas, aprofundaram consideravelmente o estudo da alma ou mente. São bem conhecidos os textos de Sócrates como *Apologia*, no *Fédon*, e na *República*, de Platão, ou ainda no *De Anima* aristotélico. Grande parte das questões suscitadas por esses expoentes do pensamento ocidental continuam como paradigmas vigentes.

Outra grande contribuição para tais questões que seguem no grande debate da contemporaneidade é a de René Descartes, que, no século XVII, proclamou o dualismo em sua obra *Meditações Metafísicas*, em que são demonstradas a existência de Deus e a distinção real entre a mente e o corpo. A ideia aparece na filosofia ocidental, já nos escritos de Platão e Aristóteles, que afirmam, por diferentes razões, que a inteligência do homem (uma faculdade do espírito ou da alma) não pode ser assimilada ao seu corpo, nem entendida como uma realidade física.

O que se busca desenvolver neste trabalho encontra grande ressonância na célebre conferência inaugural de Popper, proferida na reunião da *Aristotelian Society*, realizada em Londres, nos idos de 1958, intitulada de forma significativa *Back to Presocratics*.

---

transformação, a que nasce e se desenvolve, o fundo eterno, perene, imortal e imperecível de onde tudo brota e para onde tudo retorna.

<sup>4</sup> O *Logos* (λόγος), no grego, significava inicialmente a palavra escrita ou falada - o Verbo. Mas a partir de filósofos gregos como Heráclito passou a ter um significado mais amplo. *Logos* passa a ser um conceito filosófico traduzido como razão, tanto como a capacidade de racionalização individual ou como um princípio cósmico da Ordem e da Beleza.

<sup>5</sup> *Psyché* ou psique era o conceito grego para o *self* ('si-mesmo'), abrangendo as ideias modernas de alma, ego e mente.

Popper propunha em sua fala a volta para a simplicidade da teoria do conhecimento que os pré-socráticos haviam elaborado. Segundo ele:

Tal teoria não poderia iniciar pela pergunta: 'como sei que isto é uma laranja?', ou 'como sei que o objeto por mim percebido é uma laranja?', mas – ao contrário – de problemas do tipo 'como sabemos que o mundo é feito de água?', ou 'como sabemos que o mundo é cheio de deuses?', ou ainda 'como podemos saber qualquer coisas acerca dos deuses?' (ROBINSON, 2010).

Assim, Popper encontra nos filósofos gregos a ousadia necessária de formular as questões mais importantes, e talvez as únicas pelas quais valha a pena filosofar: como moldamos e compreendemos o mundo e a nós mesmos?

O advento e o estabelecimento de uma ciência da consciência que possa ser consistente não devem e não podem ser baseados em termos do modelo atual vigente no mundo acadêmico. Faz-se necessária a busca de diálogos constantes com os outros ramos do conhecimento. A Neurociência atual parece apontar para caminhos semelhantes, mas ainda incipientes. Os paradigmas hodiernos, segundo criterioso estudo e análise, talvez não comportem mais o modelo ainda estabelecido no mundo acadêmico, ou seja, a 'compartimentalização' do conhecimento. Em vez dessa concepção, é necessária a adoção de um modelo interdisciplinar como ponto crucial em seu estatuto e, ainda mais, um modelo transdisciplinar, em que o fluxo de conhecimento, bem como sua síntese, que possa ser proposto de modo que não só una as diversas áreas do saber, mas que vá além delas e possa transitar harmoniosamente nesses diferentes 'mundos'. Muitos filósofos do século passado tomaram a ideia do século XIX de espaço e tempo como o único quadro em que suas descrições da experiência eram elaboradas. Alguns autores compreenderam as limitações desse quadro e apontaram que a nossa incapacidade de compreendê-las e superá-las está, provavelmente, na raiz da nossa incapacidade de compreender a consciência. Um desses autores é Alfred Whitehead (1920), que nos anos vinte traça o problema com a visão positivista do século XIX, quando a imaginação se subordina à observação e se busca apenas pelo observável e concreto através dos sentidos objetivos. Ele também acreditava que a mente e a natureza são partes do mesmo fenômeno:

O que pretendo essencialmente é protestar contra a bifurcação da natureza em dois sistemas de realidade, que, na medida em que são reais, são reais em diferentes sentidos. Uma realidade seria as entidades como os elétrons, que são objetos de estudo da física especulativa. Esta seria a realidade que está aí para conhecimento; embora essa teoria nunca conhecerá o outro tipo de realidade, que é a mímica da mente. Assim, haveria duas naturezas, uma é a conjectura e a outra é o sonho.

Outro modo de expressar esta teoria sobre a qual estou contra-argumentando é bifurcar a natureza em duas dimensões, nomeadamente sobre a natureza apreendida na consciência e da natureza que é a causa da consciência. A natureza, que é o fato apreendido na consciência, detém dentro dela o verde das árvores, o canto dos pássaros, o calor do sol, a dureza das cadeiras e a sensação do veludo. A natureza que é a causa da consciência, é o sistema conjecturado de moléculas e elétrons que, portanto, afeta a mente como que para produzir a consciência da natureza aparente. O ponto de encontro dessas duas naturezas é a mente, a natureza causal sendo influente e a natureza aparente efluente. (WHITHEHEAD, 1920, p. 16).

Pode-se perguntar qual o campo de investigação da Ciência? Quais seus limites? A Ciência trata apenas dos atributos materiais do universo que são passíveis de se enquadrarem dentro dos métodos científicos bem estabelecidos? Podemos conjecturar que os limites da Ciência são os limites da mente humana. Talvez a afirmação possa estar correta, pois a brilhante mente humana é capaz de construir os equipamentos mais surpreendentes e edificar os métodos investigativos mais rigorosos, mas ainda não é capaz de explicar como tal mente se relaciona com todos os fenômenos por ela investigados. Poderíamos um dia chegar a uma compreensão científica adequada desse mistério ainda sombrio? É o fenômeno da consciência humana algo que está além do escopo da investigação científica, ou o poder do método científico será capaz um dia de resolver o problema da existência de nosso eu consciente?

O presente trabalho tem como linha central a discussão das ideias contemporâneas no fértil terreno das teorias da mente e seus reflexos no profícuo cenário da inteligência artificial. O problema da consciência é abordado com enfoques completamente diferentes e, em muitas delas, diametralmente divergentes. Busca-se mostrar a possibilidade da redutibilidade da consciência a fenômenos físicos, baseada na superveniência lógica dos estados qualitativos das experiências da realidade objetiva. Presume-se que, só assim, é possível conjecturar se os fatos físicos implicam os fatos mentais e estes, então, possam ser considerados como apenas mais uma parcela do mundo concreto.



E como contraponto ao antifisicalismo proposto pelo filósofo australiano David Chalmers, será introduzido o modelo quântico de geração da consciência do físico-matemático inglês Roger Penrose e do médico anestesiológico americano Stuart Hameroff, que, apesar de carente ainda de provas definitivas no campo experimental, ecoa como uma coerente possibilidade fisicalista. Tal modelo propõe que o fluxo de consciência seja um fenômeno discreto e não contínuo e que é gerado a partir de processos quânticos no interior das microestruturas presentes nos neurônios cerebrais. Sob a ótica dos problemas e modelos apresentados, o ponto conclusivo deste trabalho procura demonstrar e propor discussões sobre a computabilidade da mente humana e seus limites e, em lados opostos, a inteligência artificial forte e do outro a inteligência artificial fraca e suas nuances. Busca-se, ao final, lançar um pouco mais de luz sobre as antigas questões que permeiam toda a dissertação e que fazem parte da curiosidade humana: como nos identificamos e nos relacionamos com o mundo a nossa volta.

Há aqueles que acreditam que podemos realmente estar perto de uma compreensão científica da consciência; que esse fenômeno é apenas complexo, mas não pode ser classificado como mistério.

Os cientistas argumentam que é meramente uma questão de complexidade extrema e sofisticação organizacional de nossos cérebros que, atualmente, limitam a nossa compreensão da mente humana. Complexidade e sofisticação não devem ser subestimadas; talvez não exista necessidade de criação de novos modelos, e, sim, o desenvolvimento dos modelos existentes, tornando-os capazes de tratar de forma mais adequada tais características.

No outro extremo, estão aqueles que sustentam que as questões da mente e do espírito humano, ou seja, a própria consciência humana, são abordagens sobre as quais não podemos ter esperança de tratar de forma adequada pelos procedimentos frios e exatos de uma ciência 'insensível'.

O objetivo deste trabalho é tentar enfocar a questão da consciência do ponto de vista científico, de modo que, ao longo da dissertação, discussões entre a irreduzibilidade da mente e suas estritas relações com o campo da inteligência

artificial são delineadas. Questões sobre a possibilidade de ‘redução’ da consciência humana a simples fenômenos físicos parece, até os dias atuais, sem solução. Alguns cientistas apontam que há fortes indícios de que um ingrediente essencial está faltando para elucidar certas lacunas importantes. Tais ingredientes seriam necessários para que as questões centrais da mente pudessem ser acomodadas dentro de uma coerente visão científica do mundo. Possivelmente, este ingrediente seja, em si, algo que não está além da ciência, mas exija que esta deva ser aperfeiçoada, resultando numa perspectiva mais ampla e consistente da nossa realidade, a fim de poder tratar fenômenos que ainda não se enquadram em seu atual escopo.

Mesmo com a nossa limitada compreensão atual da natureza desse ingrediente ‘ausente’, podemos começar a apontar caminhos onde ele possa ser elucidado e, em decorrência, passe a fornecer uma contribuição relevante para fundamentar o que são nossos sentimentos e ações conscientes.

A scientific world-view which does not profoundly come to terms with the problem of conscious minds can have no serious pretensions of completeness.

Uma visão científica do mundo que não chegar a um profundo acordo com o problema da mente consciente não pode ter pretensões sérias de completude. (PENROSE, 1994, tradução nossa)

A consciência é parte fundamental do nosso universo, de modo que qualquer teoria física que não ceder um lugar apropriado para ela pode cair fundamentalmente na armadilha de poder ou não fornecer uma descrição verdadeira do mundo. Os campos da biologia e da neurociência não têm conseguido abarcar os fenômenos da mente de forma adequada. Muitas perguntas permanecem sem respostas. Assim, o trabalho é uma tentativa de apontar caminhos mais promissores, a partir de teorias físicas ainda incipientes e da computação, no estudo da consciência humana em suas bases mais fundamentais.

O trabalho está estruturado em três capítulos, além da introdução e das considerações finais.

Inicialmente, o capítulo 2 é dedicado às questões filosóficas sobre a irreducibilidade da consciência e seus desdobramentos. Seu início descreve a árdua

tarefa de definir o que é consciência, pois até os tempos atuais não existe um consenso a esse respeito. Também nessa parte são tratados os problemas da consciência, que David Chalmers classifica em dois gêneros distintos. O primeiro, ele nomeia como os problemas fáceis os quais são passíveis de serem elucidados sob uma ótica metodológica computacional. Já os problemas difíceis são aqueles ligados à experiência, e é exatamente aí que os argumentos sobre a irreducibilidade se concentram. São identificadas algumas resistências ao reducionismo e também são levantadas algumas visões a favor do argumento materialista. E, por fim, o capítulo apresenta algumas questões e princípios relevantes acerca de uma teoria não reducionista.

No capítulo 3, a abordagem discutida é a possibilidade de a Física ser um terreno consistente em busca de uma explicação de como o fenômeno da consciência 'surge' no mundo. Pontos sobre estruturas emergentes e sistemas complexos são considerados. A maior parte desse capítulo está voltada para a possibilidade de quais fenômenos quânticos possam ser os responsáveis pelo fluxo de consciência que vivenciamos. Tal possibilidade é centrada numa das mais conhecidas, difundidas e também controversas teorias a esse respeito: a Teoria Penrose-Hameroff, também identificada como *Orchestrated Objective Reduction* (Redução Objetiva Orquestrada). Sua descrição é detalhada para que se possa ter uma clara visão de seu objetivo, que é reduzir os processos conscientes aos fenômenos físicos dentro de um terreno ainda não estabelecido na ciência atual, a tênue e ainda obscura divisa entre a Mecânica Quântica e a Relatividade Geral. Penrose defende a existência de padrões não computáveis no cérebro, que seriam responsáveis pela geração da consciência. Tais padrões estão em certos aspectos relacionados com o problema da medição da Mecânica Quântica e algumas de suas possíveis interpretações, as quais são organizadas no decorrer desta dissertação. A parte final é dirigida às críticas ao modelo proposto, principalmente as realizadas por

Pinguelli e Faber e relacionadas aos tempos de redução dos estados superpostos<sup>6</sup> do modelo e suas relações com a decoerência quântica<sup>7</sup>.

Por fim, no capítulo 4, procura-se definir as relações entre consciência e computabilidade, ou seja, a possibilidade de implementação de algoritmos computacionais que pudessem promover estados conscientes nas máquinas compostas de circuitos eletrônicos. O capítulo é iniciado com os Teoremas de Incompletude de Gödel, de onde Penrose parte para sustentar sua proposta de que o pensamento matemático não pode ser computável, e também identifica algumas objeções funcionais à inteligência artificial forte, relativas ao comportamento criativo e flexível que os seres humanos possuem. O curso do texto segue pelas Máquinas de Turing e o problema da parada, que mostra que existem problemas que não podem ser resolvidos por tais ferramentas. Nesse caso, o problema é dito indecidível<sup>8</sup>, e a tese de Church-Turing limita o que pode ser atingido por qualquer máquina que implemente métodos computacionais. Após as discussões anteriores, chega a vez das relações mais próximas entre consciência e computação, em que níveis computacionais são descritos, de acordo com a classificação de Penrose e suas ressonâncias com o argumento do ‘quarto chinês’ de Searle. Discussões e desdobramentos relacionados à implementação de algoritmos que poderiam permitir o surgimento da consciência são considerados e a forma pela qual podem ser ‘construídos’.

Cabe afirmar sobre a completa computabilidade da consciência? A irreduzibilidade é plausível ou a inteligência artificial ainda apenas não identificou potenciais mecanismos desconhecidos no cérebro humano? A tentativa desta pesquisa é contribuir para elucidar um pouco mais tais questões.

---

<sup>6</sup> Superposição é um princípio fundamental da Mecânica Quântica que afirma que, dado um sistema físico (por exemplo, um elétron), existe em todos os seus estados particulares possíveis de forma simultânea, mas, quando uma medida do referido sistema é efetuada, o sistema ‘transita’ para apenas uma de suas configurações ou estados possíveis.

<sup>7</sup> A decoerência ambiental é uma interpretação da Mecânica Quântica na qual a interação de um estado entrelaçado com o seu ambiente desfaz os estados quânticos. Para maiores detalhes vide seção 3.3.

<sup>8</sup> Problema indecidível é um problema de decisão a partir do qual é impossível construir um algoritmo que sempre possa responder corretamente sim ou não.

## 2 QUESTÕES SOBRE A IRREDUTIBILIDADE DA CONSCIÊNCIA

### 2.1. Que é Consciência?

Antes de iniciarmos as conjecturas e discussões propriamente ditas, faz-se necessário expor as dificuldades existentes para definir alguns conceitos e definições sobre a consciência. Em virtude da natureza ainda ‘misteriosa’ da mente humana, atualmente ainda não há acordo entre os acadêmicos sobre qual é a melhor aceção para o termo consciência. Talvez nenhum aspecto da mente seja mais familiar ou mais intrigante do que a consciência e a experiência consciente de si e do mundo.

O que é uma mente? Essa mesma pergunta é feita há milênios e, ainda hoje, quando os estudiosos têm procurado esclarecer o que se entende por esse termo, uma surpreendente diversidade de opiniões e teorias surgem. Alguns sustentam que a mente é um dos muitos fenômenos da natureza e, como tal, deve ser passível de um estudo metodológico científico. Outros extrapolam o terreno da ciência e sustentam uma visão mística ou religiosa do que se entende pelo termo e posicionam esse fenômeno fora dos domínios reivindicados pela ciência formal. Pode haver uma ciência da mente? Em caso afirmativo, quais são os objetos próprios dessa ciência e quais são os limites para o conhecimento adquirido através de tal ciência?

Mesmo entre a diversidade de enfoques sobre a mente, há alguns pontos em comum. O primeiro deles é o fenômeno do pensamento. Pode-se aceitar a afirmação axiomática de que o pensamento é um atributo da mente. Se não houvesse evidência do pensar no mundo, o termo ‘mente’ teria pouco ou nenhum significado.

Outra ideia comum a todos os conceitos de mente é a de conhecimento ou saber. Enquanto cada pessoa sabe o que é conhecer algo novo, não se consegue exprimir como ocorre esse processo; assim, a natureza do conhecimento e a natureza do saber têm sido e continuam a ser temas de debate. Todos admitem,

porém, a existência de algo chamado de saber e o estado de ser que é chamado de conhecimento.

Um terceiro fator prontamente admitido por todos é que o pensamento é capaz de ser reflexivo, ou seja, o pensamento é capaz de considerar a si mesmo. Termos como 'autoconsciência' ou 'autoconhecimento' são por vezes usados como, ou quase como sinônimos, para esse poder reflexivo da mente. Se a mente não possuísse tal propriedade reflexiva, as teorias da mente<sup>9</sup> nunca poderiam surgir.

E, por último, outro fator comum a todos os conceitos da mente é a ideia de finalidade ou a chamada de intenção. Propósito ou intenção refere-se ao ato de planejar um curso de ação com conhecimento prévio do objetivo desta, ou em trabalhar de qualquer outra forma para um objetivo desejado e previsto. Esse fenômeno parece exigir algo acima e além dos sentidos e da memória. Os sentidos parecem ter um caráter imediato, existentes no agora. Memória, por outro lado, é um termo que denota a recordação de eventos passados e experiências vividas. Torna-se complexa a tarefa de explicar como a combinação dos sentidos e da memória poderia originar uma ação deliberada e direcionada ao futuro. Assim, a ideia de propósito aparece como um elemento comum em todas as teorias da mente e, às vezes, é chamado também como o poder ou a faculdade da vontade.

Pensamento, conhecimento, autoconsciência, vontade, para todos os fins parecem ser fenômenos comuns a todas as teorias da mente. Além desses, há uma série de outros termos usados por alguns teóricos sobre outros aspectos relacionados à mente. Tais termos incluem: a razão, o intelecto, o entendimento, a inteligência e a psique.

Em alguns casos, a palavra alma também é considerada como um tema abrangido pelo conceito de mente. Outros termos, como sentidos e memória, muitas vezes parecem ocupar um espaço intermediário entre a ideia de mente e a ideia de corpo. Dependendo do enfoque, sentidos e memória podem ser ligados à ideia

---

<sup>9</sup> A teoria da mente (TdM) é definida, em Psicologia, como a capacidade para imputar estados mentais aos outros e a si próprio.

de corpo, ou relacionados à ideia de mente ou a algum lugar entre esses dois domínios, que ainda parecem ser completamente diferentes um do outro.

Apesar de pensamento, conhecimento, autoconsciência e finalidade parecerem comuns a todas as teorias da mente, a diversidade entre estas começam assim que são definidos cada um dos termos anteriores. As discussões acerca do tema já se arrastam por mais de vinte séculos, quando Homero define o conceito de alma, e o debate moderno não é tão distante daqueles dos gregos antigos como se poderia supor.

Provavelmente seja impossível traçar uma fronteira classificatória consistente que possa separar as diferentes escolas de pensamento que têm como objetivo o estudo da mente e suas relações. Imperfeições na tentativa de classificar as diversas perspectivas tendem a sempre existir. Mas pode-se distinguir uma fronteira que separa três grandes grupos de estudo. Cumpre dividir aqui tais correntes, nomeando e descrevendo-as de forma genérica:

a) Materialismo - Na filosofia, a teoria do materialismo sustenta que a única coisa que existe é matéria ou energia, e que todas as coisas são delas compostas e todos os fenômenos (incluindo a consciência) são o resultado de interações materiais. Em outras palavras, a matéria é a única substância, e a realidade sobrevém dos estados de energia e matéria. Para muitos filósofos, materialismo e fisicalismo são sinônimos. Os materialistas defendem que os seres vivos são, em última análise, 'apenas' máquinas e procuram defender que a consciência e outros ainda 'obscuros' fenômenos psicológicos são apenas efeitos da estrutura biológica do cérebro.

b) Empirismo - O pilar erigido pelos empiristas acredita que os fenômenos possuem naturezas distintas. Uma é a própria realidade do mundo físico e a outra que é apreendida pelas propriedades dos sentidos que fazem o papel de 'ponte' para que os fatos do mundo possam chegar até nossa mente. O empirismo<sup>10</sup> é uma teoria do conhecimento que afirma que este se realiza basicamente através da experiência sensorial. Um dos vários ângulos da epistemologia, o empirismo,

---

<sup>10</sup> Baird e Kaufmann (2008).

ênfatiza o papel da experiênça e das provas, especialmente a percepção sensorial, na formação de ideias, sobre a noção de ideias inatas ou tradições.

c) Idealismo<sup>11</sup> – Na filosofia, o idealismo é o conjunto de abordagens que afirmam que a realidade é sobretudo mental, mentalmente construída ou, de outra forma, imaterial. Epistemologicamente, o idealismo manifesta-se como um ceticismo sobre a possibilidade de conhecer qualquer coisa independente da mente. Como uma doutrina ontológica, o idealismo vai mais longe em sua versão mais radical, afirmando que todas as entidades são compostas de mente ou ‘espírito’.

Basicamente, dessas três aceções surgem inúmeras vertentes filosóficas ligadas ao conhecimento humano e suas relações com o mundo. Na filosofia da mente, o dualismo é a suposição de que os fenômenos mentais são, em alguns aspectos, não físicos ou que a mente e o corpo sejam coisas distintas. Assim, engloba um conjunto de teses sobre a relação entre mente e matéria, e é contrastada com outras posições, como o fisicalismo, no problema mente-corpo.

Os três sistemas interagem e fornecem boas relações de suporte mútuo, mas não total. No entanto, parece ser o suficiente para colocar em segundo plano os pontos frágeis do cenário geral relacionado à consciência. As fronteiras entre cada um deles podem ser bastante tênues, mas não serão pormenorizadas, pois não fazem parte do escopo do presente trabalho.

As palavras ‘consciente’ e ‘consciência’ são termos genéricos que cobrem grande variedade de fenômenos mentais. Ambos são usados sob uma diversidade de significados. O adjetivo ‘consciente’ é heterogêneo em sua escala, sendo aplicado tanto para organismos ‘inteiros’ - consciência da criatura - e em particular, a estados e processos mentais - a consciência do estado. A palavra ‘consciência’ tem sido utilizada para a habilidade de discriminar estímulos, reportar informação, monitorar estados internos ou controlar o comportamento. Assim, um ser é considerado consciente em vários sentidos diferentes, tais como:

---

<sup>11</sup> Robinson - Encyclopædia Britannica.



a) Senciência - Pode ser consciente no sentido genérico de ser simplesmente uma criatura sensível, capaz de perceber e responder ao seu mundo (Armstrong, 1981). Ser consciente neste sentido pode admitir um modelo de gradação em que se pudessem classificar de maneira mais clara os vários níveis de consciência, dos organismos mais simples aos níveis mais complexos de uma vida consciente. De qualquer modo, as capacidades sensoriais são suficientes, mas não podem ser definidas com nitidez.

b) Vigília - Quando o organismo exerce a capacidade de 'ser', em vez apenas de ter a capacidade ou a disposição para realizá-la (vigília). Dessa forma, pode-se considerá-lo consciente quando acordados e, normalmente, alertas. Em estados alterados como o sono e quadros de coma mais profundos, os organismos não estariam conscientes ou, pelo menos, não como se entende o fenômeno da consciência até aqui. Mais uma vez, os limites de tais experiências podem ser dissolvidos, e casos intermediários podem ser válidos dentro do escopo discutido. Pode-se considerar um ser consciente sob certas circunstâncias como sonhar, estar hipnotizado ou em estado de fuga?

c) Autoconsciência - Um terceiro e ainda mais exigente sentido pode definir criaturas conscientes como aquelas que não são apenas conscientes, mas também conscientes de que estão cientes e, assim, tratar consciência da criatura como uma forma de autoconsciência (Carruthers, 2000). A exigência de autoconhecimento pode ser interpretada de várias maneiras, e assim a forma como as criaturas se qualificariam como conscientes, no sentido relevante, irá variar conseqüentemente. Em âmbito estritamente conceitual, pode-se acreditar que muitos animais e até mesmo crianças na tenra idade não podem ser classificadas como autoconscientes. Mas, se apenas as mais rudimentares formas implícitas de autoconhecimento forem necessárias, talvez uma grande variedade de criaturas possam ser qualificadas como autoconscientes.

d) Experiência - Thomas Nagel (1974) cunhou o famoso termo '*what is like*' como critério que visa capturar talvez a noção mais subjetiva de ser um ser consciente. De acordo com Nagel, um ser é consciente apenas se houver

‘algo que é como’<sup>12</sup> ser aquela criatura, isto é, de alguma forma subjetiva, o mundo parece ou aparece a partir do ponto de vista mental ou experiência individual da criatura. No exemplo de Nagel, os morcegos são conscientes porque há ‘algo que é como’ para um morcego, sua experiência do mundo através de seu sistema de eco-localização, mesmo que nós, humanos, do nosso ponto de vista humano, não possamos enfaticamente entender o que seja esse tal modo de consciência, que é o próprio ponto de vista do morcego.

Nagel<sup>13</sup> aceita que o fenômeno da consciência pode ser causado por estados físicos no cérebro, mas nega que qualquer análise de tais estados possa ser completa. Qualquer análise reducionista omite o caráter subjetivo da experiência consciente. Segundo ele, todas as tentativas realizadas, via análise reducionista, foram igualmente compatíveis com a ausência da experiência consciente. Habita na assimetria entre o objetivo e o subjetivo, que, embora seja correto examinar objetivamente um fenômeno como um relâmpago, em uma tentativa de reduzi-lo a descarga elétrica, a ideia de analisar a consciência através da experiência objetiva não é coerente. Segundo ele, a experiência só pode ser examinada pelo organismo que a tem. O que restaria do que é ser um morcego se fosse removido o ponto de vista do morcego? <sup>14</sup>

O professor Pinguelli<sup>15</sup> também acredita que atualmente ainda não exista um acordo entre os acadêmicos de qual seja a melhor acepção para o termo consciência. Concisa e claramente, define alguns conceitos importantes ligados também ao referido termo:

- a) Consciência – tem um aspecto passivo, que o físico-matemático britânico Sir Roger Penrose<sup>16</sup> chama de *awareness*, de sabermos quem somos e termos um sentimento de identidade presente na nossa

---

<sup>12</sup> Nagel (1974).

<sup>13</sup> Ibidem.

<sup>14</sup> Ibidem.

<sup>15</sup> Notas das aulas ministradas na disciplina ‘Teoria do Conhecimento I’ do HCTE.

<sup>16</sup> Penrose (1994).

mente, e outro ativo, que é o sentimento de livre-arbítrio e a capacidade de formular planos de ação para o futuro.

- b) Mente - é difícil defini-la com precisão, sendo preferível reduzi-la ao conceito de consciência, ao qual se associa; a mente pode ser vista como uma propriedade do ser humano por ter uma consciência muito desenvolvida.
- c) Entendimento – exige a presença da consciência, não se resume a ‘saber’ como fazer algo, mas entender por que se faz e para que.
- d) Inteligência - para ser genuína tem de envolver o entendimento no sentido acima, logo, ela tem como pré-condição a consciência.

Muitas explicações para consciência procuram tratar da junção sináptica entre as células nervosas existentes no cérebro como elemento básico da memória e da atividade mental. Mas essas propriedades das células parecem não ter nenhum dos potenciais rigidamente reprodutíveis do fenômeno.

Por outro lado, aqueles que veem o espírito humano como algo especial, que transcende a matéria por completo, também são deixados com um dilema desagradável. Se o atributo ‘mental-espiritual’ é imaterial, enquanto o cérebro é material, será então que nosso cérebro material tem qualquer papel ‘metafísico-mental’ em tudo? Se não, parece então negar completamente a possibilidade de uma explicação científica para a mente.

Neste cenário onde o cérebro tem algum papel auxiliar no mundo mental, então, como e onde poderia haver a interface com o componente ‘imaterial-espiritual’? Este é o problema do dualismo, que está intimamente associado com René Descartes (1641), que sustentou que a mente seria uma substância não física, ou seja, de outra ‘natureza’. Descartes identificou claramente a mente com consciência e distinguiu esta do cérebro como a sede da inteligência. Assim, ele foi o primeiro a formular o problema mente-corpo tal como existe hoje.

## 2.2. Os Problemas da Consciência

Apesar de todo o excelente trabalho que os neurocientistas têm realizado, parece que ainda não existe explicação consistente para além dos chamados problemas fáceis<sup>17</sup> da consciência, ou seja, não há muitas perspectivas que ultrapassem os aspectos estruturais e ‘mecanicistas’ do funcionamento cerebral, tais como o processamento de sinais e reconhecimento de padrões. Enquanto isso, humor, criatividade, intuição, experiência, parecem totalmente fora do escopo anterior. Estes são os chamados de problemas difíceis, que têm um caráter subjetivo e que parece confundir e frustrar qualquer tentativa de resolvê-los. É difícil visualizar como tais características do espírito humano poderiam ser incorporadas em qualquer sistema meramente mecanicista.

Em qualquer período histórico, pode-se ser cético sobre as capacidades de sua própria tecnologia contemporânea, ou dos processos biológicos que têm de ser avaliados nesses termos. A tecnologia do relógio entre os anos 1600 e 1700 certamente teve seus admiradores, mas talvez nunca tenha sido um caminho adequado para explicar o espírito humano. A favor desta visão, muitos pensadores, tais como Descartes, Hume, Kant procuravam respostas que pudessem ir além das engrenagens mecanicistas. Muitos outros os sucederam em tal empreitada, mas ainda hoje as perspectivas não parecem estar perto de uma resposta consistente para o problema.

A neurociência moderna remonta ao final do século XIX, nomeadamente através das investigações do microscópio de Ramón y Cajal<sup>18</sup>, mas era e continua sendo muito difícil conceber como apenas as conexões nervosas poderiam explicar os mais superficiais problemas da mente.

Atualmente, sabe-se muito mais sobre os detalhes dessas fibras nervosas e suas conexões sinápticas, no entanto, não é de todo óbvio que se esteja muito mais perto de explicar a consciência e seus aspectos subjetivos nesses termos. É

---

<sup>17</sup> Chalmers (1996).

<sup>18</sup> Santiago Ramón y Cajal (1852 - 1934) foi um médico e histologista espanhol. Considerado o pai da Neurociência moderna. Recebeu o Nobel de Fisiologia ou Medicina de 1906.

evidente que esses elementos estão envolvidos em atividades do cérebro de várias maneiras, mas não há ainda explicação eficaz que possa esclarecer mesmo o mais simples enigma subjetivo da mente.

David Chalmers (1996) formulou a noção do problema difícil da consciência. Assim, podem-se dividir as questões da consciência em duas frentes distintas, ou seja, a dos problemas fáceis e a dos problemas difíceis. Os problemas fáceis são aqueles que parecem diretamente suscetíveis aos métodos convencionais da ciência cognitiva, em que um fenômeno pode ser explicado em termos de mecanismos computacionais ou neurais. Os problemas difíceis são aqueles que parecem resistir, insistentemente, a esses métodos.

Os chamados problemas fáceis da consciência dizem respeito àqueles que são passíveis de explicar os seguintes fenômenos:

- Capacidade de discriminar, categorizar e reagir a estímulos ambientais;
- Integração de informações por um sistema cognitivo;
- Descrição de estados mentais;
- Capacidade de um sistema para acessar seus próprios estados internos;
- Foco de atenção;
- Diferença entre a vigília e o sono.

Ainda não existe um consenso sobre se todos os fenômenos descritos acima podem ser explicados cientificamente. Mas todos eles parecem ser diretamente passíveis de uma explanação segundo a metodologia científica atual.

O problema realmente difícil da consciência é o problema da experiência e suas decorrências (conforme citadas anteriormente), ou também chamado o problema da experiência consciente. Quando pensamos e percebemos, há todo um processo fisiológico de processamento de informação, mas há também um aspecto subjetivo, e este é o ponto central.

Como Nagel<sup>19</sup> ressaltou, ‘há algo que é como ser um organismo consciente’. Esse aspecto subjetivo é a experiência. Por exemplo, quando vemos ou experimentamos sensações visuais, a qualidade sentida de vermelhidão, a experiência de luz e escuridão, a qualidade de profundidade em um campo visual. Outras experiências estão inseridas com a percepção em diferentes modalidades: o som de um violino, o cheiro de gasolina. Depois, há sensações corporais de dor, orgasmo; imagens mentais que são evocadas internamente, a qualidade da emoção sentida, e a experiência de uma corrente de pensamento consciente. Todas elas podem ser consideradas como estados de experiência.

Decorrente ainda do contexto do problema difícil, encontra-se também o chamado *binding problem*, ou o problema de ligação. Em primeiro lugar, há a questão prática de como os cérebros segregam elementos em padrões complexos de dados. Por exemplo, ‘Quando eu vejo um quadrado azul e um círculo amarelo, que mecanismos neurais garantem que a detecção de azul é acoplada de alguma forma ao quadrado e que como o amarelo é acoplado ao círculo?’ Em segundo lugar, há o problema mais fundamental de ‘como a unidade da percepção consciente é provocada pelas atividades distribuídas do sistema nervoso central’. A primeira pergunta é uma pergunta difícil, mas suscetível a uma explicação analítico-funcional dentro da ciência física, que poderia igualmente ser aplicada a um computador ou qualquer outro sistema complexo, constituído de entradas e saídas. A segunda questão é metafísica, no sentido de que a ‘unidade de percepção consciente’ pode ser uma ideia fora da ciência física que requer uma base metafísica ou ontológica. Assim, ‘unidade’, nesse sentido, não tem significado físico, mas tem um significado crucial na experiência subjetiva.

O problema levantado por Nagel<sup>20</sup>, também chamado de *gap* explanatório, é a alegação de que a consciência e as experiências humanas qualitativas ou *qualia*<sup>21</sup>,

---

<sup>19</sup> Nagel (1974).

<sup>20</sup> Ibidem.

<sup>21</sup> Sensações e experiência. Quando se considera a experiência visual individual, como você olha para uma mancha brilhante de cor azul-turquesa em uma loja de tintas. Há algo que é pra você subjetivamente se submeter a essa experiência. O que é como se submeter a experiência é muito diferente de como é para você experimentar uma amostra de azul-turquesa. Essa diferença é uma diferença chamada de ‘caráter fenomenal’. O caráter fenomenal de uma experiência é o que é subjetivamente se submeter à experiência.

não podem ser totalmente explicadas apenas através da identificação dos correspondentes processos físicos (neurais). Essa lacuna é conhecida como o 'problema difícil' ou *gap* explanatório, e tem intrigado filósofos e pesquisadores da área da inteligência artificial (aqui designada IA) ao longo de décadas e tem sido ainda objeto de consideráveis conjecturas e debates.

O problema difícil da consciência muitas vezes é designado como consciência fenomenal ou *qualia*, mas alguns autores, entre eles Chalmers<sup>22</sup> e Newell<sup>23</sup> preferem designá-lo como 'experiência consciente' ou simplesmente 'experiência'. Outra possibilidade útil para evitar confusão é reservar o termo 'consciência' para os fenômenos de experiência e usar o termo 'percepção' para os fenômenos mais simples e passíveis de explanação como descrito anteriormente.

### **2.2.1. Os Aspectos Fenomenais da Consciência**

O nexos principal da relação entre a consciência e a cognição encontra-se nos chamados conceitos fenomenais. A experiência consciente não reside em um vazio fenomenal isolado. A pessoa está ciente da sua experiência e seu conteúdo, estabelece julgamentos sobre ela e é levada a fazer declarações sobre tal fato. Quando se tem uma sensação de vermelho, forma-se uma crença em relação a ela que pode resultar num relato verbal.

Pode-se considerar em primeiro lugar o conceito expresso por 'vermelho' em 'experiência de vermelho' como a expressão pública da língua, ou o conceito expresso pela 'vermelhidão fenomenal'. A referência dessas expressões é fixada através de uma relação a coisas vermelhas no mundo externo e, finalmente, através de uma relação a certos objetos vermelhos que são expressos ao aprender o termo público 'vermelho' da língua. Um aprendiz da língua aprende a chamar as experiências normalmente provocadas pelo 'vermelho' desses objetos (no sentido fenomenal) e a chamar os objetos que normalmente trazem o 'vermelho' nessas experiências (no sentido externo). Assim, o conceito fenomenal aqui é relacional, na medida em que tem sua referência fixa, uma relação com objetos externos. A

---

<sup>22</sup> Chalmers (1996).

<sup>23</sup> Newell (1990).

propriedade que é referida não precisa ser relacional, no entanto. O conceito fenomenal plausivelmente designa uma propriedade intrínseca de forma rígida, mas admite a possibilidade de existirem mundos contrafactuais<sup>24</sup> em que as experiências de ‘vermelho’ nunca são causadas por coisas vermelhas.

Pode-se distinguir, pelo menos, dois conceitos fenomenais relacionais, dependendo da referência, que são fixados por relações a partir de um conjunto global, ou por relações restritas ao assunto em questão. O primeiro é o que se pode chamar o conceito relacional global, por exemplo, o ‘vermelho G’. Essa afirmação pode ser explicada, de modo aproximado, como a qualidade tipicamente fenomenal causada em indivíduos normais a partir da comunidade de coisas vermelhas. O segundo é o que se pode chamar o conceito individual relacional, ou ‘vermelho I’. Neste caso, ele pode ser explicado como a qualidade tipicamente fenomenal causada em mim pelas coisas vermelhas. Os dois conceitos G e I irão referenciar as mesmas coisas para indivíduos normais, mas para os indivíduos fora de um padrão de normalidade podem levar a resultados diferentes. Por exemplo, o conceito vermelho G em um sujeito com um determinado grau de daltonismo, para quem a percepção de verde e vermelho é invertida, irá se referir ao que normalmente é denominado (o que outros chamam) ‘vermelhidão’ fenomenal, mas o vermelho de seu próprio conceito, o vermelho I, terá um caráter de ‘verdidão’ fenomenal.

### 2.3. Resistências ao Reduccionismo

O filósofo americano John Searle constrói suas ideias sobre a consciência a partir do conceito de intencionalidade. Searle<sup>25</sup> define intencionalidade como a capacidade da mente em representar as coisas, suas propriedades e seus estados no mundo. Baseando seus pontos de vista sobre a intencionalidade, Searle apresenta uma visão sobre a consciência em seu livro *The Rediscovery of the*

---

<sup>24</sup> Em lógica modal e metafísica, contrafactual representa a situação ou evento que não aconteceu, mas poderia ter acontecido. A situação ou evento que aconteceu é chamado de atual. Diz-se que o evento contrafactual faz parte de um mundo possível, enquanto o evento atual faz parte do mundo atual.

<sup>25</sup> Searle (1983).



*Mind*<sup>26</sup>. Nesse livro, Searle argumenta que grande parte da filosofia moderna tentou negar a existência da consciência, mas com pouco sucesso.

A partir da intencionalidade, ele faz analogias interessantes de algumas teorias, substituindo suas considerações acerca da intencionalidade através do exemplo do uso da mão. Afirma que seria muito limitante dizer que:

Ter uma mão é apenas estar disposto a certos tipos de comportamento, como agarrar algum objeto - behaviorismo. (SEARLE, 1994, p. 263)

A mão pode ser definida inteiramente em termos de suas causas e efeitos - funcionalismo. (SEARLE, 1994, p. 263)

Searle<sup>27</sup> argumenta que a filosofia tem sido aprisionada por uma pretensa falsa dicotomia: por um lado, o mundo consiste apenas de partículas físicas e campos de força de caráter objetivo e, por outro lado, a existência da consciência, que é claramente uma experiência subjetiva de primeira pessoa. Reforça que ambos são verdadeiros: a consciência é uma verdadeira experiência subjetiva, possivelmente causada pelos processos físicos do cérebro. Esse é o posicionamento que assume e é conhecido por naturalismo biológico.

O naturalismo biológico pretender ser uma teoria monista<sup>28</sup>, uma abordagem para o problema mente-corpo e pode ser definida por duas teses principais:

- 1) Todos os fenômenos mentais (dor, sensação, pensamento, etc.) são causados pelos processos neurobiológicos de baixo nível no cérebro;
- 2) Os fenômenos mentais são características de alto nível do cérebro.

Tais assertivas implicam que o cérebro tem capacidade causal para produzir intencionalidade. No entanto, o naturalismo biológico de Searle não implica que somente os cérebros podem causar consciência. Searle é prudente ao salientar que, mesmo parecendo que certas funções cerebrais são suficientes para a produção de estados de consciência, o nosso estado atual do conhecimento neurobiológico nos

---

<sup>26</sup> Searle (1994).

<sup>27</sup> Ibidem.

<sup>28</sup> Monismo é qualquer visão filosófica que sustenta que há uma unidade em um determinado campo de investigação. Sob esta ótica, pode-se julgar que o universo é uno; não possui qualquer caráter dualista, ou seja, diferentes naturezas.

impede de concluir que elas sejam necessárias e exclusivas para produzir consciência. Em suas próprias palavras:

O fato de que os processos cerebrais causam estados conscientes não implica que o cérebro possa ser consciente. O cérebro é uma máquina biológica, e poderíamos construir uma máquina artificial que fosse consciente. Da mesma forma como o coração é uma máquina, e temos construído corações artificiais. Como não sabemos exatamente o funcionamento do cérebro, ainda não estamos em posição de construí-lo de forma artificial. (SEARLE, 2007, p. 328)

Searle nega qualquer tipo de dualismo, a alternativa tradicional para o monismo, alegando que tal distinção é um erro. Acredita que a consciência é tanto a causa dos acontecimentos no corpo, bem como uma resposta aos eventos do corpo. Não trata a consciência como um fantasma na máquina, mas como um estado do cérebro. A interação causal entre mente e cérebro pode ser assim descrita, em termos naturalistas: Eventos no nível micro (nível neuronal) causam a consciência. Alterações no nível macro (todo o cérebro) constituem a consciência. Micro mudanças causam mudanças sistêmicas no cérebro e também são impactadas pelos efeitos dessas mudanças ao nível macro.

Searle<sup>29</sup> articula uma distinção ao termo filosófico 'reduzível', pois, segundo ele, este é ambíguo. Afirma que a consciência é causalmente reduzível aos processos cerebrais, porém, sem ser passível a uma reduzibilidade ontológica. Espera que, ao fazer esta distinção, possa escapar do dilema tradicional entre o materialismo reducionista e o dualismo de substância.

Parece uma via conclusiva natural ver tal teoria sustentando um posicionamento de dualismo de propriedades, uma vez que as propriedades mentais de uma pessoa são categoricamente diferentes de suas micropropriedades físicas. Estas últimas têm 'ontologia de terceira pessoa', enquanto as primeiras, uma 'ontologia de primeira pessoa'. As microestruturas são acessíveis objetivamente, como por exemplo, através de um cirurgião, mas a dor, o desejo ou uma crença são acessíveis subjetivamente, somente pela pessoa que as experimenta. Tenta fugir ao dualismo afirmando que as propriedades mentais são uma espécie de propriedade física, mas de uma ontologia de primeira pessoa.

---

<sup>29</sup> Searle (2007).

## 2.4. A Questão da Superveniência da Consciência e o Mundo Físico

Acredita-se que os fatos mais fundamentais sobre o universo são fatos físicos e que todos os outros fatos são dependentes destes. Em um sentido fraco, de dependência, essa constatação pode ser quase trivialmente verdadeira; porém, em um sentido forte, é controverso. Há uma variedade complexa de relações de dependência entre fatos de alto e baixo níveis dos fatos em geral, e do tipo de relação de dependência que mantém em um domínio como a biologia, estes não podem exercer em outro, como o da experiência consciente. A noção filosófica de superveniência fornece um quadro unificador em que essas relações de dependência podem ser discutidas.

A noção de superveniência formaliza a ideia intuitiva de que um conjunto de fatos pode determinar totalmente outro conjunto deles. Os fatos físicos sobre o mundo parecem determinar os fatos biológicos. Por exemplo, uma vez que todos os fatos físicos sobre o mundo são fixos não há espaço para os fatos biológicos variarem. Tal evidência fornece uma caracterização grosseira do sentido em que as propriedades biológicas sobrevivem das propriedades físicas. Em geral, a superveniência é uma relação entre dois conjuntos de propriedades. Por exemplo, propriedades de baixo nível (nível atômico) e propriedades de alto nível (os tecidos biológicos).

Como explicar a consciência? Quase tudo no mundo pode ser explicado em termos físicos, por essa razão é natural esperar que a consciência assim o possa também. No entanto, parece que a consciência escapa da rede de explicação reducionista. Até o momento, nenhuma explicação consistente baseada apenas em termos físicos parece contemplar a emergência da experiência consciente.

Assim, pode-se notar que a consciência não é logicamente superveniente ao plano físico. Em princípio, é necessário mostrar que ela não seja superveniente globalmente, isto é, que todos os fatos microfísicos no mundo não implicam os fatos sobre a consciência. Na prática, é mais fácil processar o argumento localmente, no caso, argumentando que em um indivíduo, onde os fatos microfísicos não implicam

os fatos sobre a consciência. De acordo com Chalmers<sup>30</sup>, quando se trata da consciência, a superveniência local e global parecem não se sustentar, conseqüentemente não importa muito o caminho percorrido: se a consciência é superveniente de forma global, é coerente dizer que também o é localmente.

#### **2.4.1. Argumentos contra o Fisicalismo**

Como argumentar que a consciência não é logicamente superveniente ao plano físico? Segundo Chalmers<sup>31</sup>, existem várias maneiras:

a) Pode-se pensar sobre o que é concebível, a fim de discutir diretamente a possibilidade lógica de uma situação em que os fatos físicos são os mesmos, mas os fatos sobre a experiência são diferentes;

b) Convocar a epistemologia, argumentando que o tipo certo de ligação entre o conhecimento de fatos físicos e de conhecimento da consciência está ausente;

c) Pode-se também recorrer diretamente ao conceito de consciência, argumentando que não há análise do conceito de que poderia fundamentar uma vinculação do físico para o fenomenal.

A seguir, são apresentados alguns argumentos contra a ideia materialista da mente. No presente trabalho, o foco é concentrado nos argumentos pertencentes aos itens *a* e *b*. O primeiro é essencialmente o argumento de conceitabilidade, o segundo e o terceiro são relacionados à epistemologia. Há algum elemento de redundância entre eles, mas, juntos, fazem um argumento relevante.

#### **1º Argumento: A possibilidade lógica de zumbis**

Segundo Chalmers<sup>32</sup>, a maneira mais óbvia (embora não a única) para investigar a superveniência lógica da consciência é considerar a possibilidade lógica de um zumbi: alguém ou alguma coisa fisicamente idêntica a um ser consciente, no caso um ser humano comum, mas que lhe falta consciência da

---

<sup>30</sup> Chalmers (1996).

<sup>31</sup> Idibem.

<sup>32</sup> Idibem.

experiência consciente. Em âmbito global, pode-se considerar a possibilidade lógica de um mundo zumbi: um mundo fisicamente idêntico ao nosso, mas onde não há experiências conscientes. Em tal mundo, todos são zumbis.

Continuando a argumentação, são considerados dois irmãos gêmeos, sendo o segundo um zumbi. O segundo, em relação ao primeiro, é 'idêntico' com respeito à estrutura biológica e, assim sendo, as mesmas propriedades de baixo nível postuladas pela física estabelecida, mas falta-lhe a experiência consciente. Pode-se imaginar o primeiro olhando para fora de uma janela, experimentando algumas sensações agradáveis, por exemplo, o verde das árvores lá fora ou tendo uma experiência do sabor agradável através do consumo de uma barra de chocolate e sentindo uma sensação de dor desagradável em seu ombro direito.

O que está acontecendo com o irmão zumbi? Ele é fisicamente idêntico ao primeiro, e pode-se supor que esteja inserido em um ambiente idêntico. Ele certamente será funcionalmente idêntico ao primeiro: realizará o mesmo tipo de processamento de informação, reagindo de modo similar aos *inputs* externos, e suas 'configurações' internas sendo modificadas de forma adequada e com um comportamento indistinguível resultante. Ele vai ser psicologicamente idêntico ao primeiro.

O irmão zumbi irá perceber as árvores no sentido funcional, e a degustação do chocolate no sentido psicológico. Tudo isso decorre logicamente do fato de que ele é fisicamente idêntico ao primeiro, em virtude das análises funcionais. Ele vai mesmo ser 'consciente' nos sentidos funcionais descritos anteriormente, capaz de relatar o conteúdo de seus estados internos, capaz de concentrar a atenção em vários lugares, assim por diante. Ocorre que nada desse funcionamento será acompanhado por qualquer experiência real consciente. Não haverá sensação fenomenal. Esse tipo de zumbi é bastante diferente dos zumbis encontrados em filmes, que tendem a ter deficiências funcionais significativas. O tipo de consciência que os zumbis dos filmes mais carecem é uma versão psicológica: normalmente, têm pouca capacidade de introspecção e lhes falta uma habilidade voluntária refinada de controlar o comportamento.

Conforme Chalmers<sup>33</sup> mesmo concorda, a ideia dos zumbis soa bastante estranha, e é improvável que ela seja empiricamente possível. Na prática, é provável que qualquer réplica do primeiro gêmeo fosse consciente. Mas a questão não é se é plausível que zumbis poderiam existir em nosso mundo, a questão é saber se a noção de um zumbi é coerente. Em face dessa assertiva, a noção parece inteiramente inteligível. Se tal hipótese estiver correta, a conclusão é estabelecida.

Argumentando a favor de uma possibilidade lógica que não é inteiramente clara, parece apenas, embora tal coisa não exista no mundo real, a descrição certamente reveste-se de coerência. Seria similar conceber a existência de uma bicicleta gigantesca, com quilômetros de altura. Fica bastante claro que não há contradição oculta na descrição.

Para algumas pessoas, intuir sobre a possibilidade lógica de uma réplica sem experiência consciente parece menos clara no início, talvez porque a ocorrência simultânea da bioquímica e da consciência pode levar a supor uma ligação conceitual. Mas uma vez aceita a possibilidade de que essas réplicas funcionais não conscientes sejam logicamente possíveis, a conclusão correspondente sobre uma réplica física não pode ser evitada.

Por exemplo, pode-se indiretamente apoiar a alegação de que os zumbis são logicamente possíveis, levando em conta interações fora dos padrões da própria organização funcional. O padrão de organização causal incorporado nos mecanismos responsáveis pela produção do comportamento pode, em princípio, ser realizado de modos não convencionais. Para utilizar um exemplo hipotético utilizado por Block<sup>34</sup>, o povo de uma grande nação como a China poderia se organizar de modo que pudesse desempenhar uma organização isomórfica causal, onde cada pessoa pudesse simular o comportamento de um único neurônio e fosse conectada uma à outra, através de conexões de ondas de rádio, correspondentes às sinapses. A população poderia controlar o corpo de um robô, equipado com transdutores sensoriais e estrutura motora.

---

<sup>33</sup> Chalmers (1996).

<sup>34</sup> Block (1978).

Muitas pessoas acham improvável que uma configuração como essa daria origem à experiência consciente, que de alguma forma uma 'mente coletiva' surgiria a partir do sistema global. Tudo o que se precisa aqui é que a possibilidade de que tal sistema não tenha a experiência consciente seja coerente. A possibilidade significativa está sendo expressa, e é uma questão aberta se a consciência surge ou não a partir desse fato.

Pode-se considerar de forma semelhante, o que Chalmers<sup>35</sup> chama de isomorfo de silício, que é organizado como um ser humano comum, mas que possui chips de silício onde o ser comum possui neurônios. Se tal isomorfo seria de fato consciente é controverso, mas parece que a maioria das pessoas que nega tal fato está expressando uma possibilidade coerente. A partir desses casos, conclui-se que a existência da experiência consciente não é logicamente decorrente dos fatos sobre a organização funcional.

Mas dado que é conceitualmente coerente que o exemplo do isomorfo de silício poderia carecer da experiência consciente, segue-se que o irmão gêmeo zumbi também se torna uma possibilidade igualmente coerente. Pois é claro que não há mais de um vínculo conceitual da bioquímica para a consciência do que há de silício para o homúnculo chinês. Se o isomorfo de silício, sem a experiência consciente é concebível, precisa-se apenas de neurônios substitutos para o silício, deixando a organização funcional constante, e tem-se assim o gêmeo zumbi. Nada nessa substituição poderia forçar o caráter da experiência nessa concepção. Essas diferenças na implementação simplesmente não seriam algo que poderia ser conceitualmente relevante para a experiência. Assim, a consciência não pode ser logicamente superveniente ao plano físico.

Alguns podem pensar que os argumentos de conceitabilidade não sejam suficientes. É razoável afirmar que não é possível imaginar em detalhes os bilhões de neurônios do cérebro humano. Não é necessário imaginar cada um dos neurônios para construir um cenário coerente. Complexidade simples entre os neurônios não poderia conceitualmente implicar a consciência, pois se toda a

---

<sup>35</sup> Chalmers (1996)

estrutura neural é relevante para a consciência, isso deve acontecer em virtude de algumas propriedades de nível superior que ela proporciona.

Assim, o fato de que existe consciência no mundo real deve ser um fato para além de todos os fatos físicos. O materialismo, que está comprometido com a afirmação de que todos os fatos são fatos físicos, é, portanto, falso.

## **2º Argumento: O Argumento do Conhecimento**

Concebido por Frank Jackson em seu artigo de 1992, intitulado *Epiphenomenal qualia*, no qual imagina:

Uma neurocientista chamada Mary, que é forçada a investigar o mundo de dentro de um quarto preto e branco, aparelhado com um monitor de televisão preto e branco, sem jamais ter tido acesso ao mundo externo colorido. Mary é especialista em neurofisiologia da visão e adquire toda a informação física que se pode obter sobre as cores e sobre todos os processos cerebrais envolvidos na identificação de cores – por exemplo, sobre como os diversos comprimentos de onda provenientes do céu estimulam a retina e como estes afetam o sistema nervoso central até a verbalização da frase ‘O céu é azul’.

A questão crucial para o argumento de Jackson é: o que acontece quando Mary sai da sala em preto e branco pela primeira vez? Se Mary realmente aprende algo novo ao experimentar como é a sensação do vermelho ou do azul, etc., então seu conhecimento anterior, baseado em fatos físicos apenas, era incompleto. Jackson conclui: “Se toda a informação acerca de fatos físicos não é suficiente para conhecermos fatos sobre a consciência, então o materialismo é falso”.<sup>36</sup>

Segue-se então que os fatos sobre a experiência subjetiva da visão das cores não são decorrentes dos fatos físicos. Se assim fossem, Mary poderia, em princípio, vir a saber o que é gostar de ver a cor vermelha, com base em seu conhecimento dos fatos físicos. Mas ela não pode. Talvez Mary poderia vir a saber o que é gostar de ver vermelho por algum método indireto, tal como a manipulação de seu cérebro da maneira apropriada. O ponto, entretanto, é que o conhecimento não segue a partir do conhecimento do mundo físico e suas propriedades. O conhecimento de todos os fatos físicos, em princípio, permitiria a Mary deduzir todas as reações de um sistema, através de suas habilidades e

---

<sup>36</sup>Jackson (1992).



capacidades cognitivas, mas ela ainda permanece completamente sem saber como é ter a experiência de vermelho.

Dessa maneira, pode-se considerar um computador como um agente cognitivo simples, semelhante aos humanos em certos aspectos, como a sua capacidade de discriminação sensorial. Tem a capacidade de categorizar estímulos de cor de modo bastante semelhante ao do ser humano, o agrupamento de coisas que poderiam ser classificadas como 'vermelhas' em uma categoria, e as coisas 'verdes' sob outra. Mesmo conhecendo todos os detalhes sobre os circuitos constituintes do computador, são questões em aberto:

- 1 - O computador está experimentando alguma coisa quando categoriza rosas?
- 2 - Se sim, ele experimenta a qualidade sensorial da cor da mesma forma quando um ser humano vê uma rosa, ou processa algo muito diferente?

Essas são questões significativas e conhecer todos os fatos físicos não implica uma resposta definitiva. Os fatos físicos, portanto, não implicam logicamente os fatos sobre a experiência consciente.

Ainda de acordo com Chalmers<sup>37</sup>, Jackson<sup>38</sup> apresentou argumento do conhecimento como um argumento contra o fisicalismo e não contra uma explicação reducionista. É interessante notar que a maioria das objeções ao argumento contra o fisicalismo admitia o ponto que é relevante para o argumento contra a explicação reducionista: que o conhecimento do que é vermelho, como é do conhecimento factual, não está implicado em um conhecimento *a priori* dos fatos físicos. A única maneira que essa conclusão pode ser evitada é a de negar que saber o que é a experiência de ver o vermelho acarreta o conhecimento de um novo fato. Essa é a estratégia tomada por Lewis<sup>39</sup> e Nemirow<sup>40</sup>, que argumentam que tudo o que falta a Mary é a habilidade de reconhecer as coisas vermelhas. Mas parece lógico que quando vê o vermelho pela primeira vez, ela acrescenta algo novo sobre como o

---

<sup>37</sup> Chalmers (1996).

<sup>38</sup> Jackson (1992).

<sup>39</sup> Lewis (1990)

<sup>40</sup> Nemirow (1990)

mundo é; parece razoável que o conhecimento que ela está ganhando é o conhecimento de um fato.

### **3º Argumento: Assimetria Epistêmica**

As bases para a crença na consciência derivam direta e exclusivamente da nossa própria experiência em relação ao mundo que nos cerca. Mesmo conhecendo todos os fatos, todos os detalhes sobre a física do universo, as relações de causalidade entre todos os campos e partículas no tecido espaço temporal, tais informações não nos levaria a postular a existência da experiência consciente. O conhecimento da consciência provém da experiência individual e não de nenhuma observação externa.

A partir dos fatos de baixo nível sobre as configurações físicas e suas relações causais, pode-se, em princípio, derivar todos os tipos de fatos de alto nível relacionados aos sistemas macroscópicos, sua organização, e o nexos de causalidade entre eles. Pode-se determinar também as correlações existentes entre as funções biológicas e o comportamento humano e sobre os mecanismos cerebrais pelos quais ele é causado. Mas nada nessa ampla cadeia causal levaria à conclusão da existência de algum tipo de consciência.

É razoável afirmar que os fatos sobre o mundo físico podem fornecer alguma evidência indireta para a existência da consciência. Há uma assimetria epistêmica em nosso conhecimento da consciência que não está presente no nosso conhecimento de outros fenômenos. Nosso conhecimento de que a experiência consciente existe deriva principalmente da nossa própria perspectiva, com a evidência externa desempenhando, na melhor das hipóteses, um papel secundário.

O argumento também pode ser demonstrado ao recorrer para a existência de outras mentes. Mesmo quando se sabe tudo sobre os aspectos físicos e biológicos das outras criaturas, não há como afirmar com certeza se estão conscientes ou quais são suas experiências. Ainda segundo Chalmers<sup>41</sup>, quando o alvo da questão muda, como, por exemplo, o problema de 'outras economias', ou de 'outras

---

<sup>41</sup> Chalmers (1996).

fisiologias', não há assimetria epistêmica nesses casos, precisamente porque esses fenômenos são logicamente supervenientes ao plano físico.

A assimetria epistêmica no conhecimento da consciência deixa claro que a consciência não pode ser logicamente superveniente. Se assim fosse, não haveria tal assimetria epistêmica; uma propriedade logicamente superveniente pode ser detectada diretamente com base em evidências externas, e não há nenhum papel para o caso da experiência de primeira pessoa. A assimetria epistêmica associada com a consciência é muito mais fundamental, e essa circunstância revela que nenhum conjunto de fatos sobre a cadeia complexa causal em sistemas físicos pode resultar num processo consciente.

Dentro do contexto apresentado, um fenômeno é redutível em termos físicos somente se for logicamente superveniente às propriedades físicas.

Assim, Chalmers<sup>42</sup> adota uma versão do dualismo de propriedade que ele chama de dualismo naturalista. Embora aceite a alegação materialista de que as únicas substâncias no mundo são as físicas, ele se afasta do materialismo ao afirmar que algumas dessas substâncias têm propriedades fenomenais que não podem ser reduzidas em termos de propriedades físicas. Importante, entretanto, a irreduzibilidade das propriedades fenomenais implica apenas que elas são logicamente independentes de propriedades físicas. Não é preciso concluir que as propriedades fenomenais sejam realmente independentes das propriedades físicas, e, de fato, ele nega que sejam.

Em suma, embora negue que a consciência sobrevém logicamente ao mundo físico, ele acredita que ela sobrevém naturalmente ao físico (daí o termo dualismo naturalista).

Não há como negar que essa posição afirma não apenas uma tênue linha entre esse ponto de vista, supostamente dualista, e os pontos de vista de muitos dos seus adversários que se dizem materialistas. Curiosamente, em alguns aspectos,

---

<sup>42</sup> Ibidem.

parece não haver tal divisão. Nesses casos, no entanto, Chalmers<sup>43</sup> afirma que seus adversários estão se enganando sobre os compromissos ontológicos de suas teorias.

## 2.5. A Visão Materialista

A tese do filósofo Daniel Dennett<sup>44</sup> em seu livro *Consciousness Explained* é que nossa compreensão moderna da consciência é indevidamente influenciada pelas ideias de René Descartes. O livro apresenta o chamado 'Modelo de Rascunhos Múltiplos' da consciência, com o intuito de mostrar que não há um único lugar central, por ele chamado de 'teatro cartesiano', onde ocorre a experiência consciente. Ao invés disso, há inúmeros eventos de fixação de conteúdo ocorrendo em vários lugares em vários momentos no cérebro.

O cérebro consiste de um pacote de atividades semi-independentes. Quando os conteúdos de fixação têm lugar em uma dessas atividades, seus efeitos podem se propagar para levar a emissão de sentenças que compõem o caráter central do personagem do *self*. Sua visão sobre a consciência considera aparentemente uma série de paralelismos que ocorrem de forma subjacente no cérebro.

Uma das declarações mais controversas do livro é a de que o conceito de *qualia* não pode existir. Seu principal argumento é que as várias propriedades atribuídas aos *qualias* pelos filósofos são supostamente incorrigíveis, inefáveis, diretamente acessíveis, e assim por diante. Segundo ele, parece haver uma incompatibilidade no conceito, por isso a noção de *qualia* é incoerente. A não existência de *qualia* significaria que não há o problema difícil da consciência, e os 'zumbis filosóficos', que supostamente agiriam como um ser humano em todos os sentidos, mas ao mesmo tempo lhes faltaria o *qualia*, não podem, portanto, existir. Assim, Dennett observa, com ironia, que está comprometido com a crença de que somos todos zumbis e acrescenta que sua observação é muito aberta a

---

<sup>43</sup> Chalmers (1996).

<sup>44</sup> Dennett (1991).

interpretações erradas. Segundo ele, o argumento do conhecimento fenomenal (*qualia*) não constitui uma válida refutação ao fisicalismo.

Dennett tem interesse na capacidade da Teoria da Evolução para explicar algumas das características de produção de conteúdos da consciência. Defende uma teoria conhecida por alguns como darwinismo neural. Procura abordar as questões a partir de uma perspectiva evolucionária.

Afirma que nossos cérebros mantêm apenas alguns detalhes relevantes sobre o mundo, e que esta é a única razão pela qual somos capazes de ‘funcionar’. Assim, não armazenamos imagens elaboradas na memória de curto prazo, pois essa tarefa não seria necessária e iria consumir considerável poder de processamento. Em vez de tal ação, registramos apenas o que mudou e assumimos o ‘resto’ como o mesmo, resultando na perda de alguns detalhes, como demonstrado em vários experimentos esboçados por Dennett<sup>45</sup> e outros cientistas, entre eles Laura Spinney<sup>46</sup>.

Um dos experimentos mais famosos é o chamado ‘cegueira da mudança’ (*change blindness*), que aparentemente é um fenômeno normal do cérebro que demonstra que este não tem uma representação precisa do mundo, mas uma representação lacunar, feita de detalhes parciais. Apesar do nome, esse fenômeno não afeta os olhos, mas o cérebro e, como tal, talvez possa se correlacionar também aos outros sentidos humanos. Esse fenômeno ainda está em pesquisa, mas os resultados sugerem que o cérebro estima a importância e a utilidade das informações antes de decidir armazená-las ou não. Outra questão é que o cérebro não pode ver uma mudança acontecendo a um elemento que ainda não tenha armazenado. As primeiras investigações sobre a ‘cegueira de mudança’ parecem ter sido conduzidas por George McConkie<sup>47</sup> e seus colegas no final da

---

<sup>45</sup> Dennett (1991).

<sup>46</sup> Spinney (2000).

<sup>47</sup> McConkie (1996).

década de 1970, e incidiu sobre as alterações feitas a palavras e textos durante os movimentos oculares sacádicos<sup>48</sup>.

De acordo com o modelo, há uma variedade de estímulos sensoriais a partir de um determinado evento e também uma variedade de interpretações dessas entradas. Tais entradas sensoriais chegam ao cérebro e são interpretadas em diferentes momentos, assim um determinado evento pode gerar uma sucessão de discriminações, constituindo o equivalente a vários rascunhos de uma história. Logo que cada discriminação é realizada, torna-se disponível para resultar em um determinado comportamento.

Dennett<sup>49</sup> propõe uma explicação de alto nível da consciência que seria plausível a um suporte consistente para a inteligência artificial forte. Descreve a teoria como operacionalismo em primeira pessoa. Afirma que, no modelo dos rascunhos múltiplos, o processo de 'escrever' na memória serve como base de avaliação para a consciência. Não há realidade da experiência consciente, independente dos efeitos de vários veículos de conteúdo em ação subsequente no cérebro. Cunha o termo 'consciência narrativa' para os estados que podem ser concebidos como conscientes em sentido narrativo, recorrendo para a noção de 'fluxo de consciência', considerado como uma narrativa em curso mais ou menos em série de episódios a partir da perspectiva de um eu meramente virtual. A ideia seria equiparar os estados mentais do sujeito aos que surgem em tal fluxo.

O modelo dos rascunhos múltiplos corrobora com a necessidade de um certo tempo para a experiência consciente ocorrer, de modo que percepções não instantâneas surjam na mente, em sua total riqueza de detalhes. A distinção é que a teoria de Dennett nega qualquer fronteira clara e inequívoca à separação entre experiências conscientes de todas as outras transformações. De acordo com ele, a

---

<sup>48</sup> Movimentos sacádicos são movimentos rápidos do olho. Também pode ser uma mudança demasiadamente rápida na frequência de um sinal. Movimentos rápidos e simultâneos de ambos os olhos na mesma direção. São relacionados aos lobos frontal e parietal do cérebro, servindo como um mecanismo de fixação do movimento rápido dos olhos. A palavra parece ter sido cunhada em 1880 pelo oftalmologista francês Émile Javal, que usou um espelho de um lado de uma página para observar o movimento dos olhos na leitura silenciosa e descobriu que ele envolve uma sucessão de movimentos individuais descontínuos.

<sup>49</sup> Dennett (1991).

consciência encontra-se nas ações e fluxos de informação de local para local, em vez de alguma visão singular que contenha nossa experiência. Não há experimentador central que confira um selo de aprovação durável sobre qualquer ação em particular.

Diferentes partes do processamento neural parecem tomar o controle do processo em momentos diferentes. A consciência é a propriedade de possuir influência suficiente para afetar o que a boca vai dizer e o que as mãos farão. Quais entradas serão 'editadas' em nossos rascunhos não é um ato exógeno de supervisão, mas apenas parte do funcionamento de auto-organização da rede, da mesma forma que um circuito eletrônico transmite informações de baixo para cima.

Críticos de sua abordagem, tais como David Chalmers e Thomas Nagel, apontam que o argumento de Dennett se perde completamente ao redefinir a consciência como uma propriedade externa e ao ignorar completamente seu aspecto subjetivo.

Dennett e seus partidários, no entanto, respondem que o aspecto subjetivo comumente utilizado não existe e não possui caráter científico, e que a sua redefinição é a única descrição coerente da consciência.

John Searle<sup>50</sup> ressalta que Dennett, ao insistir que discutir a subjetividade é um absurdo, pois esta não é científica, e a objetividade pressupõe a ciência, está fazendo um erro de categoria. Searle argumenta que o objetivo da ciência é estabelecer e validar afirmações que são epistemicamente objetivas (cuja verdade pode ser descoberta e avaliada por qualquer interessado), mas não são, necessariamente, ontologicamente objetivas. Chama qualquer juízo de valor subjetivo epistemicamente. Diz que, na visão de Dennett, não há consciência, além de recursos computacionais, porque a consciência para ele é apenas um simples efeito da máquina virtual de von Neumann, implementada numa arquitetura paralela e, portanto, implica que os chamados estados conscientes são ilusórios. Mas

---

<sup>50</sup> Searle (1997).

Searle<sup>51</sup> salienta: “[...] para onde a consciência estiver direcionada, a existência da aparência é a realidade”. (SEARLE, 1997, p. 112).

O papel do processamento da informação na consciência tem sido criticado por John Searle desde seu argumento do quarto chinês<sup>52</sup>. Afirma que não consegue encontrar nada que pudesse ser reconhecido como experiência consciente em um sistema que se baseia exclusivamente em movimentos de coisas de lugar para lugar. Dennett considera esse argumento como enganoso, uma vez que a consciência não pode ser encontrada em uma parte específica do sistema, mas nas ações do todo, um fenômeno emergente.

A questão de a consciência ser logicamente superveniente ao mundo físico parece fracassar, possibilitando imaginar que nenhuma explicação reducionista pode ser completa. Dada qualquer consideração que os processos físicos subjacentes englobam consciência, sempre haverá outra questão: por que esses processos são acompanhados pela experiência consciente? Para a maioria dos outros fenômenos do mundo, que logicamente sobrevivem ao mundo físico, tal pergunta pode ser respondida diretamente: os fatos físicos sobre esses processos contemplam a existência dos fenômenos. Para um fenômeno como a vida, por exemplo, os fatos físicos implicam que certas funções serão realizadas, e o desempenho delas é o suficiente para explicar a vida. Mas nenhuma resposta é suficiente, por enquanto, para a consciência.

## 2.6. Direções para uma Teoria Não Reducionista

Os argumentos a favor do materialismo possuem também algum fundamento. A primeira razão para preferir o materialismo é a simplicidade. Cabe evocar a bem conhecida Navalha de Occam<sup>53</sup>, que postula que não se deve multiplicar entidades

---

<sup>51</sup> Searle (1997).

<sup>52</sup> Ver seção 4.2.3.

<sup>53</sup> Navalha de Occam ou Ockham é às vezes expressa em latim como *lex parsimoniae* (a lei da parcimônia, economia ou concisão). É um princípio que recomenda que, dentre as hipóteses concorrentes, selecionando a que faz o menor número de novos pressupostos normalmente fornece o correto e que a explicação mais simples será a mais plausível até que uma evidência em contrário apareça para refutá-la.



sem necessidade. Mas, o materialismo ainda não pode explicar os fenômenos que precisam ser explicados. Talvez, um bom caminho, segundo Chalmers (1996), seja o sacrifício de uma visão de mundo fisicalista em direção a uma nova visão que possa abarcar os fenômenos subjetivos da mente humana.

Dada a natureza do nosso acesso aos fenômenos externos, devemos esperar uma explicação materialista apropriada de tais fenômenos. Nosso conhecimento desses fenômenos é fisicamente mediado pela luz, pelo som e por outros meios de percepção. Em virtude do fechamento causal do mundo físico, devemos esperar que os fenômenos que observamos por esses meios sejam logicamente supervenientes ao físico, caso contrário nunca iríamos saber sobre eles. Mas o nosso acesso epistêmico à experiência consciente é de um tipo inteiramente diferente. A consciência está no próprio centro de nosso universo epistêmico, e nosso acesso a ela não é mediado perceptivamente. A razão para se esperar uma explicação materialista dos fenômenos externos, portanto, se desfaz no caso de consciência.

Chalmers<sup>54</sup> advoga que para a construção de uma teoria consistente da consciência, há necessidade de posicionar a questão da experiência como pedra fundamental em tal iniciativa. Pois ela, assim como a massa, carga, estrutura espaço-tempo são características fundamentais do mundo a nossa volta.

Como proposta, uma teoria não reducionista da experiência especificará princípios básicos, formulando como a experiência depende das características físicas do mundo. Esses princípios, chamados de psicofísicos pelo autor, não interfeririam nas leis da física, pois, como parecem, elas já formam um sistema fechado em si mesmo. Tais princípios seriam um complemento para uma nova teoria física, mais ampla.

A teoria física nos 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 de leis físicas isoladas. Os novos princípios

---

<sup>54</sup> Chalmers (1996).

forneceriam os ingredientes extras para a conexão entre os dois 'mundos' em questão.

Há um óbvio problema que ronda o desenvolvimento de uma teoria da consciência: a falta de dados objetivos. A experiência consciente não é diretamente observável em um contexto experimental, por isso não pode gerar dados sobre a relação entre ela e os processos materiais.

O autor acredita em alguns princípios básicos para que o desenvolvimento de uma teoria da consciência possa ser consistente. São eles:

a) Princípio de coerência estrutural: propõe uma coerência consistente entre experiência consciente e a estrutura cognitiva. A consciência apresenta também um caráter funcional, mas é, no entanto, intimamente ligada à experiência consciente. Onde quer que haja experiência consciente, há alguma informação correspondente no sistema cognitivo que está disponível no controle do comportamento e está disponível também para o relatório verbal, que segue uma correspondente experiência consciente. Assim, existe uma correspondência direta entre consciência e percepção. Em relação às sensações de cor, por exemplo, para cada distinção entre experiências de cor, há uma distinção correspondente no processamento. As diferentes cores fenomenais que experimentamos formam um espaço tridimensional complexo, variando em tonalidade, saturação e intensidade, entre outros atributos. As propriedades desse espaço podem ser recuperadas a partir de processamento de informações. Exames dos sistemas visuais mostram que as formas de onda da luz são discriminadas e analisadas ao longo de três eixos diferentes, e é esta informação tridimensional que é relevante para o processamento posterior. A estrutura tridimensional do espaço de cor fenomenal corresponde, portanto, diretamente à estrutura tridimensional da consciência visual.

b) Princípio de invariância organizacional: uma sugestão natural é que a consciência surja em virtude da organização funcional do cérebro. Este princípio estabelece que quaisquer dois sistemas com a mesma organização funcional, desde seus componentes mais simples, poderão ter experiências qualitativamente idênticas. Se os padrões causais da organização neural forem duplicados em silício, por

exemplo, com um *chip* de silício para cada neurónio e os mesmos padrões de interação entre eles, a experiência poderia emergir. Segundo este princípio, o que importa para o surgimento de experiência não é a composição física específica de um sistema, mas o padrão abstrato da interação causal entre seus constituintes. Nesta visão, o produto químico ou o substrato quântico do cérebro são irrelevantes para a produção da consciência. O que conta é a organização abstrata do cérebro causal, uma organização que poderá ser realizada em diversos substratos físicos. Este princípio é bastante controverso, mas pode prestar um significativo apoio no estudo e na análise de experimentos ligados aos campos da neurociência e da inteligência artificial.

c) Teoria do duplo aspecto da informação: Shannon (1948) não estava preocupado com uma noção semântica da informação. Em vez disso, concentrou-se em uma noção formal ou sintática das informações, em que a chave é o conceito de um estado selecionado a partir de um conjunto de possibilidades. O tipo mais básico de informação é o *bit*, que representa uma escolha entre duas possibilidades: um único bit (0 ou 1), selecionado a partir de um espaço de dois estados, é considerado para transportar informações. Em um caso mais complexo, uma mensagem como '0110010101', escolhida a partir de um espaço de mensagens binárias possíveis, carrega a informação de forma similar. O que é importante, para Shannon, não é qualquer interpretação desses estados, e sim a especificidade de um estado dentro de um espaço de possibilidades diferentes. Podemos formalizar essa ideia, apelando para o conceito de um espaço de informação. Um espaço de informação é um espaço abstrato que consiste em um número de estados, que podem ser chamados estados de informação, e uma estrutura básica das relações de diferença entre tais estados. O espaço mais simples é o espaço constituído de dois estados, com uma diferença primitiva entre eles. Podemos pensar nesses estados como os dois bits, 0 e 1. O fato de esses dois estados serem diferentes um do outro esgota sua natureza. Ou seja, esse espaço de informação é totalmente caracterizado por sua estrutura de desigualdade. Para Bateson (1972), a informação física é uma diferença que faz a diferença. O princípio do duplo aspecto resulta na possibilidade de que existe um isomorfismo direto entre determinados espaços de informação físicos e certos espaços fenomenais de

informação. Pode-se notar que as diferenças entre os estados fenomenais têm uma estrutura que corresponde diretamente às diferenças incorporadas aos processos físicos. Tais diferenças fazem a diferença que estabelece determinadas vias causais que implicam a disponibilidade da informação de modo global. Dessa forma, pode-se encontrar o espaço de informação abstrato incorporado no processamento de informações físicas e também na experiência consciente. Com esta ideia, parece que algum tipo de informação possui dois aspectos básicos, um aspecto físico e um aspecto fenomenal. Este princípio também possui um caráter extremamente especulativo e deixa inúmeras questões fundamentais sem resposta. Uma pergunta óbvia é se toda informação teria um aspecto fenomenal.

### 3 A FÍSICA DA MENTE?

Somos parte – corpos e mentes – de um universo que parece obedecer de forma rigorosa e com precisão acurada às mais sutis leis da física. Os corpos físicos estão sob o jugo de tais leis, mas o que dizer sobre a mente? Quando vivenciamos determinados estados mentais, parece que toda a lógica, a matemática, as leis físicas não têm o mínimo valor dentro de nós. A possibilidade de que o nosso mundo interior possa ser reduzido às leis físicas ou procedimentos computacionais parece deixar muitas pessoas inquietas, outras motivadas em encontrar a existência e a possibilidade dessas relações.

Parece natural que estados mentais possam influenciar o corpo e este, por sua vez, influenciar também tais estados. Além disso, se a mente é apenas um epifenômeno do cérebro, um ‘subproduto’ do corpo, não poderia ter nenhuma influência de volta sobre ele, que assumiria um papel preponderantemente de segundo plano, o que não parece ser o caso.

Penrose<sup>55</sup> acredita que se a mente for capaz de influenciar o corpo de maneira que este pudesse agir fora dos limites das leis da física, então isso poderia perturbar a exatidão de tais leis. Assim, é difícil aceitar a visão totalmente dualista de que a mente e o corpo obedecem a tipos de leis totalmente independentes. Mesmo que essas leis físicas, que regem a ação do corpo, permitam uma liberdade dentro da qual a mente possa afetar de forma consistente o seu comportamento, então a natureza particular dessa liberdade deve ser também um ingrediente importante dessas mesmas leis físicas. Ele faz uma conjectura bastante interessante, quando afirma: “Seja o que for que controla ou descreve a mente, deve realmente ser uma parte integrante do mesmo esquema grandioso que regula, também, todos os atributos materiais do universo”.<sup>56</sup>

Alguns afirmam que se a mente for tratada apenas como outro tipo de substância - embora diferente da matéria e satisfazendo outras classes diferentes de princípios – estaríamos simplesmente cometendo um ‘erro de categoria’. No entanto,

---

<sup>55</sup> Penrose (1994).

<sup>56</sup> Ibidem.

apenas apontando para um possível ‘erro de categoria’, isto não remove o enigma ainda existente. Há exemplos na própria física sobre conceitos que parecem, à primeira vista, um ‘erro de categoria’ quando comparados um com os outros. Intuitivamente, temos por exemplo, os conceitos de massa e energia, pois o primeiro é uma grandeza material, real enquanto o segundo representa algo mais abstrato, que possui um caráter potencial de realizar trabalho.

Pode-se buscar explicar a consciência através de uma abordagem física, o que permite relacioná-la a outros conceitos que permeiam nosso mundo objetivo.

Uma posição que particularmente assumo neste trabalho é bastante semelhante ao pensamento de Penrose<sup>57</sup>, que defende que as atividades físicas do cérebro que produzem a consciência não podem ser simuladas através de algoritmos computacionais.

### **3.1. Emergência**

Emergência, muitas vezes evocada por filósofos e cientistas (entre eles Dennett e Dawkins) como agente gerador da consciência, é um fenômeno ou também chamado de processo de formação de padrões complexos a partir de uma multiplicidade de interações de entidades simples. O conceito de emergência é normalmente associado às teorias dos Sistemas Complexos. Tais sistemas são assim conhecidos quando suas propriedades não são uma consequência direta de seus elementos constituintes vistos isoladamente. As propriedades emergentes de um sistema complexo decorrem em grande parte da relação não linear entre suas partes. Assim, parece razoável dizer que em um sistema complexo o todo é mais que a soma das partes que o constituem. Exemplos de sistemas complexos incluem sistemas sociais (cidades e redes sociais), biológicos (colônias de animais) e físicos (clima). Áreas intimamente relacionadas a sistemas complexos são a teoria do caos e sistemas multiagentes, e um embasamento teórico e filosófico para esses sistemas é encontrado no estudo da complexidade.

---

<sup>57</sup> Penrose (1989).

Entidades emergentes (propriedades ou substâncias) 'surgem' a partir de entidades mais fundamentais e ainda são 'estranhas' ou 'irreduzíveis' em relação a elas. Alguns cientistas pesquisam a consciência como sendo uma propriedade emergente do cérebro, como o caso do físico americano Barabási<sup>58</sup> e suas pesquisas na teoria de redes complexas. Parece haver um interesse crescente na emergência dentro de discussões sobre o comportamento de sistemas auto-organizáveis e debates sobre a aproximação das questões sobre causalidade mental, intencionalidade e consciência com o fisicalismo.

Completamente diferente da abordagem darwiniana, Stuart Kauffman<sup>59</sup> promove uma arrojada ideia para a explicação da vida e evolução baseada na auto-organização que emerge em sistemas termodinamicamente abertos e complexos. Nas palavras do próprio professor Pinguelli<sup>60</sup>: “De certo modo recuperamos o determinismo, não mais o do paradigma newtoniano, preditivo, mas um determinismo de base biológica, regido pelas novas leis da complexidade e da auto-organização.”

Continua citando Kauffman<sup>61</sup>: “[... as ciências da complexidade que emergem começam a sugerir que a ordem não é acidental. Leis da complexidade geram espontaneamente a maior parte da ordem do mundo natural. Só então a seleção natural entra em jogo, moldando e refinando depois.]” (KAUFFMAN, 1995).

Nesse contexto, o acaso promovido pela seleção natural sai de cena e entra o determinismo da complexidade, baseada nas leis naturais. A ordem surge de forma espontânea, esta é a base da teoria.

Kauffman<sup>62</sup> compara o determinismo newtoniano, de caráter totalmente preditivo no que concerne aos sistemas mecânicos (exceto em regime caótico), com a teoria de Darwin, que explica a evolução, mas não arrisca qualquer tipo de

---

<sup>58</sup> Barabási (2002).

<sup>59</sup> Kauffman (1995).

<sup>60</sup> Notas das aulas ministradas na disciplina 'Teoria do Conhecimento III' do HCTE.

<sup>61</sup> Kaufmann (1995).

<sup>62</sup> Kaufmann (1995).

predição. Suscita a chamada teoria do caos, em que não se podem fazer predições de longo prazo, em decorrência da extrema sensibilidade às condições iniciais.

A mecânica estatística trata de sistemas cujas propriedades macroscópicas podem ser refletidas igualmente por diversos arranjos diferentes de seus constituintes básicos, como por exemplo, as posições e velocidades das moléculas, que no gás estão em movimento randômico. Ressalta, porém, que organismos vivos não são sistemas randômicos, mas sistemas complexos heterogêneos.

Sua teoria propõe-se assim a explicar a origem da vida como uma propriedade coletiva emergente em sistemas complexos - químicos e genéticos – baseados numa intrincada rede de genes que controlam o crescimento.

Kauffman<sup>63</sup> defende uma quebra de paradigma, ao impor um determinismo biológico (não preditivo) que vai de encontro ao paradigma da incerteza da mecânica quântica que superou o mundo determinista newtoniano.

Como observa Pinguelli, Kauffman<sup>64</sup> faz uma analogia com os sistemas computacionais, onde há sistemas que são irredutíveis à algoritmos.

Segundo ele, os sistemas biológicos se enquadram nesse caso, e Pinguelli<sup>65</sup> conclui:

Há sistemas compostos de misturas de substâncias químicas diferentes que podem espontaneamente transformar-se em sistemas auto-catalíticos. Nestes são formadas moléculas em reações químicas tais que os produtos das reações catalisam as reações que os produzem. Isto pode se dar em uma intrincada cadeia de reações. Estes sistemas se sustentam a si próprios e se reproduzem. Para Kauffman a vida emergiu da diversidade molecular em um sistema auto-catalítico que ultrapassou um limiar de complexidade, como que sofrendo uma transição de fase.

Dentro dessa visão baseada na complexidade, o cérebro e, por consequência, a consciência estariam também subordinados a tal ideia. Nesse cenário, baseado

---

<sup>63</sup> Ibidem.

<sup>64</sup> Ibidem.

<sup>65</sup> Notas das aulas ministradas na disciplina 'Teoria do Conhecimento III' do HCTE.



numa simples lógica, a consciência poderia ser apenas um epifenômeno da estrutura biológica cerebral, e tal proposta deixa inúmeras questões em aberto, como as já conjecturadas no capítulo 2 deste trabalho.

### 3.2. A Teoria Quântica e o Cérebro

Em 1989, Roger Penrose publicou *A Nova Mente do Rei*, que foi seguido em 1994 por *Shadows of the Mind*. Critica em ambas as obras as perspectivas da chamada Inteligência Artificial (IA) forte. Argumentou, partindo do teorema de Gödel e outras considerações, que certos aspectos da consciência humana, como a compreensão, devem estar fora dos limites de qualquer sistema computacional, ou seja, teriam aspectos de funcionamento 'não computáveis'.

Segundo Penrose, existe um aspecto não computável necessário para o 'aparecimento' da consciência humana, que, segundo suas pesquisas, deveria estar dentro do escopo da física, pois esta poderia explicar o que a neurociência ainda não conseguiu. O ponto de partida deveria ser de onde todos os fenômenos da realidade 'surtem', em uma área onde nossas atuais teorias físicas são fundamentalmente incompletas, embora pudesse apresentar uma relevância para as escalas subatômicas de funcionamento de nossos cérebros. A única possibilidade plausível para o autor foi a incompletude da teoria quântica, uma incompletude que tanto Einstein e Schrödinger tinham reconhecido, apesar de a teoria ter sido frequentemente defendida de forma a representar o ponto mais alto da realização científica no século XX. Essa incompletude é a questão não resolvida, referida como o 'problema da medição'<sup>66</sup>. Uma maneira, proposta por Penrose, de resolvê-lo seria uma mudança da estrutura padrão da mecânica quântica mediante a introdução de uma forma objetiva de redução de estados quânticos chamada 'OR' (redução objetiva)<sup>67</sup>, que será mais bem detalhada adiante.

---

<sup>66</sup> Ver seção 3.3.

<sup>67</sup> Ver seção 3.4.

### 3.3. A Natureza do Mundo Quântico e o Problema Fundamental

Na teoria quântica, as unidades fundamentais, os *quanta*, são em alguns aspectos bem diferentes dos objetos que são encontrados no mundo macroscópico, descrito pelas equações da física clássica. O termo *quantum* está relacionado à discretização de grandezas físicas observadas em sistemas macroscópicos. A famosa fórmula de Max Planck relacionou a energia  $E$  de um oscilador harmônico no estudo da radiação emitida por um corpo à sua frequência  $\nu$ :

$$E = n \times h \times \nu$$

Onde  $h$  é a constante de Planck e  $n$  é um número inteiro.

A relação entre os níveis de energia discretos e as frequências de oscilação subjaz a dualidade onda/partícula inerente aos fenômenos quânticos. Historicamente, De Broglie<sup>68</sup> associou uma onda ao movimento de uma partícula, como o elétron, invertendo a associação feita por Einstein do fóton às ondas eletromagnéticas de luz para explicar o efeito fotoelétrico, usando a relação de Planck. Com a equação de Schrödinger, os estados de sistemas físicos microscópicos são descritos por funções de onda, cujo módulo quadrado fornece uma distribuição de probabilidades. Assim, talvez nem a palavra 'partícula' nem a palavra 'onda' transmitam de adequadamente a verdadeira natureza de uma entidade quântica básica.

O mundo das partículas subatômicas é regido por leis sensivelmente diferentes das que regulam nosso mundo clássico cotidiano. Os *quanta*, quando suficientemente isolados do ambiente, podem ser descritos como ondas. No entanto, estas não são as mesmas ondas de matéria, tais como ondas no mar. As ondas quânticas são essencialmente ondas de probabilidade, que fornecem um cenário de posições possíveis de uma partícula dentro de um campo determinado de estudo. O pico da onda indica o local com probabilidade máxima onde uma determinada partícula pode ser encontrada. Assim, um de seus incompreensíveis aspectos afirma que partículas quânticas podem existir em dois ou mais estados ou locais simultaneamente. Estes estados estão em uma superposição, constituída de

---

<sup>68</sup> Louis de Broglie (1892 - 1987), foi um físico francês ganhador do Prêmio Nobel de Física em 1929, decorrente de sua teoria sobre a dualidade onda-partícula.

‘realidades’ coexistentes (ponderadas através de números complexos), caracterizadas formalmente como uma função de onda. Quando os sistemas são objeto de medições, a característica de onda é perdida, e uma partícula é encontrada em um ponto específico. Essa transição onda-partícula é comumente referida como o colapso da função de onda.

Quando o colapso ocorre, o resultado é aleatório. Este é um ponto de discrepância em relação à mecânica clássica. Aparentemente não há relações de causa-efeito ou sistema de algoritmos que possam descrever a escolha da posição para a partícula de forma determinística.

Outra propriedade bastante estranha aos nossos ‘olhos clássicos’ é chamada de emaranhamento (também chamada de não localidade), em que componentes completamente separados de um sistema são governados por uma mesma função de onda<sup>69</sup>. Por exemplo, duas partículas separadas espacialmente entre si apresentam correlações entre suas propriedades físicas observáveis.

Superposições quânticas de estados podem ser representadas por *bits* (ou *bits* quânticos - *qubits*), que, por sua vez, podem ser interligadas entre si através da propriedade de emaranhamento em computadores quânticos. Mas uma das características desta propriedade é que em estados emaranhados não há transmissão de informação, fazendo necessária uma sinalização clássica para atingir o referido objetivo.

O motivo pelo qual os estados superpostos não são percebidos diretamente está ligado ao chamado problema da medição. Tal problema nasce do conflito entre dois procedimentos fundamentais da mecânica quântica. Um desses procedimentos, conhecido como evolução unitária ( $U$ ), é a evolução contínua determinística de um estado quântico (isto é, da função de onda de todo o sistema) no tempo de acordo com a equação fundamental de Schrödinger. O outro procedimento, conhecido como operador  $R$ , tem um caráter totalmente probabilístico e é adotado sempre que uma medida do sistema ou uma observação dele ocorre, quando o estado quântico é descontínuo.

---

<sup>69</sup> Tittel (1998).

O problema também é conhecido como paradoxo da medição, pois os aparelhos designados a tal tarefa são construídos a partir do mesmo tipo de ingredientes quânticos (elétrons, fótons, prótons, nêutrons, quarks, glúons, etc.). Assim o sistema em observação deve estar sujeito às mesmas leis quânticas, sendo estes descritos em termos do operador  $U$ , contínuo e determinista. Por outro lado, como pode então a interação (medição) resultar num estado  $R$ , descontínuo e probabilístico?

Os físicos têm tentado encontrar respostas adequadas para o conflito. Por exemplo, no início do século XX, o físico dinamarquês Niels Bohr<sup>70</sup>, juntamente com Werner Heisenberg<sup>71</sup>, propôs a conhecida 'interpretação de Copenhague'. É a interpretação mais comum da Mecânica Quântica e foi desenvolvida pelos dois cientistas que trabalhavam juntos na referida cidade em 1927. Pode ser condensada em três ideias principais:

1. As previsões probabilísticas feitas pela mecânica quântica são irreduzíveis no sentido em que não são um mero reflexo da falta de conhecimento de hipotéticas variáveis escondidas. As probabilidades são utilizadas para completar o nosso conhecimento. A interpretação de Copenhague defende que em Mecânica Quântica, os resultados são indeterminísticos.
2. Não faz sentido especular para além daquilo que pode ser medido. A interpretação de Copenhague considera sem sentido perguntas como 'onde estava a partícula antes de a sua posição ter sido medida?'. Esta concepção reflete uma atitude completamente influenciada pela postura positivista.
3. O ato de observar provoca o 'colapso da função de onda', o que significa que, embora antes da medição o estado do sistema permitisse muitas possibilidades, apenas uma delas foi escolhida aleatoriamente pelo

---

<sup>70</sup> Niels H. D. Bohr, físico de origem dinamarquesa, ganhador do Prêmio Nobel de Física em 1922, cujos trabalhos contribuíram de forma decisiva para a compreensão da estrutura atômica e da mecânica quântica.

<sup>71</sup> Werner K. Heisenberg foi um físico teórico alemão que recebeu o Prêmio Nobel de Física em 1932 por sua contribuição na elaboração da mecânica quântica, cujas aplicações levaram à descoberta, entre outras coisas, das formas alotrópicas do hidrogênio.

processo de medição, e a função de onda modifica-se instantaneamente para refletir essa escolha.

Muitos físicos e filósofos notáveis têm criticado a Interpretação de Copenhague, com base quer no fato de não ser determinista quer no fato de propor que a realidade seja criada por um processo de observação não físico.

Desde que a função de onda não atribui uma realidade física, mas é considerada apenas para se referir ao conhecimento do observador do sistema quântico, a transição é considerada simplesmente para refletir o 'salto' no estado de consciência do mesmo.

Assim, antes da intervenção da consciência do observador, as várias 'realidades' alternativas do resultado da medição, incluindo os diferentes estados do aparelho de medida, de fato, ainda coexistiriam em superposição, de acordo com a evolução normal do sistema previsto pelo operador  $U$ . Penrose<sup>72</sup> acredita que o ponto de vista de Copenhague põe a consciência fora da ciência e não trata seriamente a natureza e o papel físico da superposição em si nem a questão de como estados superpostos macroscópicos podem evoluir.

Outra interpretação mais radical da mecânica quântica é a chamada 'hipótese dos muitos mundos' de Everett<sup>73</sup>, em que cada possibilidade em uma superposição evolui para formar seu próprio universo, resultando em uma gama infinita de mundos 'paralelos'. O fluxo de consciência do observador é considerado para 'dividir' cada possibilidade em cada um dos diferentes mundos, de modo que haja uma realidade em cada um deles, quando o observador permanece vivo e consciente. Cada instância da consciência do observador experimenta um mundo separado e independente e não é diretamente consciente de qualquer um dos outros mundos.

A decoerência ambiental é uma interpretação na qual a interação de um estado superposto com o seu ambiente desfaz os estados quânticos. Em vez de

---

<sup>72</sup> Penrose (1994).

<sup>73</sup> A chamada Interpretação dos Muitos Mundos (IMM) é uma das interpretações da Mecânica Quântica que propõe a existência de múltiplos 'universos paralelos'. Foi formulada inicialmente por Hugh Everett em meados da década de 50 para a explicação de alguns processos não determinísticos (tais como o procedimento de medição) na Mecânica Quântica.

uma função de onda ser utilizada para descrever o estado, uma entidade mais complexa é utilizada, usualmente designada como uma matriz densidade<sup>74</sup>. Segundo Penrose, a decoerência não fornece uma ontologia consistente para a realidade do mundo, em relação à matriz densidade, e fornece apenas um procedimento pragmático<sup>75</sup>. Adicionalmente, esta visão parece não abordar a questão de como o operador  $R$  possa ocorrer em sistemas isolados, nem a natureza de isolamento, em que um 'ambiente' externo não estaria envolvido. Também não descreve que parte de um sistema deve ser considerado como o 'ambiente', e não prevê qualquer limite parcial deste que possa permanecer suscetível à superposição quântica.

### 3.4. A Proposta de Penrose

A teoria de Penrose<sup>76</sup> aponta uma base física para o mecanismo da consciência, onde processos não computáveis ocorrem no interior do cérebro, pois o pensamento humano parece não poder ser simplesmente reduzido a um algoritmo, baseado nas lógicas estabelecidas e, portanto, não deve se enquadrar nos procedimentos descritos pela máquina de Turing<sup>77</sup>, na qual os computadores atuais são baseados.

Um dos aspectos polêmicos da teoria proposta por Penrose<sup>78</sup> é a explicação para a consciência a partir da Mecânica Quântica e da reformulação desta na fronteira com a física clássica. Propõe uma mudança na estrutura da Mecânica Quântica, buscando uni-la de forma adequada à Relatividade Geral, isto é, na ainda aparentemente distante *teoria quântica da gravitação*. Seus críticos dizem que os efeitos de qualquer gravidade, seja ela clássica ou quântica, não teria papel decisivo nos processos cerebrais, pois estes, em seus níveis químicos e elétricos, atuam nas

---

<sup>74</sup> Na Mecânica Quântica, a matriz densidade, descreve o estado estatístico de um sistema quântico. Apesar de algumas divergências, seu formalismo foi introduzido por John von Neumann em 1927.

<sup>75</sup> Penrose (1994).

<sup>76</sup> Ibidem.

<sup>77</sup> Vide seção 4.1.

<sup>78</sup> Penrose (1994).

escalas de  $10^{-12}$ m, enquanto a gravidade quântica atua em níveis muito mais elementares, no chamado comprimento de Planck<sup>79</sup>, escala esta muito menor que as partículas subatômicas. Penrose<sup>80</sup> preocupa-se em como a teoria espaço-temporal de Einstein pode ter alguma influência sobre as estruturas propostas pela Mecânica Quântica. Ressalta a incompatibilidade entre os dois procedimentos básicos *U* e *R* presentes no mundo quântico.

Tal incompatibilidade, no modelo atual da Mecânica Quântica e suas interpretações, não poderia ser enquadrada adequadamente ao modelo de consciência proposto, pois a componente *R* não deveria ‘emergir’, uma vez que não existe uma intervenção de um ‘observador’. Na teoria apresentada, segundo ele, as componentes *U* e *R* são consideradas aproximações diferentes de um procedimento único, muito mais abrangente e exato. Essa abordagem penrosiana parece apontar para uma teoria monista, cujo elemento unificador tem um caráter essencialmente não algorítmico, não permitindo espaço para uma computabilidade do futuro a partir do presente, embora pudesse ser determinado por este. Importante ressaltar aqui que Penrose, reconhecidamente, se declara um platônico.

A Mecânica Quântica, que descreve o nível fundamental da matéria e energia e por isso pode ser, apesar de sua complexidade, um bom caminho na procura pelas respostas que faltam sobre como a consciência ‘surge’ no mundo. E um dos trabalhos mais amplamente difundidos nesta área, são as propostas de Penrose<sup>81</sup> juntamente com o médico anestesista americano e Stuart Hameroff<sup>82</sup>.

Penrose<sup>83</sup> propõe que as ideias existentes sobre o colapso da função de onda podem apenas se aplicar às situações em que os *quanta* sejam objetos de medição. Considerou o caso dos *quanta* que não são objetos de medições ou interações, mas

---

<sup>79</sup> É uma escala de distância ( $10^{-35}$ m =  $\sqrt{\hbar G c^{-3}}$ ) na qual as chamadas ‘flutuações quânticas’ (baseadas no princípio da incerteza de Heisenberg) que ocorrem na própria métrica do espaço-tempo seriam tão grandes que a ideia normal de um espaço-tempo contínuo deixa de ser aplicável.

<sup>80</sup> Penrose (1994).

<sup>81</sup> Penrose e Hameroff (1996).

<sup>82</sup> Ibidem.

<sup>83</sup> Penrose (1994).

permanecem isolados, e propôs que eles possam estar sujeitos a uma forma diferente de colapso da função de onda. Baseia-se em conjecturas possíveis, mas controversas, nas relações entre a Mecânica Quântica e a Teoria Geral da Relatividade de Einstein e em suas próprias conclusões sobre a estrutura do espaço-tempo. A Relatividade Geral afirma que o espaço-tempo é curvado por objetos massivos. O autor, na tentativa de conciliar as duas teorias, sugeriu que, em escalas muito pequenas, a estrutura deste espaço-tempo curvo não é contínua, mas se constitui através de uma estrutura discreta, como uma rede.

Postula que cada superposição quântica tem a sua própria porção de curvatura do espaço-tempo. Segundo sua teoria, esses *bits* diferentes de curvatura do espaço-tempo são separados um do outro, e constituem uma espécie de ‘bolhas’ no tecido espaço-temporal. Seu limite seria a minúscula escala de Planck ( $10^{-35}$  m). Acima desse tamanho, sugere que o espaço-tempo pode ser visto como contínuo e que a gravidade começa a exercer sua força na ‘bolha’. Isto é sugerido para a estrutura se tornar instável acima da escala de Planck<sup>84</sup> e ao colapso da função de onda de modo a escolher apenas uma das localizações possíveis para a partícula. Penrose<sup>85</sup> chama isso de redução objetiva do evento (OR).

Uma característica importante da redução objetiva é que o tempo de colapso é uma função da massa/energia do objeto em foco. Assim, quanto maior a superposição, mais rápido ele passará por OR e vice-versa. Superposições minúsculas, por exemplo, um elétron isolado exigiria 10 milhões de anos para alcançar um limiar. Um objeto macroscópico isolado atinge o limiar em apenas  $10^{-37}$  segundos. No entanto, segundo ele, objetos em algum lugar entre a escala de um elétron e a escala macroscópica podem entrar em colapso dentro de uma escala de tempo que seja relevante para o processamento neural.

---

<sup>84</sup> Na área da Física de Partículas e na Cosmologia, a energia na escala de Planck corresponde à aproximadamente  $1,22 \times 10^{19}$  GeV (que corresponde à equivalência massa-energia para a massa de Planck  $2,17645 \times 10^{-8}$  kg) em que os efeitos quânticos da gravidade se tornam relevantes.

<sup>85</sup> Penrose (1994).



Lança mão do fenômeno da coerência<sup>86</sup>, em que muitas partículas ocupam o mesmo estado quântico, como nos raios *laser* e na supercondutividade. O sistema entra em um regime cooperativo entre seus componentes e funciona como um único bloco. Nesse contexto, ocorreria no cérebro um fenômeno semelhante de coerência, similar ao descrito anteriormente. Existem algumas contingências, como por exemplo, a supercondutividade só ocorre a temperaturas extremamente baixas, diferentemente das condições do cérebro. Faz-se necessário buscar indícios nos resultados experimentais para tais conjecturas.

### 3.4.1. Microtúbulos e Computadores

Ná década de 90, durante suas pesquisas, Hameroff (Hameroff e Watt, 1982; Hameroff, 1987) constatou o comportamento aparentemente diferenciado realizado pelas estruturas chamadas de microtúbulos, constituídos por polímeros de proteínas no interior do cérebro. Essas estruturas formam o citoesqueleto das células neuronais. São responsáveis por uma precisa separação de cromossomos na divisão celular, e a regulação das sinapses nos neurônios do cérebro (*ver Figura 1*). Essas constatações, de funções aparentemente inteligentes e sua estrutura, por parte dos pesquisadores constituíram a hipótese de que os microtúbulos poderiam funcionar como algum tipo de computador biomolecular.

Microtúbulos são estruturas formadas por tubulinas, estas constituídas por polímeros de proteínas e, em sua morfologia, assemelham-se a um grão de feijão. Cada tubulina é composta por um dímero<sup>87</sup>. Cada dímero da tubulina é composto por um monômero alfa e beta. Treze cadeias lineares de tubulinas (protofilamentos<sup>88</sup>) se alinham lado a lado para formar os cilindros 'ocos' dos microtúbulos (~25 nanômetros de diâmetro) com dois tipos de padrão reticulado. Sob a forma reticulada (grades), o microtúbulo apresenta múltiplos padrões sinuosos que se cruzam em protofilamentos em intervalos específicos correspondentes à sequência

---

<sup>86</sup> Coerência é uma propriedade de ondas que possibilitam interferência estacionária (ou seja, temporal e espacialmente constante). Geralmente, a coerência descreve todas as propriedades de correlação entre as quantidades físicas de uma onda.

<sup>87</sup> Dímero (formado de duas partes) é uma molécula composta por duas unidades similares ou monômeros unidos. É um caso especial de polímero.

<sup>88</sup> Filamento formado pela agregação de tubulina na forma de microtúbulos.

de Fibonacci<sup>89</sup>, encontrados amplamente na natureza, e possui uma simetria helicoidal.

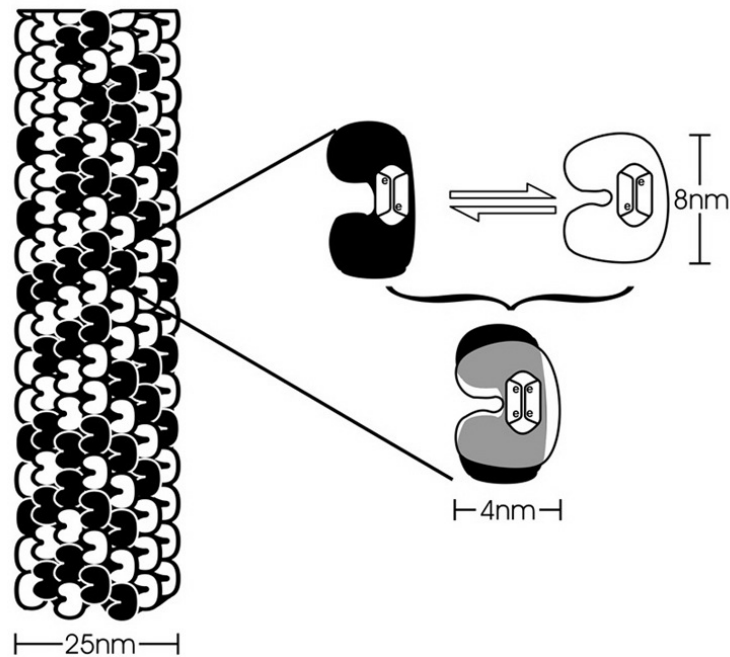


Figura 1. Penrose-Hameroff, 2011 (p. 05). Esquerda: Vista parcial de um microtúbulo composto pelas tubulinas, e estas em dímeros de proteínas. Direita, acima: Tubulina inserida numa estrutura em rede nos microtúbulos onde podem alternar seus estados (*bits*), são acopladas em nuvens de elétrons com a atuação das forças London na bolsa hidrofóbica de proteína interna. Direita, abaixo: De acordo com a OOR, cada tubulina pode estar em uma superposição quântica (*quantum bit*, ou '*qubit*') de ambos os estados.

Junto com outras estruturas constituintes do citoesqueleto das células, os microtúbulos estabelecem a forma das células e suas funções. Vários tipos de microtúbulos de proteínas associadas ('MAPs') se ligam em locais específicos, formando 'pontes' para outros microtúbulos, e definem a arquitetura das células, como vigas em um edifício. Proteínas movem-se rapidamente ao longo dos microtúbulos transportando moléculas.

Microtúbulos também se estabelecem lado a lado, em duplas ou trios. Nove duplas ou trios alinham-se para formar estruturas cilíndricas maiores chamadas

<sup>89</sup> A sequência de Fibonacci é uma sucessão de números naturais, na qual os primeiros dois termos são 0 e 1, e cada termo subsequente corresponde à soma dos dois precedentes. A sequência tem o nome do matemático pisano do século XIII, Leonardo de Pisa, conhecido como Leonardo Fibonacci, e os termos da sequência são chamados números de Fibonacci.

centríolos, organelas responsáveis pela locomoção e divisão celular. Individualmente ou em conjuntos maiores, os microtúbulos são responsáveis por movimentos celulares e intracelulares que requerem uma organização espaço-temporal inteligente. Microtúbulos têm uma estrutura reticulada, como uma grade, que apresenta semelhanças com os sistemas computacionais baseados em autômatos celulares<sup>90</sup>.

A hipótese de que os microtúbulos poderiam processar informações foi levantada por Sherrington (1957) e Atema (1973). Juntamente com alguns físicos, Hameroff desenvolveu o modelo de microtúbulos como estruturas de processamento de informação, similares a autômatos celulares, que são dispositivos computacionais auto-organizáveis. Autômatos celulares são sistemas computacionais em que as unidades fundamentais, ou 'células', em uma rede (grade reticulada) podem assumir estados específicos, por exemplo, 1 ou 0, em um determinado momento (Wolfram, 2002). Cada célula interage com as células vizinhas em passos de tempo discretos e sincronizados. O estado de cada célula num intervalo temporal é determinado pelo seu estado e os estados das células vizinhas no intervalo de tempo anterior, de acordo com as regras que regem as interações. Assim, utilizando essa estrutura, os microtúbulos poderiam realizar processos computacionais e gerar padrões complexos.

As células constituintes de autômatos celulares implicam unidades fundamentais. Mas as células biológicas não são necessariamente simples. Autômatos moleculares (no caso, os microtúbulos) são autômatos celulares em que as unidades fundamentais, *bits* ou células, são estados das moléculas, que são muito menores do que as células biológicas. Desse modo, a dinâmica interativa baseada na estrutura reticulada (estrutura em grade) torna-se necessária.

---

<sup>90</sup> Autômato celular é um modelo discreto estudado na teoria da computabilidade, matemática, e biologia teórica. Consiste de uma grade infinita e regular de células, cada uma podendo estar em um número finito de estados, que variam de acordo com regras determinísticas. Teoricamente, a grade pode apresentar qualquer número finito de dimensões.

Os estados discretos das tubulinas foram sugeridos para atuar como *bits*, alternando entre estados e interagindo (via acoplamento dipolo-dipolo<sup>91</sup>) com estados das tubulinas vizinhas, compondo uma computação baseada em 'autômatos moleculares' (Hameroff e Watt, 1982; Rasmussen, 1990; Tuszynski, 1995). O mecanismo de *bits* alternando o nível de cada tubulina foi proposto como dependente das forças de Van der Waals<sup>92</sup>, especificamente de sua componente London nas 'bolsas hidrofóbicas'<sup>93</sup> dentro de cada tubulina. As forças London nas bolsas hidrofóbicas de várias proteínas neuronais são os mecanismos pelos quais os gases anestésicos seletivamente 'desativam' a consciência (Franks e Lieb, 1984).

Para sincronizar os passos de tempo discretos nos autômatos (microtúbulos), considera-se que as tubulinas, no interior destes últimos, oscilam de forma síncrona, como proposto por Fröhlich para a coerência biológica. O biofísico Herbert Fröhlich (1968; 1970; 1975) havia sugerido que dipolos biomoleculares, limitados em uma geometria comum e num certo campo de tensão, deveriam oscilar de forma coerente, ocasionando um acoplamento vibratório. Propôs que a rede de dipolos biomoleculares poderia converter a energia do ambiente de coerência, sincronizados, por exemplo, na faixa de frequência de GHz( $10^{-9}$ ). A coerência proposta por Fröhlich pode ter um caráter quântico (por exemplo, a condensação de Bose-Einstein<sup>94</sup>, mas que é prevista apenas em temperaturas extremamente baixas, próxima do zero absoluto) ou clássico (Reimers et al, 2009).

---

<sup>91</sup> A interação química dipolo-dipolo ou dipolo permanente é uma das propriedades das forças de Van der Waals. É basicamente a força de atração que ocorre entre duas moléculas polares, ligando-as pelos seus respectivos polos, ou seja, o polo positivo de uma molécula se liga ao polo negativo da outra molécula

<sup>92</sup> São forças intermoleculares. Constituídas por três componentes, cada uma correspondendo a um tipo diferente de interação: Interações de Keeson, interações de Debye e interações de London. Numa molécula apolar, no instante em que a sua nuvem eletrônica estiver mais deslocada para um dos polos da molécula, pode-se dizer que se formou um dipolo instantâneo que gera uma pequena força intermolecular de atração. Ou seja, por um pequeno período aparecem dois polos na molécula.

<sup>93</sup> Bolsas hidrofóbicas de proteínas possuem papel importante em muitos processos que ocorrem no meio 'aquoso' cerebral, entre eles o enovelamento de proteínas, formação de membranas e complexos macromoleculares e agregação de certas proteínas em certas regiões do cérebro.

<sup>94</sup> Condensado de Bose-Einstein é uma fase da matéria formada por bósons a uma temperatura muito próxima do zero absoluto. Nessas condições, uma grande fração de átomos atinge o mais baixo estado quântico, e os efeitos quânticos podem ser observados à escala macroscópica.

Nos últimos anos, excitações coerentes foram encontradas em células vivas provenientes de microtúbulos em 8MHz (Pokorny, 2001; 2004). Bandyopadhyay (2011) encontrou uma série de picos de coerência em microtúbulos variando de 12KHz à 8MHz. Assim, os microtúbulos dentro dos neurônios poderiam fornecer outro nível de processamento de informações no cérebro.

A alta capacidade de computação baseada em microtúbulos dentro dos neurônios do cérebro poderia ser responsável pela organização sináptica, aprendizagem e memória, e talvez atuar como substrato para a consciência. Mas tal capacidade de processamento parece ainda longe de responder as questões sobre o surgimento da consciência.

### **3.4.2. O Modelo de Redução Objetiva Orquestrada (OOR)**

Como já mencionado, Penrose propõe que sua teoria seja baseada na redução objetiva (OR), que poderia representar um ingrediente fundamental na questão da incompletude da Mecânica Quântica conhecida pelo problema da medição. Seu modelo OR baseia-se no critério Diósi-Penrose<sup>95</sup>.

A redução objetiva orquestrada, doravante chamada de OOR, propõe uma extensão da mecânica quântica, que deveria formar uma ligação entre os níveis clássico e quântico, tal como a proposta quântico-gravitacional busca. A OOR está em contraste com as outras interpretações convencionais da mecânica quântica, tais como: 'decoerência ambiental', 'observação por um observador consciente', 'escolha entre mundos alternativos'. A interpretação do termo 'gravidade quântica' na teoria proposta por Penrose-Hameroff<sup>96</sup> parece sofrer variações em relação aos padrões vigentes, embora tais padrões ainda sejam objeto de muitos estudos e poucos progressos, razão pela qual ainda não são aceitos por muitos pesquisadores.

O esquema de OR soa como uma alternativa ao operador  $R$ , pois defende este, como um fenômeno físico real, não faz parte do formalismo do operador

---

<sup>95</sup> O critério Diósi-Penrose defende a existência de um limiar físico, o colapso de estados quânticos induzidos pela gravidade, que forneceria uma janela temporal plausível para a existência de estados superpostos. A proposta é também conhecida como um dos esquemas, da ainda incipiente 'gravidade quântica'.

<sup>96</sup> Penrose-Hameroff (1996).

unitário  $U$ . Assim, a OR não deve surgir apenas como uma conveniência resultante da decoerência ambiental, por exemplo.

A OR é considerada como uma das consequências da fusão dos princípios da Relatividade Geral e da Mecânica Quântica, especificamente ao modelo do operador unitário  $U$ , relacionado à evolução temporal e determinística de um sistema.

No modelo OR, qualquer medida quântica – através da qual um campo de probabilidades superpostas produzido de acordo com o formalismo  $U$ , torna-se uma única ocorrência no campo real – é um fenômeno físico real e que é tido como resultado do deslocamento de massa entre as alternativas superpostas, sendo suficiente, em termos gravitacionais, para a superposição se tornar instável.

No esquema DP (Diósi-Penrose) introduzido no modelo OR, a superposição reduz-se a uma das alternativas em uma escala de tempo  $\tau$ , que pode ser estimada de acordo com a fórmula:

$$\tau \approx \hbar / E_G$$

Onde  $\hbar (= h/2\pi)$  é a forma de Dirac da constante de Planck ( $h$ ) e  $E_G$  é a autoenergia gravitacional proveniente da diferença entre duas distribuições de massa da superposição de estados.

De acordo com o modelo para a consciência OOR, a redução objetiva não é um processo totalmente aleatório como estabelecido pela teoria padrão, mas atua segundo critérios físicos não computacionais. A ideia é que a consciência seja associada ao processo gravitacional, mas ocorre de forma significativa apenas quando as alternativas sejam parte de uma estrutura altamente organizada, de modo que tais ocorrências de OR acontecem de maneira extremamente orquestrada. Somente depois é que um evento reconhecidamente consciente ocorre. Logo, há margem para considerar que qualquer ocorrência individual de OR possa ser um elemento de protoconsciência.

O processo de OR ocorre quando superposições quânticas entre diferentes e sutilíssimas estruturas espaço-tempo se estabelecem, diferindo uma da outra por uma medida de espaço-tempo integrada, a qual se contrapõe com a escala

fundamental de Planck. Como a escala de Planck é uma entidade quadridimensional que envolve tempo e espaço, tem-se que a medida do tempo seria particularmente ínfima quando o espaço considerado é relativamente grande. De acordo com a proposta, o contrário também é verdadeiro, pois em diminutas porções do espaço resulta em escalas de tempo bastante significativas, como uma fração significativa de um segundo.

O teoria de Penrose considera que os elementos essenciais para a protoconsciência estão intimamente associados às estruturas básicas da geometria espaço-temporal de Planck, em que as escalas são extremamente ínfimas, da ordem de  $10^{-35}$ m e  $10^{-43}$ s. Estas escalas referem-se apenas às diferenças normalmente diminutas no tecido espaço-tempo entre os diferentes estados em superposição e o modelo OR tem lugar quando tais diferenças atingem o nível de Planck. As forças gravitacionais nesses níveis de realidade são extremamente fracas em comparação com as forças das reações químicas e interações elétricas presentes nas estruturas biológicas. Assim, o quantum  $E_G$  de energia é muito menor do que qualquer energia que possa surgir diretamente de processos biológicos. No modelo, o  $E_G$  não é considerado como uma força concorrente de nenhuma das forças biológicas mencionadas, uma vez que desempenha um papel completamente diferente, o que atribui um aspecto de incerteza necessária e permite, conseqüentemente, uma escolha ser realizada entre as geometrias espaço-tempo sutilmente separadas. Assim, torna-se o ingrediente-chave da computação do tempo de redução  $\tau$ . No entanto, a extrema fraqueza da gravidade aponta que deve haver uma quantidade considerável de material envolvido no deslocamento de massa coerente entre as estruturas sobrepostas a fim de que  $\tau$  possa ter o papel necessário nos relevantes processos de OR no cérebro (Figura 2).

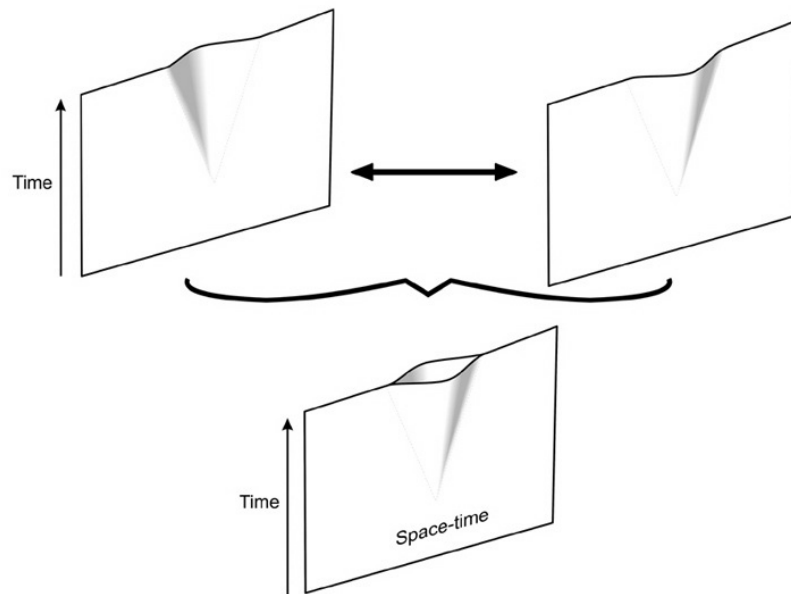


Figura 2. Penrose, 1994 (p. 338). Parte superior. Estrutura quadridimensional do espaço-tempo, aqui condensada como uma 'folha' bidimensional, onde a curvatura desta representa a presença de massa, de acordo com a teoria da Relatividade Geral. Inferior: Representação de uma 'bifurcação' do espaço-tempo na união de duas estruturas discretas espaço-temporais. Assim, uma superposição quântica de localizações alternativas pode ser representada como uma separação na geometria do espaço-tempo fundamental.

Nessas estruturas superpostas também deve ocorrer o processamento de informações e regulação da fisiologia neuronal. De acordo com a OOR, os microtúbulos são centrais para tais estruturas, e alguma forma de computação quântica biológica nestes teria que ter evoluído para fornecer uma conexão sutil, porém direta para escala de Planck, levando eventualmente a momentos discretos da experiência consciente real.

O grau de separação entre as estruturas ('folhas') do espaço-tempo é matematicamente descrita em termos de certas medidas no espaço de quatro dimensões (Penrose, 1993). A separação é, como já mencionado acima, uma separação espaço-tempo e não apenas um espaço. Assim, o tempo de separação contribui, bem como o deslocamento espacial. Grosseiramente falando, é o produto da separação temporal  $T$  com a separação espacial  $S$  que mede o referido grau total do processo, e OR ocorre quando esta separação total atinge um nível crítico. Para  $S$  pequeno, o tempo de vida  $\tau \approx T$  do estado superposto será grande; por outro lado, se  $S$  for grande, então  $\tau$  será pequeno.



Para estimar  $S$ , calcula-se a  $E_G$  gravitacional da diferença entre as distribuições de massa dos dois estados superpostos. A variável  $S$  é então dada por:

$$S \approx E_G$$

e  $T \approx \tau$ , portanto:

$$\tau \approx \hbar/E_G, \text{ e } E_G \approx \hbar/\tau.$$

Assim, a expectativa do critério DP (Diósi-Penrose) é que a OR ocorra com a resolução de uma geometria do espaço-tempo determinada a partir da superposição anterior quando, em média,  $\tau \approx \hbar/E_G$ . Além disso, segundo a OOR, isto é acompanhado por um elemento de protoconsciência.

Apesar de controversa, a ideia é que a decoerência ambiental não desempenhe papel na redução de estado. A proposta é que a redução do estado simplesmente aconteça de forma espontânea, de acordo com este critério. Em muitas situações físicas reais, haveria muito material no ambiente que seria envolvido com o estado quântico, que poderia tornar-se o principal responsável pelo deslocamento de massa - portanto, maior contribuição para  $E_G$  - assim a redução ocorreria no ambiente e não no sistema considerado. Desde que o ambiente seja quanticamente entrelaçado com o sistema, o estado de redução no ambiente irá efetuar uma redução simultânea no sistema.

Isso poderia encurtar significativamente o tempo para a redução de estado. Seria também a introdução de um elemento aleatório incontrolável no resultado da redução, de modo que qualquer elemento não aleatório (embora não computável, de acordo com a OOR) que influencia a escolha particular de estado é determinado a partir da superposição, onde esta poderia ser completamente mascarada por essa aleatoriedade. Nestas circunstâncias, a OR seria indistinguível do operador  $R$  da Mecânica Quântica convencional. Se os efeitos não computáveis sugeridos dessa proposta devem ser postos à prova, se  $E_G$  for capaz de evoluir e ser orquestrada por momentos conscientes, um isolamento significativo do meio ambiente torna-se necessário.

Atualmente, nenhum experimento foi refinado o bastante para determinar se esse modelo de OR está realmente presente na Natureza, mas, de acordo com Penrose-Hameroff<sup>97</sup>, o teste experimental do sistema é bastante próximo ao limite do que pode ser conseguido com a tecnologia atual (Marshall et al, 2003). Segundo eles, deve-se começar a ver os efeitos dessa proposta se um objeto pequeno, como um cubo de 10 *micron* ( $10^{-6}$ m) de material cristalino, puder ser mantido em uma superposição, diferindo em cerca do diâmetro de um núcleo atômico, por alguns segundos ou talvez minutos.

Para a OOR ser viável operacionalmente no cérebro, são necessárias superposições coerentes de quantidades suficientes de material, sem interferência do entrelaçamento ambiental, que reduz em conformidade com a proposta de OR, em uma escala de tempo condizente para uma experiência consciente ter lugar. Para o tipo comum de experiência do cotidiano, isso pode ser algo em torno de  $\tau = 10^{-1}$ s (10 Hz), que concorda com os possíveis correlatos neurais da consciência, tais como as frequências presentes na eletroencefalografia (EEG).

Penrose (1989, 1994) sugere que os processos de computação quântica ocorrem no cérebro delimitados pela OR. Em computadores quânticos (Benioff 1982, Deutsch 1985, Feynman 1986), a informação é representada não apenas como *bits* 1 ou 0, mas também como superposição quântica de ambos 1 e 0 juntos (*bits* quânticos, ou *qubits*), onde a superposição de grande escala de *qubits* também estaria presente.

Os *qubits* interagem seguindo a equação de Schrödinger, potencialmente permitem o processamento paralelo complexo e altamente eficiente da informação. Como previsto, no momento em que uma medição é realizada, causando a redução do estado (com alguma aleatoriedade introduzida), os *qubits* 'colapsam' para *bits* clássicos e geram estados definidos como saídas do sistema.

Segundo o modelo proposto, alguma forma de computação quântica poderia atuar no cérebro. Esse processo seria regido pela equação de Schrödinger sem

---

<sup>97</sup> Penrose-Hameroff (2011).

decoerência, até que um limiar para o autocolapso, em função de uma espécie de OR não computável, poderia ser alcançado.

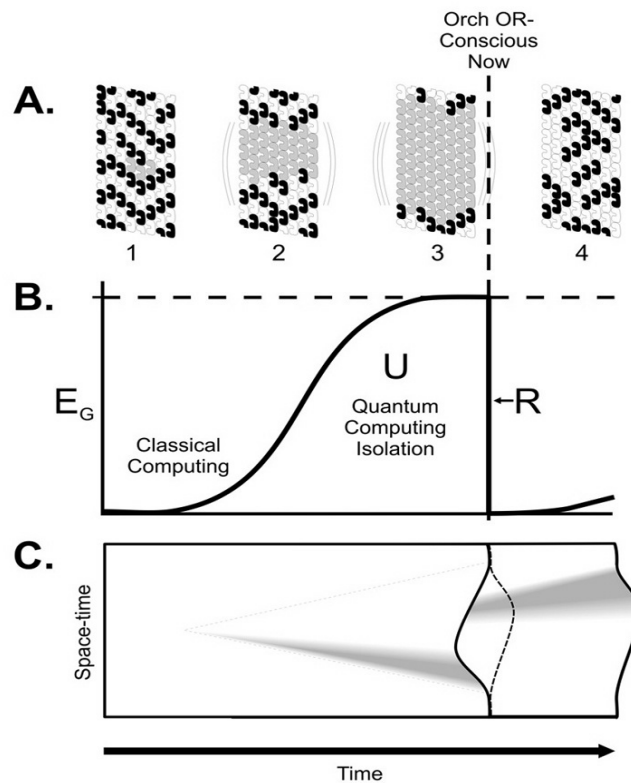


Figura 3. Penrose-Hameroff, 2011 (p. 20). Três descrições de um evento consciente baseado na OOR e regido pela equação  $EG = \hbar/\tau$ . A. Microtúbulo – tubulinas em estados quânticos (cinza) evoluem para até alcançar um limiar depois do passo 3, onde um momento consciente ocorre, os estados das tubulinas são selecionados. Para um evento real (ex. 25 msec ~40Hz), bilhões de tubulinas são necessárias; um número ínfimo é mostrado na ilustração. B. Esquema que ilustra as transições do operador  $U$  para  $R$  onde ocorrem também respectivamente as mudanças nos níveis de computação quântica para a forma clássica. C. Rede de espaço-tempo discreta em superposição alcança o limiar e uma das curvaturas ou realidade é selecionada.

Como então que  $\tau \approx \hbar/E_G$  relaciona-se com a consciência? O modelo OOR considera a consciência como uma sequência de eventos discretos em sincronia com atividades nos níveis neuronais. Em  $\tau \approx \hbar / E_G$ ,  $\tau$  é tido como o tempo para a evolução da função de onda quântica pré-consciente entre os eventos de OR, ou seja, o intervalo de tempo entre momentos conscientes, durante o qual as

superposições quânticas de estados nos microtúbulos evoluem de acordo com a equação contínua de Schrödinger, antes de alcançarem o limiar da OR em  $\tau \approx \hbar E_G$ . Tal limiar temporal é onde ocorre a redução de estado e, por consequência deste, um momento consciente ‘emerge’.

A melhor correlação temporal conhecida para a consciência, e também controversa, é a chamada sincronia gama em EEG<sup>98</sup>, de 30 a 90 Hz, muitas vezes referida ao redor de 40 Hz. Um ponto de vista possível seria aproveitar essa oscilação para representar uma sucessão de quarenta ou mais momentos conscientes por segundo ( $\tau = 25$  milissegundos). Isso seria razoavelmente consistente com algumas correntes da neurociência e também algumas ideias já expressas por alguns filósofos como Whitehead, que no final da década de vinte chamava as sucessões de momentos conscientes de ‘ocasiões de experiência’.

Na equação  $E_G \approx \hbar / \tau$ , em que  $E_G$  é derivada de interações sobre a separação da superposição em relação à distribuição de massa no sistema. Três tipos de separação de massa foram considerados na teoria, onde o efeito mais plausível pode ser calculado na separação ao nível dos núcleos atômicos, resultando numa  $E_G$  em superposição de  $2 \times 10^{10}$  tubulinas atingindo o limiar de OR adequado (~25 msec).

Cada um dos neurônios presentes no cérebro contém cerca de 108 tubulinas. Portanto, apenas algumas centenas de neurônios seriam necessários para uma sincronia gama de 25msec em um evento de OR. Se todas as tubulinas nos neurônios estiverem num estado superposto a decoerência pode ser evitada. Mas segundo alguns dados empíricos, parece mais provável que apenas uma fração de tubulinas por neurônio permaneça em estados superpostos. Estados macroscópicos de supercondutividade resultam da coerência quântica entre frações muito pequenas de seus componentes. Se um por cento dos tubulinas dentro de um determinado conjunto de neurônios forem coerentes para a janela temporal de 25 milissegundos, 20.000 neurônios seriam necessários para provocar uma OR. No entanto, alguns

---

<sup>98</sup> Eletroencefalografia (EEG) é o estudo do registro gráfico das correntes elétricas desenvolvidas no encéfalo, realizado através de eletrodos aplicados no couro cabeludo, na superfície encefálica, ou até mesmo dentro da substância encefálica.

pesquisadores apontam que a cognição e a consciência exigem algumas dezenas de milhares de neurônios (~10.000 a 100.000).

São muitas as especulações de que forma os processos quânticos computacionais poderiam evitar a decoerência nos microtúbulos. Decoerência deve ser evitada durante a evolução  $\tau (\approx \hbar / E_G)$ , de modo que os aspectos não aleatórios e não computáveis da OR passam a atuar. Argumentos contra a decoerência dentro de um ambiente 'hostil' como o cérebro parece ser realmente um grande desafio para a teoria Penrose-Hameroff.

No artigo de 1996 de Penrose e Hameroff, foi sugerido que os estados quânticos nos microtúbulos fossem protegidos por pulsos tal como o laser, através da ressonância de Fröhlich, e também apresentariam uma blindagem entre outros componentes, através do gel de actina<sup>99</sup> e fortes campos elétricos mitocondriais. Além disso, os estados quânticos na OOR são originados nas bolsas hidrofóbicas no interior das tubulinas, isoladas a partir de interações polares. Há também as potenciais ressonâncias geométricas nos microtúbulos, através das estruturas helicoidais baseadas nas proporções de Fibonacci, que são sugeridas para permitir que a computação quântica topológica pudesse evitar a decoerência indefinidamente (Hameroff, 2002) como em um supercondutor.

Pesquisas mais recentes, a partir de 2003, apontam para uma possibilidade de existência de coerência quântica em sistemas biológicos com temperaturas mais quentes. Em 2003, os pesquisadores Ouyang e Awschalom mostraram que a transferência de *spin*<sup>100</sup> através de anéis de fenil (os mesmos que nos bolsos hidrofóbicos de proteínas) é aumentada em temperaturas cada vez mais quentes. Outros estudos mostraram que coerência quântica ocorreu à temperatura ambiente em proteínas envolvidas na fotossíntese, e que as plantas usam rotineiramente coerência quântica para produzir energia química (Engel et al, 2007) entre outros

---

<sup>99</sup> Proteína que, em conjunto com a miosina e moléculas de ATP (trifosfato de adenosina – composta responsável pelo armazenamento de energia em suas ligações químicas), gera movimentos celulares e musculares. A actina polimerizada forma os microfilamentos de actina importantes na composição do citoesqueleto.

<sup>100</sup> De forma não rigorosa, no contexto da Mecânica Quântica, o termo *spin* está ligado às possíveis orientações que partículas subatômicas carregadas, tais como prótons, elétrons e alguns núcleos atômicos podem apresentar, quando imersas em um campo magnético.

estudos. Todos esses estudos ainda necessitam de um grande caminho para alcançarem resultados mais consistentes.

Várias questões ainda permanecem abertas, principalmente a não decoerência nos microtúbulos. Como já mencionado, o isolamento proposto das estruturas envolvidas na modelo parece ainda estar baseado em argumentos não suficientemente sólidos. Tal modelo propõe também que exista uma alternância entre as fases isoladas de computação quântica com fases de interação ambiental clássica, por exemplo, durante a sincronia gama, cerca de 40 vezes por segundo.

Com relação às saídas resultantes dos processos que ocorrem em nível dos microtúbulos no modelo OOR de computação quântica, os dentritos nas conexões sinápticas recebem e integram sinais de entrada durante a fase clássica. Eles tornam-se alvo de processos quânticos computacionais e evoluem até atingir o tempo limiar (dado por  $\tau \approx \hbar / E_G$ ) no qual ocorre a redução dos estados superpostos. Os estados das tubulinas escolhidos para a redução podem desencadear o disparo axonal, ajustando-o, realizando a regulação das sinapses e codificando a memória. Assim Penrose e Hameroff<sup>101</sup> acreditam que o modelo OOR pode também ter eficácia causal nas ações e comportamento, bem como proporcionar a experiência consciente e consolidação da memória.

A proposta levanta ainda a possibilidade que durante os processos intermediários da OOR, atividades neuronais não conscientes podem proceder através computação baseada no modelo clássico no interior dos microtúbulos. Adicionalmente, pode haver atividades computacionais quânticas que não atingem o limiar e, assim, também permanecem inconscientes.

Em alguns momentos, Penrose e Hameroff fazem conjecturas das mais diversas, muitas sem nenhum respaldo argumentativo consistente para as lacunas que ainda permanecem na teoria. Afirmam que, em virtude da seleção natural, algumas estruturas dos microtúbulos são relativamente primitivas e, portanto, ainda incapazes de suportar a computação quântica, o que resulta em processos que não são conscientes, devidos a tal lacuna evolutiva.

---

<sup>101</sup> Penrose-Hameroff (2011).

Segundo eles, somente numa fase evolutiva muito posterior, uma capacidade cognitiva seria uma vantagem seletiva. Isso exigiria recursos não computáveis da OOR, que vão além da computação quântica, e dependem da infraestrutura de maior escala de eficiência nos microtúbulos, capazes de operar os processos quântico-computacionais. Assim, a evolução forneceu maiores conjuntos de microtúbulos (maior  $E_G$ ) capazes de serem isolados da decoerência, permitindo pelo  $\tau \approx \hbar / E_G$ , momentos mais frequentes e mais intensos da experiência consciente.

Assumindo que as fases emaranhadas nos microtúbulos são isoladas em um neurônio específico, o estado quântico se estende através das estruturas cerebrais através do fenômeno conhecido como tunelamento<sup>102</sup>, levando os estados quânticos para outras áreas adjacentes do cérebro.

### 3.5. Críticas ao modelo de Penrose-Hameroff

O professor Pinguelli e o pesquisador Jean Faber escreveram em 2004 um artigo no qual revelam argumentos consistentes contra a possibilidade de existir no cérebro o modelo de colapso gravitacional da OOR, defendido por Penrose-Hameroff. Sugerem substituir tal ideia pelo processo de decoerência natural para produzir a consciência. Os pesquisadores mostram um erro sobre o regime de temperatura do processo de decoerência que foi considerado por Hagan, Hameroff e Tuszynski<sup>103</sup> em sua crítica ao cálculo da taxa de decoerência realizada por Tegmark<sup>104</sup>.

Um dos pontos mais frágeis na teoria proposta da OOR é a interação do sistema, no caso o cérebro, com o ambiente. Segundo a teoria da decoerência, objetos do mundo macroscópico obedecem também à mecânica quântica. A interação com o ambiente 'desfaz' os efeitos quânticos presentes nos objetos

---

<sup>102</sup> Tunelamento quântico é um fenômeno da mecânica quântica em que uma partícula tunela através de uma barreira que classicamente não conseguia superar. Esta desempenha um papel essencial em diversos fenômenos físicos, tais como a fusão nuclear que ocorre em estrelas como o Sol. Seu efeito foi previsto no início do século XX, e sua aceitação como um fenômeno físico geral veio em meados do século passado.

<sup>103</sup> Hagan, Hameroff e Tuszynski (2002).

<sup>104</sup> Tegmark (2000).

macroscópicos. Por consequência, estes acabam se comportando de acordo com as leis previstas pela física clássica e a mecânica newtoniana, em uma aproximação da teoria quântica. Na visão de Pinguelli e Faber, o ínfimo período de superposição quântica e a decoerência daí resultante constituem o mecanismo fundamental gerador da consciência. A superposição quântica coerente de estados é sustentável em nível macroscópico por um tempo diminuto. Quando o sistema físico atinge um nível crítico, a coerência (estados quânticos superpostos) é desfeita pela decoerência, por causa da interação com o meio ambiente.

Os autores acreditam que se houver uma relação satisfatória entre os efeitos da decoerência e da teoria biológica de Fröhlich<sup>105</sup>, pode-se construir um modelo quântico da mente. Ao invés de negarem a tese de Penrose-Hameroff, os autores preferem seguir outro caminho, discutindo o tempo de decoerência no cérebro calculado por Tegmark, cuja conclusão foi desconsiderar a relevância da abordagem quântica do cérebro.

Na teoria proposta por Penrose-Hameroff, o período de tempo em que a coerência é sustentada no cérebro foi estimado em aproximadamente 25 msec, baseado no tempo biológico de resposta do cérebro a estímulos externos. Esse período é muitas vezes maior do que o tempo 'padrão' de decoerência, considerado um processo extremamente rápido. Pinguelli e Faber lembram o tempo de redução pelo efeito gravitacional, calculado por Penrose, em gotas de água de  $10^{-4}$  cm e  $10^{-3}$  cm de diâmetro é de, respectivamente, 0,1 s e  $10^{-6}$  s.

Segundo os autores Pinguelli e Faber, parece que tais cálculos não estão corretos, pois o tempo típico de decoerência previsto em tais estruturas deve ser muito menor. Para sistemas com raios de  $10^{-3}$  cm e  $10^{-5}$  cm, os tempos experimentais são  $10^{-23}$  e  $10^{-9}$ s, respectivamente. Isso contraria sensivelmente o proposto por Penrose.

---

<sup>105</sup> A teoria de Fröhlich descreve a superposição quântica em macromoléculas biológicas, considerando a polarização e certas propriedades vibratórias destas moléculas. Vibrações em sistema polarizados geram campos eletromagnéticos que podem mediar interações entre as unidades do sistema. A não linearidade de seu modelo induz a uma transferência de energia entre os modos de vibração. Este processo pode fornecer uma condensação da energia e, conseqüentemente, a coerência em sistemas vivos.



Hagan, Hameroff e Tuszynski<sup>106</sup> escrevem um artigo a fim de tentar fornecer uma resposta ao trabalho de Tegmark, no qual concordavam parcialmente com este em relação às conclusões ao nível neuronal, em escala macroscópica, mas não em escala microscópica, onde as tubulinas e microtúbulos desempenham o papel fundamental na OOR. Uma razão é que o modelo Tegmark leva em consideração a interação de ondas sóliton<sup>107</sup> ao longo do microtúbulo, enquanto que a OOR considera a superposição de diferentes estados conformacionais de um dímero de tubulina. Apesar de suas críticas, Hagan, Hameroff e Tuszynski usaram o modelo de Tegmark com algumas modificações.

Na teoria da OOR há poucos argumentos relacionados a como os estados coerentes são sustentados no interior dos microtúbulos e como se mantêm isolados. Pinguelli e Faber acreditam que exista o processo como descrito por Fröhlich, agindo no cérebro para sustentar os estados coerentes. Em tal processo, a energia fornecida a um sistema com vibração não linear, equilibrado por perdas de energia em decorrência da interação com o ambiente, não é imediatamente transformada em calor e pode perturbar o equilíbrio termodinâmico através da energia de condensação. Dependendo da energia fornecida e da energia transferida para o meio ambiente, o sistema muda o regime de incoerente para um tempo limitado de coerência e daí em fases posteriores e estáveis de coerência, o que parece apontar um novo rumo para as possibilidades quânticas agindo no cérebro.

---

<sup>106</sup> Hagan, Hameroff e Tuszynski (2002).

<sup>107</sup> Sóliton é uma onda de auto-reforço unitária (um pacote de onda ou pulso) que mantém a sua forma enquanto viaja a uma velocidade constante. São causadas por um cancelamento de efeitos não lineares e dispersivos no meio. O fenômeno foi primeiramente identificado por John Scott Russell (1808-1882).

## 4 COMPUTABILIDADE E INTELIGÊNCIA ARTIFICIAL

Na sua forma padrão, a teoria da computação lida inteiramente com objetos abstratos: máquinas de Turing<sup>108</sup>, autômatos, e assim por diante. Estas são entidades habitantes do espaço matemático. Sistemas cognitivos no mundo real, por outro lado, são objetos concretos, fisicamente delimitados e interagem causalmente com outros objetos no mundo físico. Muitas vezes procuramos utilizar a teoria da computação para tirar conclusões sobre objetos concretos no mundo real. Para tanto, precisamos de uma ponte entre os domínios abstrato e concreto. Essa ponte é a noção de computabilidade: a relação entre o abstrato computacional, cujos objetos são ‘cálculos’ e instruções, e sistemas físicos que são acionados quando ‘percebem’ uma determinada ‘entrada’, ou quando um dado algoritmo ‘descreve’ um sistema físico. Em sua maioria, os modelos computacionais são implementados sintaticamente, isto é, a partir de regras e procedimentos bem definidos. Descrições computacionais são usadas para dar sentido a sistemas físicos em diversos domínios do conhecimento humano. Sempre que isto acontece, uma noção de implementação (computabilidade) está implícita ou explícita, realizando tal tarefa.

### 4.1. Gödel e a Máquina de Turing

Ainda jovem, Kurt Friedrich Gödel, matemático e lógico alemão, no seu artigo histórico *On Formally Undecidable Sentences of Principia Mathematica and Related Systems* de 1931, mostrou que nenhum sistema de regras composto de provas lógicas é suficiente para estabelecer todas as proposições verdadeiras da aritmética e, conseqüentemente, da matemática.

O teorema de Gödel tornou-se valioso na discussão da inteligência artificial. Penrose<sup>109</sup> considera o teorema de Gödel o mais importante da lógica matemática, porque mostrou que o entendimento humano não se reduz a um conjunto de regras computacionais ou algoritmos. E arrola isto contra a pretensão de se chegar a uma

---

<sup>108</sup> Vide seção 4.1.

<sup>109</sup> Penrose (1994).

genuína inteligência artificial em substituição à humana ou até mesmo da ‘humanização’ de um robô. O que Gödel mostrou, e Penrose<sup>110</sup> na discussão da mente e da inteligência artificial utiliza como uma ruptura de paradigma na lógica, é que: “Nenhum sistema formal  $F$  pode ser ao mesmo tempo consistente e completo”.

Para maior clareza, algumas definições são necessárias:

1 - Os enunciados decorrentes de um sistema formal  $F$ , como Hilbert<sup>111</sup> propôs podem ser verdadeiros ou falsos.

2 - Se um enunciado puder ser estabelecido por  $F$  que não seja verdadeiro nem falso, ele é dito indecidível.

3 - Por outro lado,  $F$  é dito consistente se nenhum enunciado é provado ao mesmo tempo ser verdadeiro e falso.

4 - Finalmente  $F$  é chamado de completo se todo enunciado matemático formulado em  $F$  sempre for ou verdadeiro ou falso, isto é, se  $F$  não contém enunciados indecidíveis.

#### 4.1.1. O Teorema de Gödel e o Problema da Parada de Turing

Ao nosso intelecto, a ideia de provar que algo não pode ser demonstrado parece ser realmente fascinante, e em algumas raras situações os pesquisadores têm oportunidade de encontrar fenômenos matemáticos dessa natureza.

Em geral, tais argumentos só ocorriam nas provas de impossibilidade de solução dos ‘problemas clássicos gregos’ (trisseção de ângulo, quadratura do círculo e duplicação do volume do cubo, usando apenas régua e compasso). Estes já tinham sido provados indemonstráveis independentemente dos teoremas de Gödel.

---

<sup>110</sup> Ibidem.

<sup>111</sup> Em 1928, Hilbert e Wilhelm Ackermann publicaram ‘*Grundzüge der Logik theoretischen*’ (*Principles of Mathematical Logic*), uma introdução à lógica de primeira ordem em que o problema da completude foi apresentado: Os axiomas de um sistema formal são suficientes para tirar qualquer declaração que seja verdade em todos os modelos do sistema?

Antes de Gödel, fazia parte de um amplo projeto de trabalho liderado por David Hilbert, conhecido como Programa de Hilbert, que buscava provar que todo problema matemático fosse solúvel. Nas próprias palavras de Hilbert, num congresso em Münster, em 1925:

Se existe um problema, ache a solução; você pode encontrá-la apenas pensando, pois não há *ignorabimus* em matemática. (CARNIELLI, 2006, p. 23).

Hilbert pretendia tornar axiomático todo o corpo do conhecimento matemático, e um de seus objetivos era provar que todo problema matemático pudesse ter uma solução, isto é, provar, por meios estritamente finitários, que a axiomática pretendida fosse consistente.

Num domingo, no dia 7 de setembro de 1930, na cidade de Königsberg (mesma cidade onde viveu Immanuel Kant) Kurt Gödel timidamente anunciou:

Pode-se, de fato, exibir sentenças verdadeiras, mas que são indemonstráveis no sistema formal da matemática.

Os famosos Teoremas da Incompletude de Gödel são:

1º. Teorema de Incompletude: Em todo sistema formal consistente  $S$ , com um mínimo de aritmética, é possível formalizar uma sentença  $U$ , tal que  $U$  possa ser interpretada intuitivamente como a afirmação de que ela própria é indemonstrável em  $S$ .

Esta sentença  $U$  é formalmente indemonstrável em  $S$  e, portanto, ela expressa um fato verdadeiro. Contrariando o que Hilbert acreditava,  $U$  constitui um legítimo *ignorabimus* na Matemática. Em segundo lugar, Gödel mostrou também que:

2º. Teorema da Incompletude: A prova da consistência para sistemas formais (envolvendo um pouco de aritmética, nas condições que Hilbert gostaria) não pode ser formalizada dentro do próprio sistema.

Um exemplo que ilustra a ideia de Gödel (no 1º. Teorema) é o seguinte: “Um conhecido político que havia mentido durante toda sua vida pública, antes de

retirar-se da carreira política, afirmou o seguinte, tentando salvar sua memória: ‘Eu sempre menti em toda minha carreira, e esse é um fato que pode ser provado’.”

Adversários: quiseram provar que ele realmente havia mentido a vida toda. No entanto, depararam-se com um problema: se inclusive essa última frase fosse mentira, ou ele não havia mentido a vida toda, ou esse fato não poderia ser provado.

Correligionários: seus comparsas, tentando argumentar que ele não havia mentido a vida toda, encontraram um problema semelhante: sabendo que o político era mentiroso, a única esperança era argumentar que a última frase fosse verdadeira, o que seria admitir que se pudesse provar que ele realmente mentiu sempre.

Um fato verdadeiro, mas indemonstrável.

Portanto o fato  $U$ :

O político sempre mentiu.

É tal que nem  $U$  e nem  $\sim U$  (onde ‘ $\sim$ ’ significa ‘negação’) podem ser provados, mas é contudo verdadeiro, porque ele sempre mentiu, inclusive ao afirmar que  $U$  poderia ser provado.

Outro exemplo bastante conhecido de problema indecidível pode ser encontrado no paradoxo do mentiroso:

O cretense diz:

Todos os cretenses são mentirosos.

Isto é uma mentira.

Não há como se decidir se é Verdadeira ou Falsa a proposição.

Se for Verdadeira, então ela é uma mentira, portanto, é Falsa.

Se for Falsa, então não é uma mentira, portanto, é Verdadeira.

Por consequência, o problema é dito indecidível.

#### **4.1.2. O Problema da Parada**

Dentro da teoria da computação, o experimento mental do problema da parada é um problema sobre decisão que pode ser assim declarado de maneira informal: Dada uma descrição de um programa e uma entrada finita, decida se o programa termina de rodar ou rodará indefinidamente, dada a referida entrada.

Alan Turing provou em 1936 que um algoritmo genérico para resolver o problema da parada, para todos pares programa-entrada possíveis, não pode existir. Dizemos que o problema da parada é indecidível nas Máquinas de Turing.

O problema da parada é um problema de decisão sobre as propriedades de programas de computadores em um determinado modelo de computação de Turing. O problema é determinar para uma dada entrada se o programa irá parar com ela. Nessa área de trabalho abstrata não há limitações de memória ou tempo necessário para a execução de um programa, pois poderão ser necessários tempo e espaço arbitrários para o programa parar. A questão é se o programa simplesmente poderá parar a partir de uma determinada entrada.

Um programa mais complexo pode ser mais difícil de analisar. O programa pode rodar por um tempo fixo e, se ele não parar, não há como saber se o programa irá parar eventualmente ou se ele irá continuar rodando indefinidamente. Dessa forma, Turing provou que não há um algoritmo que pode ser aplicado a qualquer programa arbitrário, com uma entrada, para decidir se o programa para ou não a partir de tal entrada.

#### 4.1.3. Enunciado Informal do Problema da Parada

Um problema de decisão é um conjunto de números naturais; o 'problema' é determinar se um número em particular pertence ao conjunto.

Dada uma enumeração de Gödel  $\varphi$  de uma função computável (como os números de descrição de Turing) e uma função de pareamento  $(i,x)$ , o problema da parada é o problema de decisão para o conjunto:

$$K_{\varphi}^0 := \{\langle i, x \rangle \mid \varphi_i(x) \text{ existe}\}$$

com  $\varphi_i$  a  $i$ -ésima função computável na enumeração de Gödel  $\varphi$ .

Embora o conjunto  $K$  seja recursivamente enumerável<sup>112</sup>, o problema da parada não é solúvel por uma função computável.

Existem diversas formulações equivalentes do problema da parada; qualquer conjunto, cujo grau de insolubilidade da Teoria da Computação (no inglês, *Turing degree*) seja o mesmo que o do problema da parada, pode ser considerado com tal formulação. Exemplos de tais conjuntos incluem:

$$\{i \mid \varphi_i(0) \text{ para}\}$$

$$\{i \mid \exists n(\varphi_i(n) \text{ para})\}$$

O problema da parada tem um cunho histórico, pois foi um dos primeiros a ser provado como indecidível. Esse fato aconteceu em maio de 1936, enquanto a prova de Alonzo Church sobre a indecidibilidade de um problema no cálculo lambda<sup>113</sup> já havia sido lançada em abril do mesmo ano.

Posteriormente, outros problemas também foram classificados como indecidíveis. O método típico utilizado para provar tal característica é a chamada técnica de redução. Esta procura evidenciar que se uma solução para o novo problema foi encontrada, ela poderia ser usada para decidir um problema indecidível (transformando instâncias do problema indecidível em instâncias do novo problema). Dessa forma, como se sabe de antemão que nenhum método pode decidir o problema antigo, então nenhum método pode decidir o problema novo também.

Uma consequência da indecidibilidade do problema da parada é que não pode existir um algoritmo genérico que decida se um dado enunciado sobre os números naturais é verdadeiro ou falso. A razão para isso é que a proposição que afirma que um certo algoritmo vai parar, dado uma certa entrada, pode ser

---

<sup>112</sup> Na teoria da computabilidade, um conjunto  $S$  de números naturais é chamado recursivamente enumerável se existe um algoritmo tal que o conjunto de números de entrada para qual o algoritmo para é exatamente o conjunto de números em  $S$ . Ou, de forma equivalente, existe um algoritmo que enumera os membros de  $S$ . Isto significa que sua saída é simplesmente uma lista de membros de  $S$ :  $s_1, s_2, s_3, \dots$

<sup>113</sup> Cálculo lambda, também escrito como cálculo- $\lambda$  é um sistema formal que estuda funções recursivas computáveis, relativas à teoria da computabilidade e fenômenos relacionados, como variáveis ligadas e substituição. Sua principal característica são as entidades que podem ser utilizadas como argumentos e retornadas como valores de outras funções.

convertido em um enunciado equivalente sobre os números naturais. Se nós tivéssemos um algoritmo que pudesse resolver todo enunciado sobre os números naturais, ele certamente poderia resolver tal problema; mas isso determinaria se o problema original para, o que é impossível, já que o problema da parada é indecidível.

O problema da parada demonstra que existem problemas que não podem ser resolvidos por máquinas de Turing. A tese de Church-Turing limita o que pode ser atingido por qualquer máquina que implemente métodos efetivos. Também se pergunta se qualquer um desses processos físicos desconhecidos esteja envolvido no funcionamento do cérebro humano e se humanos podem resolver o problema da parada<sup>114</sup>.

A introdução de Turing do modelo de máquina, que posteriormente ficou conhecido como Máquinas de Turing, revelou-se um modelo muito conveniente para a Teoria da Computação.

#### **4.1.4. Tese de Church-Turing**

É definida pelas palavras de Turing:

Toda função que seria naturalmente considerada computável pode ser computada por uma Máquina de Turing. (TURING, 1936)

Pela imprecisão do conceito de uma 'função que seria naturalmente considerada computável', a tese não pode ser formalmente provada, mas pode ser refutada, caso seja descoberta uma máquina teórica mais poderosa que a máquina de Turing.

Qualquer programa de computador pode ser traduzido em uma máquina de Turing, e qualquer máquina de Turing pode ser traduzida para uma linguagem de programação de cunho geral; assim, a tese é equivalente a dizer que qualquer linguagem de programação de cunho geral é suficiente para expressar qualquer algoritmo.

---

<sup>114</sup> Copeland (2004).



## 4.2. Consciência e Computação

Mesmo com os avanços no campo da neurociência, ainda há uma grande ausência de consenso relacionada aos estados do cérebro e suas relações com o fenômeno da consciência. O que parece certo é que nem todas as partes do cérebro participam de forma igualitária em sua manifestação. Para corroborar com esta asserção, o caso do cerebelo pode ser evocado. Seu funcionamento pode lembrar a de um autômato<sup>115</sup>, pois suas ações parecem ocorrer de maneira involuntária, não sendo necessária nenhuma ação da consciência para sua realização. Mesmo em atividades rotineiras, como caminhar de um ponto a outro, não temos consciência da complexa cadeia de processos neurais e motores envolvidos no movimento. O movimento reflexivo (por exemplo, ao afastar a mão do fogo) também apresenta um forte caráter inconsciente intermediado pela parte superior da medula vertebral. Parece sensato afirmar então que a consciência tenha uma maior relação com o cérebro do que com o cerebelo e a medula.

Mas o foco da discussão está nas atividades que se podem fazer de modo consciente e que não são possíveis quando um indivíduo se encontra sob um estado de inconsciência. Trata-se de uma distinção tortuosa, pois tudo o que fazemos parece nos exigir certo nível de consciência, pois muitas coisas têm de ser aprendidas primeiro, para depois serem executadas. Em casos como a formulação de julgamentos (muitos fatores inconscientes podem estar presentes na avaliação) baseados em regras não definidas antecipadamente, entendimento e compreensão de contextos no âmbito semântico, avaliação artística ainda não podem atualmente ser estabelecidas por meio computacional.

Alguns estudiosos apontam que ações, conscientes ou não, do cérebro humano sejam apenas a execução de um algoritmo muito complexo, e o objeto das ciências da mente e da IA forte seja apenas a decodificação dessas ações. Mesmo que esta proposição esteja correta, ainda fica a pergunta de como tal algoritmo surgiu, mas a resposta trivial de que a seleção natural seja a responsável por tal fato, ainda não parece razoável. Essa teoria afirma que cérebros mais adaptados (evoluídos?), isto é, que dispunham de algoritmos mais eficientes, teriam maior

---

<sup>115</sup> Modelo matemático de uma máquina de estados finitos.

probabilidade de sobreviver, resultando assim, maiores chances de gerar descendentes. Estes últimos herdariam os algoritmos mais eficientes de seus primos, e o aperfeiçoamento seguiria uma cadeia causal, muitas vezes não linear, proporcionando mudanças sensíveis em sua evolução. Esse é um painel muito resumido de uma corrente que tem, entre seus partidários, pesquisadores como Richard Dawkins<sup>116</sup>.

As ideias da corrente evolucionista, descritas resumidamente acima, contam com evidências empíricas e satisfazem muitas perguntas que careciam de respostas. Mas algumas lacunas persistem, quando admitimos que a consciência tenha evoluído através da seleção natural. Acreditar que ações das consciências dos programadores de um computador são, em si, simples algoritmos, então devemos chegar à conclusão de que tenham evoluído de forma semelhante. Mas a questão fundamental neste contexto refere-se ao problema da parada da máquina de Turing<sup>117</sup>, pois decidir se um algoritmo funciona ou não requer um processo intuitivo e não apenas de um algoritmo; obviamente dentro dos limites atuais de como o concebemos.

Continuando a utilizar a seleção natural como sendo capaz de produzir algoritmos aproximadamente válidos, pode-se concluir que um processo semelhante poderia apenas interferir nos resultados por ela produzidos, e não diretamente sobre as ideias subjacentes às suas ações. Tal processo não apresenta uma característica eficiente e nem mesmo viável. Para ilustrar tal ocorrência, Penrose<sup>118</sup> cita um exemplo bastante recorrente na informática:

É uma tarefa extremamente complexa verificar o que é realmente um algoritmo apenas examinando seu resultado. Seria fácil construir duas ações, bem diferentes, de máquinas de Turing simples, para as quais as fitas de resultados não diferissem até, por exemplo, o  $(2^{65536})^o$  lugar – e essa diferença poderia não ser nunca percebida na história do universo. Além disso, a menor ‘mutação’ de um algoritmo (por exemplo, uma modificação na especificação de uma máquina de Turing ou na sua fita de entrada) tenderia a torná-lo totalmente inútil, sendo difícil ver como aperfeiçoamentos nos algoritmos pudessem surgir de forma aleatória. Mesmo as melhorias deliberadas são difíceis sem que existam ‘significados’. Isso é confirmado

---

<sup>116</sup> Dawkins (2001).

<sup>117</sup> Ver seção 4.1.2.

<sup>118</sup> Penrose (1989).

particularmente pelas circunstâncias não raras em que um programa de computador, complicado e inadequadamente documentado, precisa ser alterado ou corrigido; e o programador original já não se encontra presente ou talvez tenha falecido. Em lugar de tentar decifrar todos os vários significados e intenções de que o programa dependia implicitamente, talvez seja mais fácil jogá-lo fora e começar tudo de novo. (PENROSE, 1989, p. 535).

Talvez o próprio processo de seleção natural, apoiado nos bilhões de anos desde o surgimento da vida, ainda contenha elementos totalmente desconhecidos que permitam uma contínua escala de aprimoramento do pretenso algoritmo (e talvez suas sub-rotinas) que produz a consciência.

#### **4.2.1. Os Níveis Computacionais e o Pensamento Consciente**

Provavelmente existam certas coisas que o cérebro realiza que não podem ser descritas em termos computacionais. Coisas como sentimentos de felicidade, amor, sensibilidade à dor, estética, vontade, entendimento etc., talvez não possam ser enquadrados apenas a imagens e procedimentos.

Em relação às conjecturas sobre as possibilidades de um computador no futuro ter uma mente como a humana, Penrose<sup>119</sup> classifica alguns níveis computacionais e os relaciona com as atividades cerebrais. Assim, identifica quatro diferentes pontos de vista extremos e possíveis para os limites da inteligência artificial. São eles:

A - Todo o pensamento reduz-se à computação e a experiência da consciência pode ser produzida através da computação apropriada.

B - A consciência decorre de processos físicos que ocorrem no cérebro e, embora elas possam ser simuladas em computador, as simulações por si só não produzem a consciência.

C - As atividades físicas do cérebro que produzem a consciência não podem ser simuladas através de algoritmos computacionais.

---

<sup>119</sup> Penrose (1994).

D - A consciência nem se reduz à computação nem às atividades físicas do cérebro e não pode ser explicada cientificamente.

O modelo *D* é o domínio do misticismo e da religiosidade. Como a tentativa é a busca de caminhos através da ciência, tal ponto deve ser completamente rejeitado em face da incompatibilidade com a metodologia científica.

O extremo oposto é a possibilidade *A*. Seus adeptos são os partidários da inteligência artificial forte (IA forte), também podendo ser identificados como funcionalistas, que parece ser, neste contexto, um tanto nebuloso. Alguns outros defensores de *A* se diferenciam na forma como interpretam o significado de 'consciência'. Alguns nem sequer permitem que haja um fenômeno como 'consciência' em tudo, enquanto outros aceitam a existência desse fenômeno, mas o consideram como apenas algum tipo de propriedade emergente que se relaciona a um determinado grau suficiente de sofisticação envolvido nos procedimentos computacionais.

O argumento operacional pregado pelo modelo *A* parece apresentar deficiências, pois se todas as manifestações externas de um cérebro consciente, incluindo respostas ao questionamento contínuo, podem de fato ser completamente imitadas por um sistema sob o controle computacional, então seria plausível aceitar que suas manifestações internas de consciência deveriam ser também consideradas em associação com tal simulação.

A aceitação desse tipo de argumento relaciona-se basicamente com o teste de Turing<sup>120</sup> que, em essência, distingue *A* e *B*. Segundo *A*, um robô controlado por qualquer computador que, depois de um questionamento sustentado, se comporta de forma convincente, como se possuísse consciência, deve ser considerado realmente consciente. De acordo com o ponto de vista *B*, um robô poderia perfeitamente se comportar exatamente como uma pessoa consciente pode se comportar sem realmente possuir qualquer qualidade mental.

---

<sup>120</sup> O teste de Turing é um teste da capacidade de uma máquina para exibir um comportamento inteligente. O teste foi introduzido por Alan Turing em 1950, em seu artigo '*Computing Machinery and Intelligence*'.

*A* e *B* permitiriam que um robô controlado por computador poderia se comportar de forma convincente como uma pessoa consciente faz, mas o modelo *C*, por sua vez, não admite a possibilidade de uma simulação totalmente eficaz de uma pessoa consciente através de um robô. Assim, de acordo com *C*, a falta real de consciência do robô deve finalmente se revelar, após um interrogatório suficientemente longo. De fato, *C* é muito mais próximo do modelo *A* sob este ponto de vista do que *B*. Penrose<sup>121</sup> considera o modelo *B* o 'senso científico comum'. Também é referido como a inteligência artificial fraca (IA fraca). Da mesma maneira que *A*, afirma que todos os objetos físicos do mundo devem se comportar de acordo com uma ciência que, em princípio, permite que possam ser computacionalmente simulados. Por outro lado, nega veementemente a alegação operacional que uma coisa que se comporta externamente como um ser consciente deve necessariamente ser consciente em si. Isso vai ao encontro da declaração de Searle<sup>122</sup> que afirma que uma simulação computacional de um processo físico é uma coisa muito diferente do processo propriamente dito. Por exemplo, a simulação em computador de um furacão não é certamente um furacão.

O modelo *C* é o defendido por Penrose<sup>123</sup>. Declara que existem manifestações externas de objetos conscientes (por exemplo, o cérebro) que diferem das manifestações externas de um computador, os efeitos externos da consciência não podem ser adequadamente simulados computacionalmente. Nesta abordagem, talvez um novo entendimento seja de fato necessário, em uma área intermediária e sutil que conecte os domínios macro e micro do universo que nos cerca.

Talvez existam outras possibilidades de combinações dos modelos apresentados ou que possam estar localizadas entre tais modelos, como se fossem gradações dos mesmos. Existe até um ponto de vista que poderia ser considerado como uma combinação de *A* e *D* (ou talvez *B* e *D*). Segundo essa visão, a ação do cérebro é realmente a de um computador, mas um computador de tão extrema

---

<sup>121</sup> Penrose (1994).

<sup>122</sup> Searle (1980, 1994).

<sup>123</sup> Penrose (1989, 1994).

complexidade que a sua simulação estaria além das capacidades do homem e da ciência contemporânea.

#### **4.2.2. Algoritimização da Consciência**

Talvez existam razões sistemáticas pelas quais os métodos atuais da ciência cognitiva e da neurociência não conseguem abarcar os complexos processos da experiência consciente. Para descrever a experiência consciente, precisamos de um ingrediente extra na explicação. Essa sentença soa como um desafio para aqueles que buscam as soluções para o problema difícil da consciência: Qual é o seu ingrediente extra e qual a sua importância para a experiência consciente?

A conjectura anterior enseja uma pergunta: poderíamos desenvolver um computador consciente? Essa questão e outras desse teor têm sido objeto de uma enorme quantidade de debate nas últimas décadas. O campo da inteligência artificial é dedicado em grande parte ao objetivo de reproduzir a mente humana de forma integral através de circuitos eletrônicos. Até agora, o progresso tem sido bastante limitado, apesar de algumas 'proezas' que robôs e computadores têm realizado ao longo dos últimos anos.

Podem-se eleger duas formas de objeções à inteligência artificial forte:

1- Funcionais que são relativas ao comportamento criativo ou flexível que os seres humanos possuem;

2 - A não computabilidade do pensamento matemático baseado no Teorema da Incompletude de Gödel e exaustivamente defendido por Roger Penrose (1989, 1994).

As duas objeções podem ser discutidas de forma conjunta, pois em ambas o comportamento versátil e o processo criativo do cérebro humano parecem ter um papel fundamental.

Penrose baseia-se no teorema de Gödel<sup>124</sup> para tentar provar que o papel da consciência não apresenta um caráter algorítmico ao formular juízos matemáticos e

---

<sup>124</sup> Ver seção 4.1.1.

afirma que esse mesmo caráter possa também ser aplicado a circunstâncias gerais (não matemáticas) do cotidiano.

A partir do referido teorema, demonstra-se que qualquer algoritmo que um matemático pudesse utilizar para estabelecer uma verdade matemática, haverá sempre proposições, como a proposição explícita de Gödel  $P_k(K)$ <sup>125</sup> para as quais seu algoritmo não é capaz de fornecer uma resposta. Logo, se o funcionamento do cérebro do matemático fosse estritamente algorítmico, então o algoritmo (ou sistema formal) por ele utilizado para fazer seus julgamentos seria incapaz de se ocupar da proposição  $P_k(K)$  construída a partir do seu próprio algoritmo pessoal. Podemos constatar (em princípio) que  $P_k(K)$  é verdadeira, e aí surge uma contradição, já que o matemático também poderia ser capaz de fazer o mesmo, resultando assim na conclusão de que ele não estava fazendo uso de um algoritmo.

Penrose<sup>126</sup> lança mão de um exemplo para afirmar tal ideia, propondo que diferentes matemáticos utilizem algoritmos não equivalentes para decidir uma determinada verdade<sup>127</sup> matemática. O autor enfatiza que a matemática seja a única disciplina do conhecimento humano em que a verdade das proposições pode ser estabelecida pela argumentação abstrata. Assim, um matemático que seja convencido por um determinado argumento também convencerá outro colega, a partir do momento que o tenha compreendido. Nesse contexto, as proposições de Gödel podem seguir o mesmo raciocínio. Segundo ele, argumentos que estabelecem a verdade matemática são comunicáveis. No exemplo, continua afirmando a possibilidade de existência de um sistema formal universalmente empregado, equivalente a todos os diferentes algoritmos de matemáticos para julgar a verdade matemática. Diz que esse suposto sistema não pode jamais ser reconhecido como aquele que os matemáticos usam para decidir a verdade, pois, se assim o fosse, poderia se construir a sua própria proposição de Gödel e reconhecer também sua validade como verdade. A conclusão é de que o algoritmo usado pelo

---

<sup>125</sup> Proposição aritmética, cuja verdade não pode ser demonstrada através de regras matemáticas.

<sup>126</sup> Penrose (1989).

<sup>127</sup> Verdade formal é a validade de uma conclusão à qual se chega seguindo as regras de inferência a partir de postulados e axiomas aceitos.

matemático é tão complicado ou obscuro que sua própria validade jamais poderá ser conhecida.

Através do exemplo acima, Penrose<sup>128</sup> procura refutar o argumento dos partidários da IA forte que afirmam que se o matemático se utilizasse de um algoritmo muito complexo não seria possível formular a proposição de Gödel e muito menos constatar sua validade.

#### **4.2.3. O Argumento de Searle**

O argumento do ‘quarto chinês’, imaginado por Searle em 1980, busca refutar os teóricos da IA forte e do funcionalismo. Baseia-se na presunção de que a sintaxe (gramática) não é garantia de existência da semântica (sentido). Searle preocupa-se com a questão da ‘compreensão’ e se uma ação apropriada de um dado computador sofisticado pode ser realizada para alcançar essa qualidade mental.

O experimento pode ser resumido da seguinte forma: O sistema composto por um ser humano, que compreende apenas o português, equipado com um livro de regras escrito em português e diversas pilhas de papel, sendo algumas em branco e outras com inscrições indecifráveis (o ser humano é a CPU, o livro de regras o programa e o papel o dispositivo de armazenamento). O sistema está num quarto com uma pequena abertura para o exterior. Por essa abertura passam papéis com símbolos indecifráveis. O ser humano encontra símbolos correspondentes no livro de regras e segue as instruções que podem incluir símbolos em novas folhas de papel, encontrar símbolos nas pilhas, reorganizar as pilhas, etc. Eventualmente, as instruções farão com que um ou mais símbolos sejam transcritos em uma folha de papel que será repassada ao exterior do quarto. No exterior há um sistema que está recebendo as saídas na forma de instruções em chinês e está gerando respostas em chinês, que são, sem dúvida, ‘inteligentes’. A pessoa no quarto não entende o chinês (dado inicial). O livro de regras e o papel não entendem chinês. Então, não está acontecendo nenhuma compreensão do chinês. Por conseguinte, a execução do programa correto não gera necessariamente compreensão. (SEARLE, 1980, p. 417).

---

<sup>128</sup> Penrose (1989).





Figura 4 – Representação do ‘quarto chinês’. Searle (1980)

Searle argumenta, portanto, que a qualidade mental de compreensão não pode ser apenas uma questão computacional - para o sujeito humano (que não entende o chinês) que executa cada ato único de computação que o computador realiza, mas não experimenta nenhum entendimento qualquer que seja as histórias. Searle permite que uma simulação da saída dos resultados do entendimento poderia ser possível, de acordo com o ponto de vista *B*, uma vez que ele está preparado para admitir que isso pudesse ser conseguido por um computador simulando todas as ações físicas relevantes de um cérebro humano, quando este realmente entender alguma coisa. Mas pelo argumento do quarto chinês, ele insiste que a simulação não pode, em si, realmente ‘sentir’ qualquer entendimento. Assim, a compreensão real não poderia ser de fato alcançada por qualquer simulação computacional.

O argumento de Searle é dirigido contra o ponto de vista *A* (qualquer ‘simulação’ de entendimento seria equivalente à compreensão ‘real’) e é apresentado como fornecendo suporte para o nível *B* (embora igualmente sustentando os níveis *C* ou *D*). Preocupa-se com aspectos interiores e subjetivos da qualidade de entendimento. Não nega a possibilidade de uma simulação de entendimento em seus aspectos ativos, exteriores ou objetivos. Searle acredita que o cérebro é como um computador digital. Isto sugere que estaria preparado para aceitar a possibilidade de uma simulação completa da ação de um cérebro consciente no ato de ‘compreensão’ sobre alguma coisa, de onde as externas

manifestações dessa simulação seriam idênticas aos de um ser humano consciente real, de acordo com o nível *B*.

### 4.3. Computadores Conscientes

Muitos cientistas e filósofos, incluindo aí David Chalmers, acreditam que as leis da física são potencialmente computáveis, tornando totalmente plausível a simulação computacional do comportamento humano. Obviamente essa posição é bastante contestada, com argumentos como os de Penrose de que existiriam elementos não computáveis dentro do contexto das leis físicas ou o argumento de não causalidade física da consciência.

Além das objeções funcionais, há as chamadas objeções experienciais, nomeadas assim por Chalmers. Essas objeções admitem, pelo menos, por causa do argumento de que os computadores podem simular o comportamento, mas lhes faltaria a parcela fenomenal da mente humana, não possuindo vida interna, ausência de experiência consciente e verdadeiro entendimento. Assim, perante tal cenário, um computador poderia fornecer apenas uma simulação da mente humana, mas não sua replicação<sup>129</sup>. A melhor objeção conhecida deste gênero é o chamado ‘quarto chinês’ visto anteriormente. Dentro dessa visão, os computadores seriam apenas ‘caixas ocas de silício’.

Aqueles que têm uma visão não redutível da consciência têm sido atraídos pelas objeções à inteligência artificial experimental, ainda bastante incipiente. Achem difícil de acreditar que um sistema não biológico poderia dar origem à experiência consciente.

---

<sup>129</sup> A Filosofia da Mente contemporânea tem se desenvolvido através de uma estreita relação com os desdobramentos Inteligência Artificial. Uma das questões principais nesta é a de saber se os estados cognitivos humanos podem ser duplicados por computadores. Há basicamente duas posições comuns em relação a esta questão. A posição fraca, segundo a qual os estados cognitivos humanos podem ser apenas simulados por computadores e uma posição forte segundo a qual os estados cognitivos humanos podem ser por eles replicados. Segundo a tese fraca os computadores apenas simulam os estados cognitivos humanos quando perante o mesma entrada, uma mesma saída é produzida. Segundo a tese forte os computadores podem replicar as relações causais internas presentes na cognição humana.

Apesar de ser partidário da irreducibilidade da consciência, Chalmers não possui uma visão crítica da IA forte. Traça importantes considerações sobre alguns critérios de computabilidade da mente humana, de onde surgem, segundo ele, duas questões bastante distintas. A primeira diz respeito à força de conexão entre sistemas físicos e consciência: a consciência é constituída por processos físicos, ou esta simplesmente surge a partir destes processos físicos? A segunda está ligada à forma da conexão: onde a apropriada organização bastaria para que os sistemas físicos dessem lugar à consciência?

Portanto, segundo Chalmers, pode-se considerar que não é óbvio que um determinado tipo de computação possa dar origem à consciência, assim como, à primeira análise, não é óbvio que processos neurais no cérebro proporcionem a origem da consciência. Assim, não haveria nenhuma razão clara para que os computadores devam ser piores do que qualquer cérebro a esse respeito. Dada a aceitação do fato surpreendente de que o cérebro dá origem à consciência, não seria uma surpresa também descobrir que a computação também passa dar origem à consciência. Conclui-se, então, que a adoção de uma visão não reducionista da consciência deveria deixar a questão aberta.

Chalmers faz algumas conjecturas acerca do que Searle chama de inteligência artificial forte. A de que há uma classe não vazia de computação de tal modo que a implementação de qualquer algoritmo nessa classe seja suficiente para a mente e, em particular, suficiente para a existência de experiência consciente. Também é logicamente possível que qualquer computação possa ocorrer na ausência de consciência. Mas também vale para o cérebro, como visto.

Chalmers apoia-se no seu princípio da invariância organizacional<sup>130</sup> que, se estiver correto, estabelece que qualquer sistema com a forma adequada em sua organização funcional é consciente, não importando seu substrato físico. Por consequência desse argumento, uma máquina constituída a partir de componentes de silício poderia ser consciente.

---

<sup>130</sup> Ver seção 2.6.

O que ainda permanece nebuloso em tal argumento é como é realizada a conexão entre a computação e a organização funcional, a fim de estabelecer que a implementação de uma computação adequada seria suficiente para garantir a presença do respectivo ambiente para o surgimento da consciência. Isto alcançado, a IA forte seria apenas uma consequência natural.

#### 4.3.1. Implementação de Algoritmos

A teoria da computação lida inteiramente com objetos abstratos: máquinas de Turing, programas em linguagens específicas, autômatos de estado finito e assim por diante. São entidades matemáticas que habitam o espaço matemático abstrato. Os sistemas cognitivos do mundo real são objetos concretos, fisicamente incorporados, e interagem causalmente com outros objetos no mundo físico. Mas muitas vezes queremos usar a teoria da computação para tirar conclusões sobre objetos concretos no mundo real. Para fazer isso, há a necessidade de uma ponte entre os domínios abstrato e concreto.

Searle (1990b) argumentou que a computabilidade não é uma questão objetiva, mas um 'observador relativo' no qual qualquer sistema pode ser implementado por qualquer computação se interpretado apropriadamente. Chalmers<sup>131</sup> refuta tal ideia, pois, segundo ele, se assim fosse, seria difícil ver como noções computacionais poderiam desempenhar qualquer papel fundamental em uma teoria que, em última análise, trata de sistemas concretos. Assim a IA forte seria carente de conteúdo ou implicaria uma forte forma de pampsiquismo<sup>132</sup>.

Qualquer consideração sobre qual tipo de computação a ser implementada dependerá da classe de desafios em questão. Há inúmeros formalismos computacionais, com diferentes classes de implementação, como já mencionado: máquinas de Turing, autômatos de estado finito, redes conexionistas, autômatos celulares, entre outros. Em princípio, é necessária uma descrição de implementação ou computabilidade para cada um desses formalismos. Chalmers considera em sua

---

<sup>131</sup> Chalmers (1996).

<sup>132</sup> Doutrina setecentista, notabilizada por Leibniz (1646-1716), segundo a qual a matéria possui uma essência espiritual ou anímica, um fundamento psíquico inextenso subjacente aos atributos especiais ou tridimensionais.

proposta a implementação de um formalismo único, o de autômatos de estados combinatórios (CSA). Na sua opinião, essa classe de computabilidade é suficientemente geral, tanto que sua implementação associada pode ser facilmente estendida para a aplicação a outras classes.

Um autômato de estado combinatório (CSA) é um modelo mais sofisticado de um autômato de estado finito (FSA). Um autômato de estado finito (FSA) é especificado através de um conjunto finito de entradas, um conjunto finito de estados internos e um conjunto finito de saídas, fornecendo um conjunto associado de relações de transição de estado. Um estado interno de um FSA é um elemento simples  $S^i$ , sem qualquer estrutura interna, o mesmo valendo para entradas e saídas.

As relações de transição de estado especificam, para cada possível par entrada / estado interno, um novo estado interno e uma saída. Se o estado inicial de um FSA é dado, essas relações de transição de estado especificam como ele irá evoluir ao longo do tempo e quais saídas irá produzir, dependendo de quais entradas são recebidas. A estrutura computacional de um FSA consiste no conjunto relativamente simples de transições de estado das relações de um conjunto de estados não estruturados.

Autômatos de estado finito (FSA) são inadequados para representar a estrutura da maioria das implementações relevantes na prática, como os estados e suas transições. As relações entre tais transições geralmente têm estrutura interna complexa. Nenhuma descrição a partir de um FSA pode cobrir toda a estrutura presente em um programa Pascal, por exemplo, ou numa máquina de Turing, ou em um autômato celular. Portanto, é mais útil se concentrar em uma classe de autômatos que ofereça uma estrutura de estados internos estruturados.

Autômatos de estados combinatórios (CSAs) são semelhantes aos FSAs, exceto que os seus estados internos são estruturados. Um estado de um CSA é um vetor  $[S^1, S^2, \dots, S^n]$ . Esse vetor pode ser finito ou infinito. O caso considerado será o caso finito. Os elementos desses vetores podem ser considerados como componentes de estados internos, pois correspondem às células em um autômato celular ou a cabeça de impressão em uma máquina de Turing.

Cada elemento  $S^i$  pode assumir um número finito de valores  $S^i_j$ , onde  $S^i_j$  é o valor 'j-ésimo' possível do 'i-ésimo' elemento. Esses valores podem ser pensados como 'subestados' do estado geral. Entradas e saídas têm um tipo similar de estrutura complexa: uma de entrada é um vetor  $[I^1, \dots, I^k]$  e uma saída é um vetor  $[O^1, \dots, O^m]$ .

Um CSA é determinado pela especificação do conjunto de vetores de estado interno e vetores de entrada e saída, e pela especificação de um conjunto de regras de transição de estado que determinam como o estado do CSA evolui através do tempo. Para cada elemento do vetor de estado interno, uma regra de transição de estado determina como seu novo valor depende dos valores anteriores de entrada e dos vetores de estado interno.

Para cada elemento do vetor de saída, uma regra de transição de estado determina como seu novo valor depende de antigos valores do vetor de estado interno. Cada CSA finito pode ser representado como um FSA com o poder computacional igual, mas a descrição do FSA vai sacrificar a maior parte da estrutura que é crucial para um CSA. Segundo Chalmers (1996), essa estrutura é central no uso de CSAs para capturar de forma mais adequada e eficaz a organização subjacente à mente.

Assim, a questão sobre a computabilidade parece aqui apontar para uma resolução mais completa. Os CSAs são objetos abstratos, com uma estrutura formal determinada por seus estados e suas relações de transição de estado. Sistemas físicos são objetos concretos, com uma estrutura causal determinada por seus estados internos e as relações causais entre tais estados.

De maneira não formal, pode-se dizer que um sistema físico implementa uma computação quando a estrutura causal do sistema espelha a estrutura formal da computação. Ou seja, o sistema implementa a computação se existe uma forma de mapeamento dos estados deste para os estados físicos que são causalmente relacionados com os estados que são formalmente ligados aos primeiros de forma correspondente.

Essa ideia intuitiva pode ser diretamente aplicada ao rendimento de um cálculo de implementação de CSAs. Um sistema físico implementa um CSA se há uma decomposição de estados internos do sistema em subestados. Uma decomposição das entradas e saídas dentro de subestados correspondentes e um mapeamento de subestados do sistema em relação aos subestados do CSA, de tal forma que as relações de transição causal entre estados físicos, entradas e saídas possam refletir as relações de estado formal de transição entre os estados formais correspondentes.

O critério formal de implementação de CSA é o seguinte:

Um sistema físico  $P$  implementa um CSA  $M$  se há uma decomposição dos estados internos de  $P$  em componentes  $[s^1, \dots, s^n]$ , e um mapeamento  $f$  dos seus subestados  $s^j$  nos correspondentes subestados  $S^j$  de  $M$ , juntamente com vetorizações similares e mapeamentos para as entradas e saídas, de modo que para cada regra de transição de estado  $([I^1, \dots, I^k], [S^1, \dots, S^n]) \rightarrow ([S^1, \dots, S^n], [O^1, \dots, O^l])$  de  $M$ : se  $P$  esta em um estado interno  $[s^1, \dots, s^n]$  e recebendo entrada  $[i^1, \dots, i^k]$  os quais são mapeados para os estados e entradas formais  $[S^1, \dots, S^n]$  e  $[I^1, \dots, I^k]$  respectivamente, isso de forma confiável faz com que ele entre em um estado interno e produza uma saída que projeta seus efeitos para  $[S^1, \dots, S^n]$  e  $[O^1, \dots, O^l]$  respectivamente. (CHALMERS, 1996, p. 297).

Pode-se prever que, em uma decomposição do estado de um sistema físico em um vetor de subestados, o valor de cada elemento do vetor deve sobrevir em uma região separada do sistema físico, para garantir que a organização causal relacione componentes distintos do sistema. Caso contrário, não é claro que a detalhada estrutura causal está realmente presente no sistema físico.

Pode parecer que CSAs não têm muitas vantagens sobre FSAs. Pois afinal, para qualquer CSA finito, pode-se encontrar uma FSA correspondente com o mesmo comportamento de entrada / saída. Mas há algumas diferenças cruciais:

- 1- As condições de aplicação em um CSA são muito mais limitadas do que as de um FSA correspondente. Uma implementação de um CSA é necessária para compor-se em uma complexa interação causal entre um número de partes separadas. Uma descrição de CSA pode, portanto,

apreender a organização causal de um sistema de uma maneira muito mais refinada;

- 2- CSAs fornecem uma explicação unificada das condições de computabilidade para máquinas finitas ou não;
- 3- O CSA pode refletir diretamente a organização complexa formal de objetos computacionais, tais como máquinas de Turing e autômatos celulares. Nos FSAs correspondentes, grande parte dessa estrutura seria perdida.

Na verdade, pode-se utilizar essa definição de computabilidade para fornecer critérios de implementação direta para outros tipos de computação. Para especificar o que é preciso para implementar uma máquina de Turing, por exemplo, precisamos apenas descrevê-la como um CSA e aplicar as definições. Para tanto, descrevemos o estado da máquina de Turing como um vetor gigante. Um elemento do vetor representa o estado da cabeça de impressão, e há um elemento para cada quadrado (espaço) da fita, representando o símbolo no quadrado e também indicando se a cabeça de impressão ocupa ou não aquele determinado quadrado.

As regras de transição de estado entre os vetores são aquelas derivadas naturalmente das possibilidades de comportamento da cabeça de impressão e o avanço e recuo da fita. Os vetores aqui são infinitos, mas as condições de implementação, no caso, são uma extensão direta do caso finito de estados. Dada essa tradução ao formalismo do CSA a partir do formalismo da Máquina de Turing, pode-se dizer que esta é implementada sempre que o CSA correspondente é implementado. Assim, pode-se promover traduções similares em processos computacionais a partir de outros formalismos, como autômatos celulares ou programas escritos em linguagem Pascal, por exemplo, gerando condições de implementação computacional em cada uma dessas diferentes classes.

Isso gera um critério perfeitamente objetivo de computabilidade. Por exemplo, um único elemento de estado do CSA será implementado por cada sistema, e um CSA de dois estados será implementado de forma mais ampla. Também é verdade que a maioria dos sistemas irá implementar mais de uma computação, dependendo de como pode ser dividido os estados do sistema. É totalmente concebível que se



um computador pode implementar um número de diferentes computações, nosso cérebro também possa fazer o mesmo.

Um ponto importante é que não há nenhuma razão para se acreditar que cada CSA será implementado por cada sistema. Para qualquer CSA complexo, poucos sistemas físicos terão a organização causal necessária para implementá-lo. Um CSA, cujo estado de vetores tem 1000 elementos, com dez possibilidades para cada elemento, em seguida, sugerem que a chance de um conjunto arbitrário de estados físicos tendo as relações de causalidade necessárias é algo menos do que um em um montante  $(10^{1000}) \times 10^{1000}$  (na verdade, muito menos do que isso, por causa da exigência de que as relações de transição possam ser coerentes).

Não parece claro que as afirmações de Searle sobre as representações computacionais assumam o papel de um 'observador relativo'. É verdade que há um grau limitado de observação da relatividade: qualquer sistema físico irá implementar uma série de computações, e qual delas o observador escolherá dependerá de seus propósitos.

Dizer que um sistema físico implementa uma certa computação complexa  $P$  é afirmar algo muito importante sobre a estrutura causal do sistema, algo que pode ser muito útil para fornecer explicações cognitivas e, talvez, no entendimento da base da consciência. Sistemas apenas com um tipo muito específico de organização causal terão uma esperança de satisfazer as fortes restrições de computabilidade. Assim, pode não haver perigo de vacuidade, criando espaço para se acreditar que a noção de computabilidade pode fornecer uma base sólida para a análise de sistemas cognitivos.

#### **4.4. Inteligência Artificial Forte?**

O que é preciso para implementar um CSA é muito semelhante ao que é preciso para realizar uma organização funcional. Uma organização funcional é determinada, especificando-se um número de componentes abstratos, um número de estados para cada componente e um sistema de relações de dependência que

indica como o estado de cada componente depende de estados anteriores e das entradas e saídas. A noção de um CSA é efetivamente uma formalização direta dessa noção.

Uma organização funcional pode ser deduzida diretamente em um CSA. Torna-se necessário estipular quais vetores de estado do CSA têm um elemento para cada componente da organização, e as transições de estado do CSA devem corresponder às relações de dependência causal entre os componentes do sistema. Compreender a organização funcional é quase o mesmo que implementar o CSA correspondente, ficando mais forte, desse modo, a ligação entre a organização causal de um sistema com sua respectiva computação.

Portanto, quando descrições computacionais são aplicadas a sistemas físicos, estas efetivamente fornecem uma descrição formal da organização causal dos mesmos.

A linguagem de computação fornece uma linguagem perfeita em que este tipo de organização causal abstrata pode ser especificado. Pode-se argumentar que é precisamente por esse motivo que noções computacionais têm tido ampla aplicação em toda a ciência cognitiva. O que é mais relevante para a explicação do comportamento de um sistema cognitivo complexo é a organização abstrata causal do sistema, e formalismos computacionais fornecem um bom cenário dentro do qual esse tipo de organização pode ser descrita e analisada.

De acordo com Chalmers, essa conexão faz a defesa da IA forte. Baseado em seu princípio de invariância organizacional<sup>133</sup>, que afirma que para qualquer sistema dotado de experiências conscientes, um sistema com a mesma organização detalhadamente ajustada poderá ter experiências conscientes qualitativamente semelhantes, mas talvez não idênticas. Mas sabe-se que uma determinada organização funcional pode ser deduzida em um CSA, que é implementado sempre que a organização é compreendida.

---

<sup>133</sup> Ver seção 2.6.

Segue-se que para um dado sistema  $M$  consciente, a sua fina e detalhada organização funcional pode ser deduzida em um CSA  $M$ , de tal forma que qualquer sistema que implemente  $M$  vai perceber a mesma organização funcional, e, portanto, ter experiências conscientes qualitativamente indistinguíveis daquelas do sistema original. Isto estabelece a tese de inteligência artificial forte.

Por exemplo, poderíamos deduzir uma descrição neural do cérebro em um CSA, com um elemento do vetor de estado para cada neurônio, e com subestados para cada elemento que pudessem refletir a faixa relevante de estados de cada neurônio. As regras de transição de estado do CSA refletem a maneira pela qual o estado de cada neurônio depende do estado de outros neurônios, e da maneira em que os estados neurais estão relacionados com as suas respectivas entradas e saídas. Se componentes não neurais do cérebro são relevantes, pode-se considerar componentes computacionais para estes também. Mas o problema torna-se em saber quais seriam eles. Qualquer sistema físico que implemente esse CSA terá uma organização funcional tão fina e sofisticada que poderia teoricamente replicar o nível funcional dos neurônios do cérebro. Pelo princípio de invariância, esse sistema poderá ter experiências semelhantes daquelas apreendidas pelo cérebro.

Chalmers alerta para a forma, muitas vezes simplista, de como vemos certas coisas. Exemplifica essa afirmação, relatando que quando vemos um computador apenas como um dispositivo de entrada e saída e algumas operações matemáticas nele embutidas, muitas vezes deixamos de fora o fato fundamental de que existem complexas dinâmicas causais dentro de um computador, assim como há também no cérebro. De fato, em um computador comum que implemente uma simulação neurônio por neurônio do cérebro, haverá relações causais reais acontecendo entre tensões de vários circuitos que precisamente espelham os padrões de causalidade entre os neurônios. Para cada célula neuronal haverá uma alocação de memória ou uma alocação que a representa, e cada um desses locais será fisicamente representado em uma tensão em algum local físico da máquina. O autor afirma ainda que o padrão causal entre tais circuitos, assim como é o padrão causal entre os neurônios no cérebro, é responsável por qualquer experiência consciente que possa surgir.

No entanto, alguns podem questionar essa tendência. Por exemplo, talvez um número infinito de estados sejam necessários para cada neurônio, para capturar o papel vital de processamento 'contínuo'. E alguns podem afirmar que as transições entre esses estados infinitos possam ser não computáveis, não originando assim estados conscientes.

As possibilidades levantadas sobre a computabilidade da mente são ainda bastante conjecturais, apoiando-se, muitas vezes, em argumentos intuitivos ou filosóficos. As hipóteses são bastante ousadas, mas muito distantes ainda de qualquer perspectiva para a construção ou simulação de ambientes ou estruturas fenomenologicamente conscientes.

#### **4.5. Inteligência Artificial e Platonismo**

A IA forte sustenta que a mente percebe sua existência pelo processamento de um algoritmo complexo, quando este é executado por estruturas do mundo físico, no caso, a organização cerebral, não atribuindo importância a tal organização. Apenas o algoritmo parece ser valorizado, mas para que este exista independentemente de qualquer estrutura física, parece necessário um ponto de vista exterior da matemática.

Torna-se tarefa difícil para um partidário da IA forte sustentar que os conceitos matemáticos, por exemplo, sejam provenientes das mentes, já que tal concepção criaria uma circularidade, que exigiria mentes preexistentes para a existência de algoritmos e vice-versa.

Apesar de já termos levantado a questão sobre um possível caráter monista na teoria proposta por Penrose, esta dificuldade existente no escopo da IA forte encontra bastante ressonância em suas ideias. Como já identificado anteriormente que as mentes, segundo ele, são entidades não algorítmicas. Penrose acredita que a consciência esteja relacionada estreitamente com a percepção das verdades necessárias, isto é, um contato direto com o mundo platônico dos conceitos matemáticos, tarefa esta completamente não algorítmica, segundo sua proposta.

Penrose<sup>134</sup> acredita que o mundo físico ‘emerge’ do mundo platônico, e este último tenha uma parcela de si em intersecção com o mundo físico.

Em sua obra de 1989, Penrose refuta a possibilidade de que certas teorias, entre elas a Relatividade e a Mecânica Quântica, tenham surgido a partir de, como ele mesmo chama, ‘seleção natural fortuita de ideias’. Acredita numa profunda razão subjacente para a concordância entre a Matemática e a Física, para ele, entre o mundo platônico e o físico. Para tanto, atribui ao mundo de Platão algum nível de realidade comparável ao nível físico. As teorias da física moderna parecem ter imposto uma característica mais nebulosa também à realidade que nos cerca. A precisão de tais teorias atribuiu, segundo ele, uma existência matemática quase abstrata à realidade.

Serão esses dois mundos faces diferentes da mesma realidade? Através desta identificação, torna-se mais compreensível como a mente parece evidenciar e ser a responsável pelas misteriosas relações entre tais mundos. Essas relações talvez possam ser mais bem elucidadas mediante as palavras de Einstein, sobre a não verbalidade do pensamento:

As palavras da língua, tal como escritas ou faladas, não parecem desempenhar qualquer papel em meu mecanismo de pensamento. As entidades psíquicas que parecem servir como elementos do pensamento são certos sinais e imagens mais ou menos claros que podem ser produzidos e combinados ‘voluntariamente’... Os elementos mencionados são, no meu caso, do tipo visual e outros do tipo muscular. As palavras convencionais ou outros sinais têm de ser procurados laboriosamente numa segunda etapa, quando o jogo associativo de que falei já está bem estabelecido e pode ser reproduzido à vontade.

Citação de Albert Einstein na obra *A Mathematician's Mind* de Jacques S. Hadamard (1996).

---

<sup>134</sup> Penrose (1996b).

## 5 CONCLUSÃO

Ao longo da história, o homem avançou profundamente em todas as áreas, por ele concebidas, o que possibilitou um melhor e mais acurado entendimento do mundo. Desde o fogo e a roda, o espírito desbravador e a criatividade, passados tantos avanços, ainda nos assombram. Mas, apesar de tudo o que construímos e controlamos a nossa volta, talvez o aspecto mais fundamental ainda permaneça sem resposta. Mesmo após milênios de conjecturas e tentativas explanatórias, precipitadas nas mais diversas áreas do conhecimento, carecemos de soluções convincentes em nossas relações básicas com o mundo. Como mostrado no segundo capítulo, algumas questões filosóficas solapam as tentativas da ciência em responder aos problemas apresentados.

Este trabalho apresentou dois fluxos incipientes e sem comprovação metodológica, em busca da mesma resposta. Um relacionado basicamente com as questões filosóficas, que apontam numa direção *top down* em que partimos de conjecturas mentais em direção ao mundo da realidade física, do modo como estamos ‘presentes’ no mundo. O outro que, a partir do mundo subatômico das partículas elementares, a consciência ‘emergiria’ através de processos quânticos num fluxo *bottom up* em direção ao mundo cotidiano. Ambas as abordagens foram discutidas em termos de computabilidade e as perspectivas em que a inteligência artificial forte parece caminhar. Em ambas as abordagens o *gap* explanatório persiste.

Como apontado no capítulo 2 do presente trabalho, o problema da experiência parece ainda não ter uma perspectiva de resolução. Tal problema, chamado de *hard problem* da consciência, pelo filósofo da mente David Chalmers, refere-se ao problema de difícil explanação, que são as experiências qualitativas fenomenais. Como podem alguns organismos estar sujeitos à experiência e por que qualquer transformação de processos físicos poderia originar uma riquíssima e significativa vida interior? A grande questão que permeia os estudos da mente parece centrar-se em torno de qual é a metodologia mais adequada para a compreensão da origem e natureza da consciência. Como descrito na seção 4.2.3, o argumento do ‘quarto chinês’ de Searle tenta mostrar que uma máquina, como um

computador, não pode ser adequadamente descrita como capaz de ‘possuir’ uma ‘mente’ ou ter a capacidade de compreensão, independentemente de quão inteligente ela possa se comportar. O argumento vai de encontro ao ponto de vista do funcionalismo e, especificamente, da teoria computacional da mente, pois os modelos atuais e a nossa forma de compreensão não permitem ainda esclarecer o ‘fenômeno’ da consciência, pois o funcionalismo ignora completamente o *qualia*, as experiências interiores, como atributos básicos de um ser consciente.

Diferentemente dos chamados ‘problemas fáceis’ da consciência, tais como a capacidade de um ser humano discriminar estímulos sensoriais e reagir a eles de forma adequada, ou a integração das informações de muitas fontes diferentes e sua utilização para determinar certos comportamentos, o ‘problema difícil’, ao contrário, é a questão de como os processos físicos do cérebro dão origem à experiência subjetiva. Essa questão envolve os aspectos internos do pensamento e da percepção e da maneira como as coisas são ‘sentidas’ pelo sujeito, o qual é o único conhecedor de tais experiências. A experiência é um ponto crucial no estudo da mente, pois, ao se comparar estados cerebrais, que são, em princípio, objetivos e replicáveis, com experiências subjetivas que, por mais ricas que sejam para o experimentador, não estão disponíveis para um observador externo. Assim, a subjetividade da experiência do mundo representa um grave problema de diferença qualitativa, o que constitui ainda uma barreira quase intransponível.

Até hoje, a neurociência parece não apontar para a resolução da questão. Pode-se, por exemplo, lembrar a chamada oscilação gama (40 Hz) que veio a público através de Crick e Koch (1992) e outros, que poderia fornecer uma explicação para a ligação coerente conjunta de diferentes regiões cerebrais, por exemplo, os aspectos visual e auditivo de uma percepção podem explicar como o cérebro integra tarefas de processamento diferentes – problema fácil, dentro da conceituação chalmérica –, mas não explica como são evocadas as experiências subjetivas do som e da visão.

Os argumentos apresentados na seção 2.4.1 sustentam que a consciência não é superveniente ao mundo físico. A visão materialista, apesar de seus argumentos, não sustenta satisfatoriamente seu objetivo de afirmar a consciência

como apenas uma decorrência dos vários processos físico-químicos que ocorrem no cérebro. Alguns partidários da visão materialista procuram sustentar que o aspecto subjetivo, comumente utilizado, não existe e não possui caráter científico, e que a sua redefinição é a única descrição coerente da consciência. Negar simplesmente certos argumentos, como por exemplo, Dennett faz em relação ao *qualia*, não parece ser um caminho consistente na solução das questões em evidência.

Foi apresentada no capítulo 3 a teoria proposta por Roger Penrose e Stuart Hameroff, que afirma que a consciência surge a partir de processos quânticos que ocorreriam nos microtúbulos, sendo estas estruturas constituídas de proteínas no interior dos neurônios. A partir do fato que ainda hoje nem a neurociência nem tampouco as ciências cognitivas podem apontar para uma perspectiva clara sobre as questões sobre a consciência, torna-se perfeitamente plausível a possibilidade da abordagem através da Física, o campo do conhecimento humano que possui o maior domínio sobre as relações dos processos subatômicos que estão presentes em toda a natureza, inclusive no cérebro. Dessa forma, uma teoria física foi descrita no trabalho por apresentar um caminho lógico e consistente em direção a um conhecimento mais profundo sobre os milhões de processos que se apresentam no vasto campo de interações que o cérebro mantém.

O modelo Penrose-Hameroff, apesar de seu nível de detalhamento e aparente coerência em alguns de seus pontos, ainda se encontra muito longe de explicar fisicamente como a consciência torna-se 'presente' no mundo. Críticos a ele não faltam, como a excelente argumentação em que Pinguelli e Faber se debruçam e pode ser vislumbrada na seção 3.5. Considero a parte mais relevante do modelo, o ponto de vista a partir do qual Penrose inicia suas conjecturas. Esse ponto inicia-se a partir dos Teoremas de Gödel, em que o autor centra os aspectos não computáveis da mente humana. A não computabilidade por si só não aborda diretamente o *hard problem* da natureza da experiência, mas talvez possa apresentar uma pista para o tipo de atividade que está por trás dela. Em todas as tentativas de confrontar as questões da consciência, seja através de teorias físicas ou não, o caráter subjetivo da experiência consciente parece ainda longe de qualquer resolução.



No capítulo 4, foram apresentadas as discussões mais diretas sobre as possibilidades de uma computabilidade da mente e suas implicações e limites. Foram descritas duas objeções feitas por Penrose<sup>135</sup> bastante pertinentes em relação às perspectivas da inteligência artificial forte. Objeções funcionais relativas ao comportamento criativo ou flexível que os seres humanos possuem e a outra, baseada na não computabilidade das proposições aritméticas, dentro do pensamento matemático formal, constantes dos Teoremas de Gödel.

Diretamente ligadas às objeções funcionais, há as chamadas objeções experienciais. Estas admitem que, mesmo que computadores possam simular o comportamento humano, a parcela fenomenal da mente estaria ausente, ocasionando a ausência de vida interna e experiência consciente. Portanto, é razoável admitir que um computador poderia fornecer apenas uma simulação da mentalidade humana, mas não sua replicação. Essa hipótese é reforçada pelo argumento de Searle do 'quarto chinês'<sup>136</sup>, em que os computadores seriam apenas caixas ocas de silício, sem possibilidade de introspecção.

Mesmo com todos os avanços e os exemplos cotidianos das conquistas no terreno da computação e da robótica, ainda é difícil de acreditar que um sistema não biológico possa dar origem à experiência consciente. Portanto, pode-se considerar que não é óbvio que um determinado tipo de computação possa gerar consciência, assim como, à primeira análise, também não parece nada óbvio que somente processos neurais no cérebro proporcionem a sua origem. Assim, essa analogia entre o aparente fato da estrutura físico-química do cérebro originar a consciência e a possibilidade de uma determinada classe de computação fazer 'emergir' um sistema consciente é onde Chalmers apoia-se na ideia que a sofisticada coerência estrutural e a invariância organizacional do cérebro humano deveriam também estar presentes, num pretenso candidato não humano de forma a possibilitarem a emergência de experiências conscientes.

O que ainda permanece nebuloso nesse argumento é como é realizada a ligação entre a computação e a estrutura organizacional, a fim de estabelecer que a

---

<sup>135</sup> Penrose (1994).

<sup>136</sup> Vide seção 4.3.1.

implementação de uma computação adequada fosse suficiente para garantir a presença do respectivo ambiente para o surgimento da consciência. Alguns autores defendem que a existência de padrões causais em circuitos artificiais, similares aos existentes nos neurônios do cérebro, seriam os responsáveis por quaisquer experiências conscientes que pudessem emergir. Parece um tanto fraco tal argumento, pois talvez um número infinito de estados seja necessário para a replicação de cada neurônio, para capturar o papel vital do processamento destes. Adicionalmente, as transições entre os estados podem apresentar caracteres não computáveis, não ocasionando assim os pretendidos estados conscientes.

Sustentar que os conceitos matemáticos, por exemplo, sejam provenientes das mentes, pode se tornar uma árdua tarefa para um partidário da IA forte, já que isso criaria uma circularidade, que exigiria mentes preexistentes para a existência de algoritmos e vice-versa.

Em todas as relações e modelos apresentados ao longo da discussão, sejam eles físicos, filosóficos ou computacionais, nenhum até agora pôde explicar ou proporcionar algum sentido à forma como a experiência consciente, ou seja, o *hard problem* da consciência, está presente no mundo e a forma pela qual o percebemos.

Embora a descrição científica baseie-se exclusivamente sobre o universo físico objetivo, nosso contato com a realidade é essencialmente alicerçada pela nossa experiência subjetiva consciente. Do nascimento à morte, nós experimentamos apenas um fluxo de consciência através da qual toda a nossa experiência do mundo físico é adquirida. Todos os experimentos científicos realizados no mundo físico são mediados e validados pela experiência subjetiva consciente dos experimentadores.

Diante de sua natureza subjetiva, uma investigação objetiva, dentro do escopo da metodologia científica atual, parece ser totalmente incompatível. As descrições reducionistas identificam a consciência subjetiva com atributos funcionais do cérebro, inferindo que máquinas poderiam apresentar características conscientes. Na melhor das hipóteses, em tais pontos de vista, a mente consciente subjetiva ainda permanece um enigma considerado apenas como 'epifenômeno' da estrutura

biológica do cérebro. No entanto, é o mundo físico, que é secundário à nossa experiência pessoal, que parece ter um caráter consensual, a partir do qual moldamos representações subjetivas estáveis que ‘montam’ a nossa visão do mundo real. Essa perspectiva talvez não deixe claro se um universo físico, sem observadores conscientes, poderia existir além de um sentido puramente conceitual ou teórico. A consciência subjetiva pode ser necessária para a existência da realidade física e, portanto, fundamental para a existência física em um sentido cosmológico. Essa ideia foi expressa como ‘princípio antrópico cosmológico’, em que ‘observadores’ são significativos e, possivelmente, condições de contorno necessárias para a existência do universo (Barrow e Tipler, 1988).

A forma como pensamos o mundo, como um quebra-cabeça, onde cada peça deve ser reduzida de modo a ser explicada através de leis fundamentais, talvez possa estar próxima de um esgotamento. Poderíamos pensar o mundo como um conjunto de mapas, como um atlas que representa cada aspecto da realidade em que estamos inseridos. A realidade geográfica, a política, a econômica onde as informações constantes em cada uma abarca a realidade que se busca explicar. Tentar ‘reduzir’ um mapa econômico para um geográfico, dada a incongruência das informações, não possibilitaria uma explicação adequada, como também poderíamos perder muito do conteúdo da informação na transição proposta.

Se a existência da consciência não pode ser derivada apenas das leis físicas, portanto, uma teoria da física não pode ter a pretensão de ser uma ‘teoria de tudo’. Assim sendo, uma teoria final deve conter um componente adicional crucial, em que a experiência consciente possa ser considerada como uma característica fundamental e irreduzível a qualquer coisa mais básica.

Toda a dissertação foi delineada tendo com autores base Chalmers e Penrose. Tal escolha ocorreu em virtude de pontos importantes que caracterizam as ideias de cada um. Primeiro Chalmers, que pode ser considerado um dualista por suas ideias, aponta para uma teoria não reducionista, mas de forma incomum e consistente, prefere deixar a questão totalmente aberta, buscando convencer-se de que a IA forte pode ser, sim, um caminho bastante plausível nos desdobramentos futuros da ciência contemporânea. E do outro lado, Penrose, que, através do modelo

que propõe, busca explicar o surgimento da consciência no mundo a partir de fenômenos e processos físicos, no qual, à primeira vista, poderíamos enquadrá-lo dentro de um campo materialista. Defende que a mente não possa ser replicada através de procedimentos e lógicas computacionais, tendo em vista a não computabilidade de certos processos envolvidos na geração da consciência.

O trabalho chega ao seu desfecho e, acredito, após toda a discussão, que podemos sim, explicar a mente através de processos físicos, nos quais modelos, como o de Penrose podem apontar para caminhos mais sólidos e promissores em busca dos mecanismos e processos que geram os fenômenos que vivenciamos. O campo da IA tem muito ainda o que desbravar. 'Proezas' ainda maiores serão realizadas pelas máquinas, mas também acredito que certas capacidades que detemos não podem ser apreendidas por quaisquer sistemas não humanos.

A objetividade puramente científica nem sempre foi a melhor escolha para explicações de certos fenômenos a nossa volta e, talvez, também não seja para a consciência. Creio que os dois atributos da vida, o físico e o mental, não sejam dois campos diferentes, mas talvez apenas faces distintas de uma mesma moeda, apenas dois pontos de vista: interior e exterior que se complementam. Isto pode ampliar nosso campo de visão, talvez possibilitando enxergar o homem como um ser integral conectado ao mundo que o cerca.

Ao longo do processo humano no planeta, a via que utilizamos foi sempre a de fora para dentro, ou seja, a percepção do mundo e a constatação empírica como seus aspectos norteadores. Talvez, antes de continuarmos nessa longa empreitada, devamos voltar aos tempos imemoriais e recorrer ao templo de Delfos e incluirmos o '*conhece-te a ti mesmo*' ao caminho, estabelecendo a segunda via, a do dentro para fora... Mas a ciência, em seu modelo atual, não pode comportar tal abordagem...

## 6 REFERÊNCIAS

ABBAGNANO, Nicola. *Dicionário de filosofia*. Martins Fontes. São Paulo, 1998.

ARMSTRONG, D. *What is consciousness? In The Nature of Mind*. Cornell University Press. Ithaca, 1981.

ATEMA, J. *Microtubule theory of sensory transduction*. Journal of Theoretical Biology, 38, 181-90. 1973.

BANDYOPADHYAY, A. (2011) Direct experimental evidence for quantum states in microtubules and topological invariance. Abstracts: Toward a Science of Consciousness. Sockholm, Sweden, 20011. Disponível em: <<http://www.consciousness.arizona.edu/documents/FullProgramandAbstractsTSC2011Stockholm.pdf>>. Acesso em 05 nov. 2011.

BARABÁSI, A. L. *Linked: how everything is connected to everything else and what it means for business*. Science and Everyday Life. Plume-Penguin Group Inc. New York, 2002.

BARROW, J.D.; TIPLER, F.J. *The anthropic cosmological principle*. Oxford University Press. Oxford, 1986.

BATESON, G. *Steps to an ecology of mind*. Chandler Publishing. San Francisco, 1972.

BENIOFF, P. Quantum mechanical hamiltonian models of Turing Machines. *Journal of Statistical Physics*, 29, 515-46. New York, 1982.

BLOCK, N. *On a confusion about the function of consciousness*. Behavioral and Brain Sciences. 18, 227-287. Cambridge, 1995.

CARNIELLI, Walter; EPSTEIN, Richard L. *Computabilidade, funções computáveis, lógica e os fundamentos da matemática*. UNESP. São Paulo, 2006.

CARRUTHERS, P. *Phenomenal consciousness*. Cambridge University Press. Cambridge, 2000.

CHALMERS, David. *Consciousness and its place in nature*. Philosophy of Mind: Classical and Contemporary Readings. Oxford, 2002.

\_\_\_\_\_. *The conscious mind*. Search of a Fundamental Theory. Oxford University Press. New York, 1996.

\_\_\_\_\_. *How can we construct a science of consciousness?* Disponível em: <<http://.consc.net/cons-papers.html>>. Acesso em 29 set. 2011.

\_\_\_\_\_. *Facing up to the problem of consciousness*. Journal of Consciousness Studies, 2(3):200-19. Tucson, 1995.

COPELAND, B. Jack. Computation. In FLORIDI, Luciano (ed.). *The Blackwell guide to the philosophy of computing and information*. Wiley-Blackwell, 2004.

CORNELLI, Gabriele. *As origens da alma*. Os Gregos e o Conceito de Alma de Homero a Aristóteles. São Paulo: Annablume, 2010.

CRICK, F. and KOCH, C. *Toward a neurobiological theory of consciousness*. Seminars in the Neurosciences, n.2 263-275. 1990.

DAMASIO, A. *The feeling of what happens. Body and emotion in the making of consciousness*. Harcourt Inc. San Diego, 1999.

DAVIDSON, D. *Actions, Reasons, and Causes*. Oxford: Reprinted in Essays on Actions and Events, 3-19. Clarendon Press, 1963.

\_\_\_\_\_. *Mental Events*. 1970. Oxford:Reprinted in Essays on Actions and Events, 207-225. Clarendon Press. 1980.

DAWKINS, R. *The blind watchmaker*. Longman. London, 1986.

DENNETT, Daniel C. *Consciousness explained*. Back Bay Books, Brown and Company Hachette Book Group. New York, 1991.

DEUTSCH, D. *Quantum theory, the church –Turing principle and the universal quantum computer*. Proceedings of the Royal Society. A400, 97-117. London, 1985.

ENGEL, G.S.; CALHOUN T.R.; READ E.L.; AHN T.K.; MANCAL, T.; CHENG, Y.C.; BLANKENSHIP, R.E.; FLEMING, G.R. *Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems*. Nature 446:782-786. 2007.

FEYNMAN, R.P. *Quantum mechanical computers*. Foundations of Physics, 16(6), 507-31. 1986.

FRANKS, N.P.; LIEB, W.R. *Seeing the light: protein theories of general anesthesia*. Anesthesiology, Vol:101, Pages:235-237. 2004.

FRÖHLICH, H. *Coherent excitations in active biological systems*. In: Modern Bioelectrochemistry. New York: F. Gutman and H. Keyzer. Springer Verlag, 1986.

HADAMARD, J. *The Mathematician's Mind: The Psychology of Invention in the Mathematical Field*. Princeton University Press. New Jersey, 1996.

HADORN, C.; WEHNER, R. *Zoologia Geral*. Fundação Calouste Gulbenkian, 3<sup>a</sup>. ed. Lisboa, 1978.

HAGAN, S.; HAMEROFF S. R. and TUSZYNSKI, J. A. *Quantum computation in brain microtubules: decoherence and biological feasibility*. Physical Review E. v. 65, 2002.

HAMEROFF, Stuart. *Ultimate computing: biomolecular consciousness and nanotechnology*. Amsterdam: Elsevier North-Holland, 1987.

\_\_\_\_\_. *The 'conscious pilot' - dendritic synchrony moves through the brain to mediate consciousness*. J Biol Phys, 36, 71–93, Springerlink.com, 2010.

HAMEROFF, S.; WATT R.C. *Information processing in microtubules*. Journal of Theoretical Biology, 98, 549-61. 1982.

KAUFMANN Walter; BAIRD, Forrest. *From plato to derrida*. Upper saddle river. Pearson Prentice Hall. New Jersey, 2008.

LEWIS, D. *What Experience teaches*. Mind and Cognition. Oxford: Blackwell. W. Lycan, 1990.

MARSHALL, W.; SIMON, C.; PENROSE, R.; BOUWMEESTER, D. *Towards quantum superpositions of a mirror*. Physical Review Letters 91, 13-16; 130401. 2003.

MCCONKIE, George W.; CURRIE, C.B. *Visual stability across saccades while viewing complex pictures*. Journal of Experimental Psychology: Human Perception & Performance. 22 (3): 563–581. 1996.

NAGEL, T. *What is it like to be a bat?* Philosophical Review, 83: 435-456. 1974.

NEMIROW, L.. *Physicalism and the cognitive role of acquaintance*. Mind and Cognition. Oxford: Blackwell. W. Lycan, 1990.

OUYANG, M.; AWSCHALOM, D. *Coherent spin transfer between molecularly bridged quantum dots*. Science 301:1074-78. 2003.

PENROSE, Roger. *The emperor's new mind*. Concerning Computers, Minds and Laws of Physics. Oxford: University Press, 1989.

\_\_\_\_\_. *Shadows of the mind. A Search for the Missing Science of Consciousness*. Oxford: University Press, 1994.

\_\_\_\_\_. *The Large, the small, and the human mind*. Cambridge: University Press, 1996.

\_\_\_\_\_. *The road to reality: a complete guide to the laws of the universe*. Jonathan Cape. London, 2004.

PENROSE, R.; HAMEROFF, S. *Orchestrated reduction of quantum coherence in brain*. Mathematics and Computers in Simulation, 40, 453-480, Elsevier Science B.V, 1996.

\_\_\_\_\_. *Consciousness in the universe: neuroscience, quantum space-time geometry and orch OR theory*. Journal of Cosmology, v. 14. 2011.

PINGUELLI ROSA, Luiz. *Tecnociências e humanidades*. 2v. Paz e Terra. São Paulo, 2005.



PINGUELLI ROSA, Luiz; FABER, Jean. *Quantum models of mind: are they compatible with environment decoherence?* Physical Review E - Statistical Physics, Plasmas, Fluids and Related Interdisciplinary Topics, v. 70, n.031902, p. 031902, 2004.

POKORNÝ, J. *Excitation of vibration in microtubules in living cells.* Bioelectrochem. 63: 321-326. 2004.

POPPER, K. *Back to the pre-socratics: the presidential address.* Proceedings of the Aristotelian Society, New Series. v. 59. Blackwell Publ, London, 1958-1959.

RASMUSSEN, S.; KARAMPURWALA, H.; VAIDYANATH, R.; JENSEN, K.; HAMEROFF, S. (1990) *Computational connectionism within neurons: a model of cytoskeletal automata subserving neural networks.* Physica D 42:428 49. 1990.

REIMERS, J.R.; MCKEMMISH, L.K.; MCKENZIE, R.H.; MARK, A.E.; HUSH, N.S.; *Weak, strong, and coherent regimes of Fröhlich condensation and their applications to terahertz medicine and quantum consciousness.* Proceedings of the National Academy of Sciences. 106(11):4219-24. USA, 2009

ROBINSON, Thomas. *As Origens da Alma: Os gregos e o conceito de alma de Homero a Aristóteles.* Editora Annablume. São Paulo, 2010.

ROSENTHAL, D. 1986. *Two concepts of consciousness.* Philosophical Studies, 49: 329-59. 1986.

ROUTLEDGE. In: *ENCYCLOPEDIA of philosophy.* Taylor & Francis Group. Londres, 1998.

SEARLE, John. *The mystery of consciousness.* The New York Review of Books. New York, 1997.

\_\_\_\_\_. *Minds, brains and programs.* The Behavioral and Brain Sciences, v. 3 417-457. Cambridge University Press. Cambridge, 1980.

\_\_\_\_\_. *Intentionality: an essay in the philosophy of mind.* Cambridge University Press. 1983.

\_\_\_\_\_. *The rediscovery of the mind*. MIT Press. Massachusetts, 1994.

\_\_\_\_\_. *Biological naturalism*. The Blackwell Companion to Consciousness. Blackwell Publishing Ltd. Massachusetts, 2007.

SHANNON, C.E.. *A mathematical theory of communication*. Bell Systems Technical Journal 27: 379-423. 1948. Reprinted in C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press. 1949.

SHERRINGTON, C.S. *Man on his nature*. 2. Ed. Cambridge University Press. . Cambridge, 1957.

SPINNEY, L. *Blind to change*. New Scientist, n.2265, p. 29–32, November, 2000.

TEGMARK, M. *Importance of quantum decoherence in brain processes*. Phys. Rev. E 61, 4194–4206, 2000.

TITTEL, W.; BRENDDEL, J.; GISIN, B.; HERZOG, T.; ZBINDEN, H.; GISIN, N. *Experimental demonstration of quantum correlations over more than 10 km*. Physical Review A, 57:3229-32. 1998.

TURING, A. *On computable numbers, with an application to the entscheidungsproblem*. Proceedings of the London Mathematical Society, Series 2, vol. 42, 1936; reprinted in M. David (ed.), *The Undecidable*. Raven Press. Hewlett. 1965.

WHITEHEAD, A.N. *The concept of nature*. Cambridge University Press. Cambridge, 1920.

\_\_\_\_\_. *Process and reality*. Macmillan. New York, 1929.

WOLFRAM, S. *A New Kind of Science*. Wolfram Media incorporated. 2002. Disponível em <<http://www.wolframscience.com/nksonline/toc.html>>. Acesso em out.2011.

Revista Científica:

Revista Brasileira de Ensino de Física *Print version* ISSN 1806-1117. Rev. Bras. Ensino Fís. vol.25 no.3 São Paulo, Set. 2003.

Sites:

STANFORD Encyclopedia of Philosophy; *Consciousness*. Stanford: 2004. Disponível em: <<http://plato.stanford.edu/entries/consciousness/>>. Acesso em 02 maio 2011.

ROBINSON, Daniel Sommer. *Idealism*, Encyclopædia Britannica, <<http://www.britannica.com/EBchecked/topic/281802/idealism>>. Acesso em 12 dez. 2011.