

**UNIVERSIDADE ESTADUAL PAULISTA – UNESP
FACULDADE DE FILOSOFIA E CIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA
CAMPUS DE MARÍLIA**

**Um Contraste entre as Teorias Cognitivas da
Consciência de Baars e Dennett: O Espaço de
Trabalho Global seria um Teatro Cartesiano?**

Samuel de Castro Bellini-Leite

**Marília
2013**

**UNIVERSIDADE ESTADUAL PAULISTA – UNESP
FACULDADE DE FILOSOFIA E CIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA
CAMPUS DE MARÍLIA**

Um Contraste entre as Teorias Cognitivas da Consciência de Baars e Dennett: O Espaço de Trabalho Global seria um Teatro Cartesiano?

Samuel de Castro Bellini-Leite

Dissertação apresentada ao Programa de Pós-Graduação em Filosofia da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP) para Defesa de Mestrado na área de concentração em Filosofia da Mente, Epistemologia e Lógica.

Orientador: Dr. Alfredo Pereira Júnior

**Marília
2013**

Bellini-Leite, Samuel de Castro.

B444c Um contraste entre as teorias cognitivas da consciência de Baars e Dennett: o espaço de trabalho global seria um teatro cartesiano? / Samuel de Castro Bellini-Leite. - Marília, 2013.

124 f. ; 30 cm.

Dissertação (Mestrado em Filosofia) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2013.

Bibliografia: f. 119-124. Orientador:

Alfredo Pereira Júnior.

1. Consciência. 2. Evolução. 3. Cognição. 4. Filosofia da mente. 5. Baars, Bernard J. A cognitive theory of consciousness.

6. Dennett, C. D. (Daniel Clement). Consciousness explained. I. Título.

CDD 126

Samuel de Castro Bellini-Leite

**Um Contraste entre as Teorias Cognitivas da
Consciência de Baars e Dennett: O Espaço de
Trabalho Global seria um Teatro Cartesiano?**

Banca Examinadora (Qualificação)

Prof. Dr. Alfredo Pereira Júnior (Unesp/Botucatu)
(Orientador)

Prof. Dr. Leonardo Ferreira Almada (UFU)
(1º Examinador)

Profª. Drª. Maria Eunice Quilici Gonzalez (Unesp/Marília)
(2º Examinador)

Profª. Drª. Mariana Cláudia Broens (Unesp/Marília)
(suplente)

Banca Examinadora (Defesa)

Prof. Dr. Alfredo Pereira Júnior (Unesp/Botucatu)
(Orientador)

Prof. Dr. Leonardo Ferreira Almada (UFU)
(1º Examinador)

Profª. Drª. Maria Eunice Quilici Gonzalez (Unesp/Marília)
(2º Examinador)

Profª. Drª. Mariana Cláudia Broens (Unesp/Marília)
(suplente)

Prof. Fabio Augusto Furlan (Unimar)
(2º suplente)

Marília, 18 de fevereiro de 2013

*As iniciativas de estudo
em Cif?ncia Cognitiva no Brasil.*

Agradecimentos

Ao meu orientador Alfredo Pereira Júnior pelo apoio durante a realização do trabalho.

Aos professores da UNESP por terem me aceitado com disposição inacreditável.

Ao grupo GAEC pelas discussões durante a realização desta dissertação.

Aos meus pais por sempre incentivarem meus estudos.

À Marina e João Felipe por terem me esperado por quase dois anos com paciência, respeito e amor.

À CAPES pelo auxílio financeiro prestado durante as minhas pesquisas, sem o qual não seria possível ter chegado aos resultados aqui apresentados.

*“Hold on tight beginner's mind
The current is far too strong
It'll carry you along
Till you're just like everyone
Stay a while my inner child
I'd like to learn your trick
To know what makes you tick
To nurse you when you're sick”
(Conor Oberst)*

SUMÁRIO

Introdução	10
Capítulo 1 – A Evolução da Consciência	18
Introdução.....	19
1.1 Seleção Natural, Auto-Organização e Efeito Baldwin.....	20
1.2 A Evolução da Consciência Segundo Dennett.....	30
1.3 Funções Adaptativas da Consciência Segundo Baars.....	41
1.4 Consciência em Animais.....	43
1.5 Considerações Sobre o Capítulo.....	48
Capítulo 2 – Espaço de Trabalho Global e Esboços Múltiplos	51
Introdução.....	52
2.1 Noções Elementares da Teoria do Espaço de Trabalho Global	53
2.2 A Teoria do Espaço de Trabalho Global.....	63
2.3 Consciência e Informação.....	70
2.4 Considerações Sobre A Teoria do Espaço de Trabalho Global.....	72
2.5 A Heterofenomenologia.....	73
2.6 O Ponto de Vista do Observador e o Teatro Cartesiano.....	77
2.7 Esboços Múltiplos.....	79
2.8 Máquina Joyceana.....	82
2.9 Considerações Sobre a Teoria de Dennett.....	86
Capítulo 3 – O Teatro Cartesiano	89
3.1 Um Contraste entre a Teoria do Espaço de Trabalho Global e o Modelo de Esboços Múltiplos.....	90
3.2 O Conceito do Teatro Cartesiano.....	93
3.3 A Teoria do Espaço de Trabalho Global é um Teatro Cartesiano?.....	99
3.4 Há Realmente um Problema em Postular um Teatro Cartesiano?.....	106
3.5 Considerações Sobre o Capítulo.....	112
Considerações Finais	114
Referências	119

RESUMO

Este trabalho tem como objetivo geral realizar um contraste entre duas teorias cognitivas da consciência, a Teoria do Espaço de Trabalho Global de Bernard Baars (1988) e o Modelo de Esboços Múltiplos de Daniel Dennett (1991). Apesar de Dennett demonstrar apreciação pela Teoria do Espaço de Trabalho Global, sua teoria não aparenta ser muito compatível com a mesma. O objetivo específico deste trabalho é de verificar esta compatibilidade perguntando se as críticas de Dennett ao Teatro Cartesiano atingem a Teoria do Espaço de Trabalho Global. O primeiro capítulo expõe a visão de Dennett sobre como a consciência evoluiu, em contraste com a visão baseada na Teoria do Espaço de Trabalho Global. Para a primeira, a cultura e a linguagem possuem um papel central na origem da consciência, para a segunda a consciência tem principalmente uma origem biológica. No segundo capítulo, ambas as teorias da consciência são expostas, comentadas e criticadas. O terceiro capítulo realiza um contraste entre as duas teorias e analisa as implicações do conceito de Teatro Cartesiano para a Teoria do Espaço de Trabalho Global. Argumentamos que o conceito do Teatro Cartesiano é vago, e através de uma análise cautelosa encontramos 10 requisitos para uma teoria não se enquadrar em um Teatro Cartesiano, através das palavras de Dennett. Verificamos que a Teoria do Espaço de Trabalho Global preenche alguns desses requisitos. Por fim, através de uma exploração das análises de Todd (2009), concluímos que as críticas principais de Dennett a alguns aspectos do Teatro Cartesiano são fracas. Dessa forma, as críticas ao Teatro Cartesiano não são uma ameaça para a Teoria do Espaço de Trabalho Global.

Palavras-chave: Consciência; Teatro Cartesiano; Esboços Múltiplos; Teoria do Espaço de Trabalho Global.

ABSTRACT

This work has as its main goal a contrast of two renowned cognitive theories of consciousness, Bernard Baars' (1988) Global Workspace Theory and Daniel Dennett's (1991) Multiple Drafts. Although Dennett shows some appreciation to the Global Workspace Theory, his own Multiple Drafts Model does not seem very compatible with it. The specific goal of this work is to verify such compatibility by asking if the Global Workspace Theory suffers from Daniel Dennett's criticism of the Cartesian Theater. The first chapter exposes Dennett's perspective on the evolution of consciousness, in contrast to the view based on The Global Workspace Theory. The former understands that language and culture play a central role in the origin of consciousness; the latter understands consciousness has mainly a biological origin. In the second chapter, both theories of consciousness are exposed and reviewed. The third chapter focuses on a contrast of the two theories and some implications of the Cartesian Theater. Also, we noted that the concept of a Cartesian Theater is vague, and through a rigorous analysis, 10 requisites for a theory to evade the Cartesian Theater, following Dennett's words, were identified. The Global Workspace Theory was shown to meet a few of these requisites. Finally, making use of Todd's (2009) analyses, we concluded that Dennett's main critics to some aspects of the Cartesian Theater are weak. So it follows that the criticism to the Cartesian Theater does not pose serious problems for the Global Workspace Theory.

Key-words: Consciousness; Cartesian Theater; Multiple Drafts; Global Workspace Theory.

INTRODUÇÃO

O psicólogo Steven Pinker em seu livro “*How the Mind Works*” (1997) sugere que na ciência existe uma diferença entre problemas e mistérios. Alguns problemas não possuem respostas imediatas, mas existem métodos que possibilitam algumas possíveis respostas, e é possível prever como essas respostas serão. Já os mistérios não possuem resposta e nem se imagina como chegar a uma resposta, ou até mesmo qual forma esta resposta tomaria.

A consciência sempre foi um fenômeno misterioso para o entendimento do senso comum, da filosofia e da ciência. Pela tradição filosófica cartesiana, chamada por Ryle (1949) de “a doutrina oficial”, a consciência é uma característica de seres dotados de alma, tendo relação física apenas com a glândula pineal. Sendo a consciência localizada na alma, definida como algo imaterial, um dos principais métodos de acesso à consciência na filosofia e na psicologia era o da introspecção. Basicamente, através do pensamento sobre os fenômenos mentais individuais, em primeira pessoa, tentava-se entender a consciência.

Os problemas que vinham sendo trazidos e nunca resolvidos pelo método introspectivo e pela perspectiva dualista cartesiana fizeram a consciência sair de foco nas discussões da psicologia; o Behaviorismo era mais promissor por apresentar métodos científicos para o estudo do comportamento sem considerar fenômenos mentais (ATKINSON *et. al.*, 2002). Entretanto, desde o final da década de 40 e durante a de 50, nas áreas da Matemática, Computação, Linguística, Antropologia e Psicologia, pesquisadores em conjunto elaboraram a “Revolução Cognitiva”. Um dos principais avanços desta revolução em relação ao Behaviorismo consistiu em privilegiar o estudo do processo que ocorre entre o estímulo e a resposta de organismos no mundo. Este processo é chamado de cognição. Para Neisser (1967), a cognição é o processo pelo qual a informação sensorial recebida é transformada, reduzida, elaborada, armazenada, recuperada e usada, sendo fundamental para qualquer realização humana.

Essa revolução deu início à chamada Ciência Cognitiva, uma ciência que, em sua vertente mais tradicional, realiza um estudo dos fenômenos mentais por meio de uma hipótese central: a mente pode ser considerada um programa computacional por ter como função processar informação. A Ciência Cognitiva tradicional possui as seguintes

características essenciais: representação; computação; pouca ênfase no afeto, contexto, cultura e história; interdisciplinaridade; e ter raízes em problemas filosóficos clássicos.

Para explicar a ação humana e o pensamento, os cognitivistas postulam um nível das representações entre a entrada (*input*) e a saída (*output*), no qual entidades como símbolos, regras, imagens e esquemas são relacionadas, unidas ou transformadas. De acordo com Eysenck e Keane (2000), uma representação é qualquer sinal ou conjunto de símbolos que representa algo para alguém. Estas podem ser divididas primeiramente entre externas e internas. A primeira sendo qualquer representação feita em substrato material externo, como pintura, desenhos, linguagem escrita, mapas etc. A segunda diz respeito a representações mentais, internas. As representações externas podem ser divididas entre representações de imagem e de linguagem. A representação linguística é separada em diversos símbolos, possuindo símbolos explícitos para o que representam; são organizadas por uma série de regras, são abstratas e arbitrariamente relacionadas ao significado (por exemplo, a palavra maçã em nada se relaciona com a fruta). As representações de imagens não são separáveis em unidades menores, não possuem um símbolo distinto para tudo que representam, por exemplo, não há um símbolo específico para significar que algo está “em baixo”. A imagem é mais concreta do que abstrata (EYSENCK; KEANE, 2000).

Uma representação mental, ou interna, é uma estrutura com conteúdo semântico, sobre o mundo exterior e/ou interior ao sistema cognitivo. As representações internas, ou mentais, em geral possuem quatro aspectos essenciais: são realizadas por alguém que a contém; possuem conteúdo ou representam um ou mais objetos; sua relação representacional é razoavelmente estável; e podem ser interpretadas por ou servir de representação para outras pessoas. Na ciência cognitiva, os portadores da representação mental são estados computacionais (VON ECKARDT, 1999).

De acordo com Paivio (1990), o termo “representação” difere do conceito de processo. Um modo claro de diferenciar estes é através de uma analogia com a noção de estrutura e função. A representação é uma entidade estrutural nas quais os processos podem operar; os processos são as atividades envolvidas no uso funcional da informação.

As representações podem ser divididas entre simbólicas e distribuídas (EYSENCK; KEANE, 2000). A representação distribuída é adotada por uma corrente de pensamento intitulada Conexionismo. Trata-se de uma forma de representar informação por meio da ativação de um grupo de unidades, no qual cada conceito é

representado por um padrão de conexão das unidades, sendo que cada uma pode participar da representação de diversos outros conceitos.

Steven Pinker (1997) argumenta que redes conexionistas por si só não explicam a cognição. O psicólogo acredita ser a estruturação de redes em programas que manipulam símbolos a explicação para a inteligência humana. Tal manipulação de símbolos é necessária para a linguagem humana e para partes do raciocínio que interagem com esta. Esta união entre processamentos em paralelo e serial é adotada pela vários cognitivistas, sendo essencial para a teoria de consciência de Dennett (1991).

Inicialmente, a tradição da representação simbólica e serial não considerava a forma de processamento dos neurônios, pois assumiam que os modelos teóricos criados a partir de experimentos, aplicados a programas computacionais, poderiam ser o bastante para explicar a representação (EYSENCK; KEANE, 2000). Atualmente, a maioria dos cientistas acredita que a relação da teoria funcional com o cérebro é um ganho eficiente para o modelo.

As raízes da teoria da representação simbólica estão vinculadas à hipótese do sistema físico de símbolos. A hipótese do sistema físico de símbolos foi criada por Newell e Simon (1976). Estes cientistas acreditam que este sistema físico de símbolos por si só possui condição suficiente e necessária para ter ações inteligentes. O sistema está submetido às leis da física e opera em padrões físicos. Esses padrões podem ser fisicamente relacionados a ponto de criar expressões, que são estruturas de símbolos. O próprio sistema é composto por essas estruturas e de uma gama de processos que operam nas estruturas para produzir outras.

Assim as teorias de representação simbólica assumem que a cognição é um sistema físico de símbolos, como proposto por Newell e Simon. Tal sistema é dependente no sentido de que é aberto a informações do meio, mas autônomo, pois sua arquitetura principal não pode ser reorganizada. A informação do meio é o material que o sistema usa para operar, mas as mudanças de arquitetura não são determinadas pelo meio. Dessa forma, os eventos mentais são separáveis do conteúdo do mundo externo, tornando possível a criação de programas de computador que simulem tais eventos de forma satisfatória (DERRY; STEINKUEHLER, 2005).

A simulação no computador, para a Ciência Cognitiva, serve como uma prova por existência. Se uma máquina pode raciocinar, ter metas, conferir o comportamento e transformar informação, então os seres humanos podem ser caracterizados da mesma

forma. A possibilidade de computadores realizarem atividades humanas fortalece a ideia de estes últimos serem também máquinas.

A ideia de computabilidade foi apresentada de forma rigorosa pelo matemático Alan Turing em 1936. Uma computabilidade é um procedimento que pode ser realizado por uma Máquina de Turing. Uma Máquina de Turing é uma máquina abstrata composta de três partes: uma fita infinita dividida em quadrados nos quais símbolos podem ser escritos; um escâner que analisa um quadrado de cada vez; e um quadro de leis especificando o que a máquina deve fazer. A máquina pode realizar quatro ações: mover para próximo quadrado, mover para o quadrado anterior, trocar um símbolo ou apagar um símbolo. Essas ações são descritas de forma serial no quadro de leis (BODEN, 2006). Em sua implementação física (o computador), o quadro de leis é o programa escrito pelo programador, os símbolos são o zero ou um, os quais representam a liberação ou não liberação de energia. Cada combinação de zero ou um forma um padrão de liberação de energia único, que é relacionado com a função que o computador deve exercer. Para um exemplo prático, uma imagem na tela do computador é formada por um conjunto de píxeis em certa intensidade. Um código “010101110” é uma instrução única de liberação de energia para iluminar de uma determinada forma cada pixel da tela; juntos estes píxeis formam a imagem desejada.

Uma das ideias centrais da computação na Ciência Cognitiva é que nossos mecanismos mentais funcionam como processos computacionais. Assim, parte da metodologia desta ciência é de descrever o funcionamento mental em processos seriais, como um programa, para que este possa ser implementado em um computador e testado (GREEN, 1996). Grosso modo, se o programa produzir respostas semelhantes àsquelas da mente humana, pode-se dizer que ele é um bom modelo da mente humana. É possível estabelecer uma crítica a esta metodologia por existirem diversas formas de obter o mesmo resultado em uma computação, mesmo com procedimentos diferentes. Assim, a mente poderia estar realizando um procedimento e os modelos computacionais realizando outro completamente distinto. Porém, o papel atribuído à Psicologia Cognitiva, Neurociência e Neurociência Computacional é justamente o de conferir como o procedimento do cérebro/mente realmente ocorre para que os modelos sejam cada vez mais fidedignos.

A própria definição de computabilidade remete à noção da Máquina de Turing. Apesar disto, atualmente acredita-se que o cérebro não é literalmente uma Máquina de Turing. Alguns, como Dennett (1991), acreditam que o cérebro instancia uma Máquina

de Turing virtual, ou que alguns “programas” da mente funcionam como uma Máquina de Turing. Mas o próprio cérebro, por sua própria estrutura e modo de funcionamento dificilmente seria uma Máquina de Turing. Entretanto, acredita-se que o cérebro seja outro tipo de computador, o qual realiza procedimentos que cientistas da computação ainda não conseguiram replicar ou entender.

Cientistas da cognição não negam a importância do afeto, contexto, cultura e história para o comportamento humano. Entretanto, tradicionalmente, por questões de controle e pesquisa em laboratório, os pesquisadores tendiam a excluir esses fatores ou procurar processos nos quais estes interferem menos. Por causa de suas relações com todos os tipos de processos cognitivos, as emoções ganharam um foco maior no estudo da cognição desde o final do século XX (DAMASIO, 1994). Uma ideia recente de como a emoção e cognição podem se relacionar é proposta por Pereira Jr. e Furlan (2010) e Pereira Jr. e Almada (2011).

Atualmente os modelos da Ciência Cognitiva tendem a trabalhar em três níveis de análise: o comportamental (o resultado em termos de ação), o cognitivo (sistemas de processamento de informação) e o biológico (os sistemas cerebrais e suas ligações com fatores genéticos e ambientais) (GREEN, 1996). O cérebro é cada vez mais aceito como uma parte importante da explicação dos procedimentos mentais, assim a Ciência Cognitiva está cada vez mais se transformando em uma Neurociência Cognitiva.

Green (1996) define a Ciência Cognitiva como uma ciência interdisciplinar que busca entender como a mente funciona em termos de processos operando em representações. Boden (2006) a define simplesmente como o estudo interdisciplinar da mente, informada por conceitos teóricos vindos da ciência da computação.

Com base na Ciência Cognitiva e também em teorias evolutivas, Pinker (1997) afirma que a mente deixou de ser um mistério, sendo elevada ao patamar de problema. Entretanto, Baars (1988) menciona que apesar da Ciência Cognitiva ter aceitado o desafio de compreender a mente, a consciência demorou a ser considerada. Os modelos tradicionais de funcionamento da mente se especializavam por funções, como memória, raciocínio, linguagem, e deixavam para depois o papel da consciência. Até mesmo arquiteturas cognitivas elaboradas com o objetivo de unir os diferentes setores de processamento, como a ACT-R de Anderson (1983), deixavam a consciência fora da explicação. Por vezes, os fluxogramas desses modelos apresentavam funções de execução central, mas estas supunham a presença de um homúnculo para apreender as informações processadas, pela não explicação dos procedimentos realizados por este

executivo central (DENNETT, 1991). Apenas no final da década de 80 a consciência realmente começou a tomar foco nas discussões da Ciência Cognitiva, e no fim do século XX diversos modelos biológicos e funcionais foram propostos para explicar como a consciência funciona (BAARS, 1988; DENNETT, 1991; EDELMAN, 1997; CRICK, KOCH, 1990).

Através da mudança da pergunta de “o que é a consciência” para “como a consciência funciona em termos biológicos e de processamento de informação” a consciência, por um lado, também deixou de ser um mistério para se tornar um problema. Por outro lado, a antiga pergunta “o que é a consciência”, a qual remete a problemas metafísicos, dificilmente deixou de ser um mistério. Apesar de ainda ser um mistério, existem diversas correntes filosóficas com objetivo de propor uma resposta satisfatória. A corrente filosófica mais ligada à Ciência Cognitiva é o Funcionalismo.

De acordo com Maslin (2001), o Funcionalismo rejeita a noção da mente enquanto uma entidade, seja a alma ou o cérebro. Esta doutrina caracteriza a mente como um conjunto de funções. A função de um objeto é o que ele faz. Em termos matemáticos, uma função específica transforma um *input* em um *output* seguindo sempre a mesma regra. Considere a equação $y = x^2$. O *input* é a variável x , a função é multiplicar a variável por si mesma, e o *output* será o y . Em relação ao computador, o grupo de instruções que formam um programa são as funções sendo exercidas pela máquina.

Um guarda-chuva pode ser feito de diversos materiais, pode vir em diversos modelos, ter tamanho e espessura variável. Esses objetos são extremamente diferentes entre si, apenas um fator permanece constante, ele permite o bloqueio de água, esta é sua função. De acordo com Maslin (2001), há claramente uma distinção entre o grupo de arranjos que permitem um objeto exercer uma função, e a função em si. Segue-se que uma mesma função pode ser manifestada por diversos tipos de arranjos materiais. Esta é uma propriedade do Funcionalismo chamada de Múltipla Realizabilidade. Uma função não é material e nem imaterial, pois não se trata de uma entidade, e sim um processo exercido por entidades físicas.

O Funcionalismo de Dennett (1991) tem duas características específicas. De acordo com Lycan (1990), podemos chamar esta doutrina de Funcionalismo Homuncular teleológico. Homuncular pois a explicação das funções é atribuída a micro-agentes cada vez menores. Gilbert Ryle critica teorias que postulam homúnculos para explicação de certas funções, argumentando que a não explicação do homúnculo faria a

explicação recair em um regresso ao infinito. Em outras palavras, não se explica o problema, apenas se postula uma entidade misteriosa para tomar lugar da função. O Funcionalismo Homuncular de Dennett pretende fugir desta crítica, pois se as funções são explicadas em termos de grupos de homúnculos cada vez menores, deixamos de atribuir a explicação a um ser misterioso. A explicação se iguala, portanto, à metodologia reducionista da ciência, e se fosse possível explicar todos os mínimos detalhes, passaríamos de uma explicação funcional para uma explicação biológica, e da explicação biológica para a física. O Funcionalismo de Dennett é também chamado de teleológico, pela razão de que sua explicação do propósito das funções remete à sua utilidade para a evolução biológica. Por este tipo de explicação, uma característica pode ter a função de fazer “X” por ter sido selecionada por ter feito “X”.

Por dizer que a consciência é constituída apenas de funções de processamento de informação, a estratégia principal do Funcionalismo é da anulação da primeira pergunta sobre a consciência (o que é a consciência), para focar a segunda (como a consciência funciona em termos de processamento de informação). Ou seja, se a consciência é um conjunto de funções de processar informação, será preciso apenas descrever qual é o conjunto de funções na mente que lidam com a consciência para termos uma ideia completa do que é a consciência.

Este trabalho analisa teorias empíricas da consciência na dimensão meta-teórica. O objetivo geral é realizar um contraste entre duas teorias cognitivas da consciência que seguem o funcionalismo, a Teoria do Espaço de Trabalho Global de Bernard Baars (1988) e o Modelo de Esboços Múltiplos de Daniel Dennett (1991). Apesar de Dennett demonstrar apreciação pela Teoria do Espaço de Trabalho Global, sua teoria não aparenta ser muito compatível com a mesma. O objetivo específico deste trabalho é de verificar esta compatibilidade perguntando se as críticas de Dennett ao Teatro Cartesiano atingem a Teoria do Espaço de Trabalho Global.

No primeiro capítulo, a origem evolutiva da consciência, de acordo com os dois modelos teóricos, é discutida. Primeiramente são descritas três forças atuando na evolução, a seleção natural, a auto-organização e o Efeito Baldwin. Em sequência é descrito como Dennett entende a origem da consciência a partir da instalação de memes. Também apresentamos uma visão da evolução coerente com a Teoria do Espaço de Trabalho Global. Por fim, realizamos um comentário sobre as duas teorias da evolução da consciência.

No segundo capítulo, ambas as teorias da consciência são expostas, comentadas e criticadas. A exposição visa mostrar o método utilizado, os principais argumentos e evidências a favor das duas teorias. Baars afirma que a consciência pode ser entendida como um sistema funcional chamado de espaço de trabalho global. Dennett argumenta que a consciência é uma máquina virtual que faz releituras sobre os eventos em esboços múltiplos que ocorrem no cérebro.

O terceiro capítulo realiza um contraste entre as duas teorias e analisa as implicações do conceito do Teatro Cartesiano para a Teoria do Espaço de Trabalho Global. Logo, uma análise mais profunda do conceito do Teatro Cartesiano na obra de Dennett é realizada. A partir desta análise, tentamos identificar se a teoria de Baars segue um modelo baseado no Teatro Cartesiano. Questionamos, ainda, as próprias críticas de Dennett, tentando verificar se realmente existe um problema em modelos da consciência baseado no Teatro Cartesiano.

CAPÍTULO 1 – A Evolução da Consciência

“Nada no estudo da mente faz sentido
se não à luz da evolução.”
Adaptado de Dobzhansky (1973)

Introdução

O biólogo Theodosious Dobzhansky publicou em 1973 um artigo intitulado “*Nothing in Biology makes sense except in the light of evolution*” (Nada na biologia faz sentido senão à luz da evolução). Em 1983, o biólogo Richard Dawkins criou o termo “Universal Darwinism” (1983) para se referir à tendência e necessidade de estender os conceitos darwinistas para outros campos de estudos que não a biologia. A Psicologia Evolucionista e Neurociência Evolutiva são duas disciplinas que resultaram do Darwinismo Universal, assim como a Filosofia da Mente de Dennett. Poderíamos argumentar atualmente, em analogia à Dobzhansky, que nada no estudo da mente faz sentido se não à luz da evolução. Seguindo esta tradição (como fazem Dennett e Baars), a consciência precisa ser entendida à luz da evolução.

De acordo com Dennett (1995), seguindo a forma Darwinista de explicação, saímos de um momento onde não havia “x”, a partir de gradativas mudanças, e incorporações passamos por momentos nos quais é difícil dizer se há ou não “x” e chegamos a momentos em que claramente existe “x”. Acreditamos que esta forma de descrever fenômenos seja contra-intuitiva devido à falta de correspondência direta de nossos conceitos e categorias com o mundo. Quando dizemos que um animal é um lobo, por exemplo, estamos dizendo que existe um grupo de características de certos organismos que podemos classificar como “lobo”, e que o indivíduo em questão possui esse grupo de características. Mas, se voltarmos em sua história evolutiva, teremos indivíduos semelhantes que não saberemos dizer se são ou não lobos. Isto porque “lobo” é um conceito (ou categoria) inventado para nos referirmos a um conjunto de animais, mas cada indivíduo é único, e alguns podem não se encaixar no nosso conceito. Ainda, se considerarmos que membros da mesma espécie são aqueles que podem se cruzar e gerar filhos férteis, então não se pode dizer que um pastor alemão e um poodle sejam da mesma espécie, pois o acasalamento dos dois não gera filhos férteis e, ainda, um pastor alemão gera filhos férteis com alguns lobos enquanto poodles não os geram. Poderíamos até mesmo utilizar métodos mais precisos para dizer o que é uma espécie, como a partir de análises de DNA; conseguiríamos assim fazer uma classificação mais precisa, mas nunca chegaríamos a categorias que sejam mutuamente exclusivas, isso porque cada indivíduo é único. Uma criança aprendendo a falar chama todas as aves de “passarinho”, seus conceitos são imprecisos, mas também os conceitos filosóficos e

científicos ainda não chegam à precisão absoluta, pois para isso teríamos que ter um nome diferente para cada organismo (*token*) no mundo.

Dennett (1995, p.201) afirma que nada complicado o bastante para ser interessante pode ter uma essência. O filósofo usa o exemplo da vida. Poderíamos voltar em uma máquina do tempo e tentar achar um momento onde a vida começou, mas seria difícil dizer o que é vida neste momento. O mesmo pode ser dito do surgimento da agência (ações intencionais, reflexão, representações de razões, decisão), a qual é realizada por milhares de “micro-robôs” com “quasi-agência”. De acordo com Dennett, Darwin acreditava que essas características de perda de essência e foco em passos graduais geram uma revolução metafísica e epistemológica que surgiu junto com sua ciência. Desde Platão e Sócrates consideramos essencial definir nossos termos para compreender o mundo, mas Dennett (1995, p.202) afirma que isso é o que *nós* precisamos fazer, o *mundo* em si não possui essências que correspondem aos nossos termos.

Neste capítulo, primeiramente serão apresentadas hipóteses sobre como a evolução ocorre. Em seguida, serão discutidas as visões da evolução da consciência de acordo com Dennett (1991), e posteriormente, de acordo com Baars (1988), Seth, Baars & Edelman (2005) e Edelman & Seth (2009).

1.1 Seleção Natural, Auto-Organização e Efeito Baldwin

Antes de Darwin, um dos maiores mistérios a serem desvendados era o surgimento da complexidade das espécies. Richard Dawkins é um biólogo que trabalha bastante com o evolucionismo e interpreta as obras de Darwin. Um de seus livros mais famosos é “O Gene Egoísta” (1976), no qual ele descreve sua interpretação do Darwinismo enquanto “selecionismo de genes”. Esta interpretação tem um foco forte no papel dos genes, porém descobertas recentes na biologia sobre o papel de reguladores de genes e partes do DNA que não estão ligadas à produção de proteínas (HAYDEN, 2010) podem trazer problemas para a interpretação de Dawkins (a qual é em grande parte seguida por Dennett). De toda forma, em seu livro “O Relojoeiro Cego” (2001 [1986]), Dawkins descreve, os funcionamentos da seleção natural como descrita por Darwin. A partir deste livro é possível compreender o que é a seleção natural, para não nos confundirmos com mitos que cercam esta teoria. Dennett começou a se interessar pela seleção natural a partir de leituras de Dawkins e isso influenciou toda sua filosofia.

Por isso, será interessante utilizar a explicação do especialista para compreendermos a ideia fundamental que da base às defesas das ideias de Dennett.

A evolução é um fato comprovado por evidências paleontológicas e da biologia comparativa (RUSE, 1995). A forma mais clara de como a evolução pode ocorrer é através da seleção natural, da forma descrita no livro de Darwin “A Origem das Espécies” (1859). Um dos maiores mitos que Dawkins (1986) pretende desconstruir é a noção de que a seleção natural é uma teoria do acaso. De acordo com esse mito, a seleção natural ocorreria com o tempo e por diversas tentativas, por acontecimentos aleatórios a complexidade iria aos poucos surgindo. Porém uma coisa complexa é algo que não ficamos satisfeitos em dizer que surgiu de aleatoriedades, por ser muito improvável. Podem existir diversas formas de estar vivo, mas com certeza existem mais formas ainda de não estar vivo; é muito mais provável que a aleatoriedade absoluta gere o caos do que a ordem. Podemos combinar células aleatoriamente, por diversas vezes, por bilhões de anos e nenhuma vez chegar a um aglomerado de células que voe, nade, corra ou que sobreviva e evolua.

Em lugar do acaso, Dawkins (1986) explica a existência da complexidade dos seres vivos em consequência de transformações graduais e cumulativas de objetos primordiais simples, por meio de mecanismos seletivos. Essas transformações, por gerarem elementos complexos, não podem ser transformações aleatórias. Dawkins (2001 [1986], p.43) usa um exemplo para ilustrar um sistema que automaticamente gera não-aleatoriedade:

Caminhando por uma praia pedregosa, podemos notar que as pedras não estão dispostas aleatoriamente. As menores tendem a ser encontradas em zonas separadas, acompanhando a linha da praia, e as maiores em zonas ou faixas diferentes. Essas pedras foram classificadas, organizadas, selecionadas. Uma tribo que habita o litoral poderia refletir sobre esse indício de classificação ou organização no mundo e desenvolver um mito para explicá-la; talvez a atribuíssem a um Grande Espírito celeste metódico e organizado. É possível que sorríssemos com superioridade diante dessa ideia supersticiosa e explicássemos que a disposição das pedras na verdade foi produto das forças cegas da física - neste caso, da ação das ondas. As ondas não têm propósitos nem intenções, não têm mente metódica, não têm mente nenhuma. Simplesmente jogam as pedras com força na praia, e, como pedras grandes e pedras pequenas sofrem efeitos diferentes com esse tratamento, acabam parando em níveis diferentes na areia. Um pouquinho de ordem surgiu da desordem sem que alguma mente houvesse planejado esse resultado.

Outro exemplo mais simples é um buraco. Para passar por um buraco necessariamente é preciso ser menor que ele. Se colocarmos tudo o que encontrarmos

neste buraco, eventualmente teremos selecionado coisas que cabem e que não cabem no buraco. Este é um processo de geração de ordem simples. Dawkins (1986) o chama de seleção de um passo. Este processo apenas não é o suficiente para gerar organismos vivos. Para isso precisamos de um segundo processo, chamado de seleção cumulativa. A diferença do segundo processo é o acréscimo da reprodução. Assim, as entidades são selecionadas por inúmeras gerações, e o produto final (genético) de uma geração é o início de outra.

Dawkins (1986) usa de uma ilustração conhecida para explicar a diferença de potencial entre as duas formas de seleção, a qual diz que: com tempo suficiente um macaco batendo aleatoriamente em uma máquina de escrever poderia produzir todas as obras de Shakespeare. Esse tempo provavelmente teria que ser maior do que a idade do universo. Mesmo se reduzíssemos o trabalho do macaco para uma frase de 28 caracteres de um livro de Shakespeare e fosse utilizada a seleção de um passo, o resultado não melhora muito. Sentenças aleatórias seriam escolhidas, mas não continuariam se aperfeiçoando. Um tempo maior que a idade do universo também seria necessário. Dawkins pede para considerarmos como seria a mesma tarefa, mas seguindo a seleção cumulativa. O biólogo resolveu colocar o teste em prática, mas ao invés de utilizar um macaco real, fez um programa de computador para começar com uma frase de 28 caracteres aleatória. Desta vez, na segunda tentativa ele duplicava a primeira tentativa aleatória, mas com certa chance de mudança aleatória (mutação) no processo de cópia. O programa compara as duas com a frase original do livro e mantém a frase mais parecida, mesmo que seja por uma letra. Por cada geração esse processo simples é repetido. O programa que Dawkins desenvolveu demora cerca de 11 segundos para chegar a uma frase do livro original do Shakespeare a partir de seleção cumulativa. Em média, apenas 50 gerações seriam necessárias para replicar a frase.

A peça fundamental desta técnica é que cada aperfeiçoamento, mesmo parecendo insignificante, é usado como base para as próximas gerações. Como exemplificado por programas de computadores, esse processo simples, dado tempo suficiente, ganha uma proporção espetacular e gera complexidade e proficiência. Assim, as características da seleção cumulativa são justamente a não-aleatoriedade, a força da cumulação e a pressão seletiva. O Darwinismo está ligado à noção da seleção cumulativa e, portanto, à não-aleatoriedade (DAWKINS, 1986).

Dawkins (1986) lembra que uma diferença importante dos processos descritos de seleção cumulativa para a seleção natural é que no primeiro caso as gerações eram

selecionadas de acordo com um ideal distante (este é um dos erros do design inteligente, hipótese segundo a qual a seleção segue um plano ideal realizado por Deus) (DAVIS; KENYON, 1989). A vida não é assim, a evolução não tem uma meta a longo prazo. Na realidade o critério usado para seleção é sempre de curto prazo, como sobrevivência ou sucesso reprodutivo. Esta versão mais precisa da seleção natural também poderia ser simulada em computadores. Para isso seria essencial simular a morte não-aleatória, e os programas deveriam interagir no computador dentro de uma simulação de um ambiente hostil.

Uma suposição central do Darwinismo é que organismos não passam a informação que adquiriram durante a vida para seus descendentes. Esse pressuposto foi confirmado através da descoberta do princípio da barreira de Weismann: a informação hereditária passa apenas dos genes para as células do corpo, nunca ao contrário. A hereditariedade seria, portanto, apenas a cópia dos genes. Porém, existem reguladores que possuem um papel fundamental na decisão sobre como os genes irão afetar o corpo. Tradicionalmente considera-se que essa regulação se reinicia nos descendentes, porém houve evidências de herança epigenética por gerações (SCHMITZ *et al.*, 2011). De toda forma, esses estudos ainda são recentes e as evidências são poucas e principalmente em plantas. Se mais evidências forem encontradas, será possivelmente descoberto um novo mecanismo de aceleração da evolução que deverá entrar na explicação da evolução biológica.

Mesmo que o aprendido não seja transmitido para os descendentes, a seleção natural privilegia o organismo que se adaptou melhor a um dado ambiente, pois os demais não sobreviveram ou não reproduziram, e seus fenótipos dependem dos genes, os quais serão preservados. Dessa forma, por exemplo, zebras se tornam cada vez mais velozes porque as que estavam um pouco mais devagar foram devoradas e não puderam ter filhos. As mudanças são pequenas, mas como Dawkins (1986) explica, o tempo faz com que elas sejam significativas, no exemplo, aos poucos as zebras mais e mais velozes vão sendo selecionadas (os predadores também se aprimoram). A mistura de organismos que ocorre na reprodução e a aleatoriedade (mutação) gerada no processo de cópia permite que a seleção natural tenha novos candidatos para selecionar.

Dawkins (1986) explica como a seleção natural é um processo cego (sem intenções), inconsciente, automático e sem propósito, capaz de moldar diferentes espécies. A seleção natural não planeja o futuro, tem apenas metas curtas, escolhe pelo momento de adaptação. Em essência, a ideia é que dada a reprodução não-aleatória com

variação na hereditariedade, as consequências serão surpreendentes e darão a ilusão da criação se houver tempo para a acumulação. A teoria de seleção natural não é uma teoria do acaso, se baseia em acumulações graduais, transformações de passo por passo, de coisas mais simples para coisas mais complexas por meio da pressão seletiva.

Dawkins (1986) não tem como explicar o surgimento dos elementos mais simples em cima dos quais a seleção natural começou a atuar. Porém, uma nova abordagem pretende dar conta disso: são as tendências da auto-organização na evolução. Ainda, estas tendências tentam mostrar que a mutação e a aleatoriedade não são únicos fatores que determinam as possíveis organizações de um organismo no qual a pressão seletiva irá atuar. Por conseguinte, estas teorias de auto-organização podem explicar um pouco mais sobre como a ordem surge. Stuart Kauffman, um biólogo que trabalha com sistêmica, foi um dos pioneiros desta ideia. De início, as tendências da auto-organização na evolução pareciam ser contrárias às ideias da seleção natural, mas a posição seguida neste trabalho será – de acordo com os raciocínios de Pereira Júnior *et al.* (2004), Kauffmann (1991, 1993) e Dennett (1995) – que o fenômeno da auto-organização em sistemas biológicos é completamente compatível e até mesmo complementar à teoria da seleção natural darwinista.

Para entender as novidades propostas por Kauffman e como elas podem ser compatíveis com a seleção natural é preciso esclarecer brevemente o que é auto-organização. De acordo com Pereira Jr. (2010, p.102) “A Teoria da Auto-Organização é parte da Teoria de Sistemas, segundo a qual os sistemas são recortados pelo observador, conforme seus interesses, e analisados a partir das interações entre seus componentes e com o ambiente externo.” Podemos delimitar um sistema a partir de qualquer agrupamento de elementos exercendo relações entre si em um dado espaço e tempo.

Todo sistema que não for fechado é organizado tanto por hétero e auto-organização. Contudo, pode existir em um sistema maior influência de um dos dois tipos de organização. Na hétero-organização há uma predominância de fatores exógenos atuando no sistema, já na auto-organização a predominância é de fatores endógenos (PEREIRA JR., 2010).

Independentemente de qual sistema a realiza, a auto-organização parece ter certas características. Pereira Jr. (2010) cita quatro destas. A espontaneidade: a relação entre os elementos trazem novidades não contidas no início do processo. Possuem a característica de derivar seus padrões de organização das relações internas entre seus componentes e apresentar respostas construtivas às perturbações externas. Apresentam

distintos níveis de organização os quais estabelecem relações de *feedback* entre si formando uma causalidade circular. Por fim, apresentam aspectos de não-linearidade, desproporção entre magnitude de causas e efeitos no sistema, gerando um “efeito borboleta”, ou seja, pequenas mudanças, mesmo que locais, repercutem por todo o sistema, mesmo que ele volte a certo equilíbrio, por um momento, esta pequena mudança é refletida de forma global.

Capturando a essência do conceito, Debrun (1996, pg.4) argumenta que “uma organização ou forma é auto-organizada quando produz a si própria”. Para o autor, apesar da forma auto-organizada depender de seus elementos constituintes, estes não determinam mecanicamente o processo a ser desenrolado por base deles. O sistema emergente na auto-organização tem origem no próprio processo e não nas condições de partida. Assim, esta herda um começo, mas o qual apenas fornece uma orientação ou impulso numa certa direção. Ainda de acordo com Debrun (1996), o principal motor da auto-organização reside na interação entre elementos realmente distintos e “soltos”; quanto maior a discrepância entre a forma final da interação e a soma das influências, maior é o nível de auto-organização.

O estudo da auto-organização requer algumas novidades de pensamento e método. Para entender um fenômeno complexo, a ciência tradicional divide este em partes cada vez menores para que estas sejam melhores compreendidas, com o objetivo final de somar as partes e ter uma imagem do todo. O uso quase pejorativo do termo reducionismo é no mínimo equivocado; afinal, do início do século XX até o começo do XXI houve saltos imensos na quantidade de conhecimento e compreensão do mundo baseado nesse método. De toda forma, no final do século XX foram descobertos fenômenos que acontecem no sistema quando analisado como um todo, os quais não são previsíveis estudando apenas os seus elementos. Um dos motivos para a defesa do reducionismo é a dificuldade metodológica para o estudo de sistemas. Contudo, o recurso tecnológico atual permitiu um avanço metodológico inovador para o estudo de sistemas, como por exemplo, o uso de modelagem computacional e modelos matemáticos. A partir destes recentes métodos é possível a análise do comportamento de sistemas como um todo, sem a necessidade de isolar elementos. Como ambos os tipos de pensamentos possuem métodos úteis e não são conflitantes, devem ser complementares.

Para chegar a sua compreensão de seleção natural e auto-organização, Kauffman utiliza de uma visão sistêmica e de métodos computacionais. Kauffman (1991) afirma

que modelos matemáticos ajudam na compreensão de sistemas complexos de processamento em paralelo. Para estudar o comportamento de milhares de elementos interconectados, o biólogo usa uma classe de sistemas chamados de redes booleanas aleatórias. Nestas redes, cada variável é binária e regulada por outras que servem como *inputs*. O comportamento dinâmico de cada variável é controlado por uma função Booleana. Uma função Booleana ‘OR’, faz a variável responder de forma ativa se alguma dentre suas variáveis de *inputs* está ativa. Uma função ‘AND’ faz a variável ficar ativa apenas se todos seus *inputs* estão ativos.

Kauffman (1991) discute versões matematicamente idealizadas de sistemas biológicos chamadas de redes booleanas NK autônomas e aleatórias. A rede consiste de elementos (N) ligados por um número (K) de *inputs*. As funções booleanas e o valor das variáveis começam aleatoriamente. São autônomas porque os *inputs* estão dentro do sistema. Porém ao deixar estas redes sem perturbações externas elas permanecem em um de seus ciclos de estado constantemente. Se a rede for perturbada, sua trajetória pode mudar. Perturbações mínimas são mudanças no valor das variáveis, e perturbações estruturais são mudanças no número de *inputs* (K) por elemento (N) ou mudanças nas funções booleanas.

Em redes com $N=K$, ou seja, todos os elementos estão ligados a todos, o comportamento é simples e caótico, a sucessão de um estado para outro é completamente instável. Estas redes exibem sensibilidade inicial máxima, pois qualquer perturbação pode mudar a trajetória destas. Outro sinal de desordem é que quanto mais elementos são incluídos, o tamanho dos ciclos de estado aumenta de forma exponencial (KAUFFMAN, 1991, p.81).

Kauffman (1991, p.81) explica que redes com $K > 3$, ou seja, três ou mais inputs por elemento, mantêm comportamentos caóticos como das redes $N=K$. Contudo, quando se chega a redes $K = 2$, algo fantástico acontece, as redes começam a exibir ordem coletiva espontânea. Os ciclos de estado deste tipo de rede se mantêm estável com quase todas as perturbações mínimas e perturbações estruturais modificam seu comportamento apenas ligeiramente. A ordem surge sem uma organização precisar ser forçada no sistema, é espontâneo. As redes possuem qualidade Homeostática, normalmente retornam para seus ciclos de estado padronizados depois de perturbações. Estas apresentam as mesmas características descritas acima sobre auto-organização e a Homeostase é uma propriedade de todas as coisas vivas.

Kauffman (1991, p.82), seguindo a sugestão de Christopher Langton, classifica o comportamento dessas redes em três tipos, sólido, gasoso e líquido. Uma rede sólida é extremamente ordenada, mas por ser tão rígida, qualquer mudança gera problemas graves para o sistema (como um programa computacional tradicional), ela é incapaz de resistir a perturbações, e não se adapta. O sistema gasoso é caótico que, como as redes $N=K$ descritas, são instáveis. Já as redes líquidas são ordenadas, porém podem chegar à beira do caos. Sistemas líquidos apresentam características fantásticas, perturbações mínimas causam diversos pequenos avalanches e raras avalanches grandes. Assim, grupos locais se comunicam frequentemente por meio de pequenas avalanches e regiões distantes se comunicam menos por avalanches grandes. Com esses mecanismos, sistemas na beira do caos conseguem praticar computações extremamente complexas.

O estado líquido parece ter capacidade otimizada para evolução, principalmente pela facilidade de mudar e adaptar. Como Kauffman (1991, p.84, tradução nossa) explica:

Assim como ensinava Darwin, mutações e a seleção natural podem aprimorar um sistema biológico pela acumulação de variantes mínimas sucessivas, assim como um técnico pode aprimorar a tecnologia. Apesar disso, nem todos os sistemas possuem a capacidade de adaptar e se aprimorar assim. (...) Redes na fronteira entre ordem e caos podem ter a flexibilidade de adaptar rapidamente e de forma bem sucedida durante a acumulação de variações úteis. Em tais sistemas equilibrados, a maioria das mutações possuem pequenas consequências por causa da natureza homeostática do sistema. Um grupo de mutações, entretanto, causam correntes de mudança maiores. Sistemas em equilíbrio irão, portanto, tipicamente adaptar a um ambiente em mudanças gradualmente, mas se necessário, podem ocasionalmente mudar rapidamente. Essas propriedades são observadas em organismos. Se redes Booleanas de processamento paralelo equilibradas entre a ordem e o caos podem se adaptar mais facilmente, então elas podem ser o alvo inevitável da seleção natural. A habilidade de abusar da seleção natural seria uma das primeiras características selecionadas¹.

As ideias de Kauffman (1991) não são apenas hipóteses sem relação com o mundo. O pesquisador colocou redes booleanas para jogarem jogos entre si praticando

¹ As Darwin taught, mutations and natural selection can improve a biological system through the accumulation of successive minor variants, just as tinkering can improve technology. Yet not all systems have the capacity to adapt and improve in that way [...]. Networks on the boundary between order and chaos may have the flexibility to adapt rapidly and successfully through the accumulation of useful variations. In such poised systems, most mutations have small consequences because of the systems' homeostatic nature. A few mutations, however, cause larger cascades of change. Poised systems will therefore typically adapt to a changing environment gradually, but if necessary, they can occasionally change rapidly. These properties are observed in organisms. If parallel-processing Boolean networks poised between order and chaos can adapt most readily, then they may be the Inevitable target of natural selection. The ability to take advantage of natural selection would be one of the first traits selected.

seleção natural com as vencedoras e descobriu que após um tempo de evolução apenas as redes em fase líquida permaneciam. A organização das redes também é comparada a características de organismos biológicos.

A ordem dos seres vivos parece não ter surgido exatamente como Dawkins (1986) pensava, como um acaso ocorrido em elementos simples que aos poucos foram tornando-se complexos por meio da seleção natural. Parece ser mais plausível que a ordem dos seres vivos tenha surgido por processos de auto-organização específicos nos quais a seleção natural atuou. Ainda, Dawkins cita a mutação aleatória como a única forma de variabilidade possível, mas é possível que a diversidade de ordem seja possibilitada por fenômenos de auto-organização. Como Pereira Jr. *et al.* (2004, p.13) sugere:

A auto-organização genético-molecular é um processo gerador de novidades orgânicas. Estas antecedem o processo seletivo, mas, uma vez geradas, tais novidades vêm a integrar o rol de alternativas a serem selecionadas – *i.e.*, a seleção natural só pode discriminar entre variedades pré-existentes. A auto-organização genético-molecular e a seleção natural seriam, portanto, forças evolutivas complementares: a auto-organização estabeleceria o conjunto de possibilidades genotípicas e fenotípicas em cujo interior a seleção operaria, estabelecendo, no fim das contas, aquelas possibilidades que seriam efetivamente realizadas nas populações existentes.

Dennett (1995) também concorda com a visão de que a auto-organização e a seleção natural são forças complementares. O filósofo concorda que Kauffman mostrou como a ordem aparece mesmo sem a seleção, não por causa dela. De toda forma, Dennett (1995, p.221) acredita que o Darwinismo explica como ocorre o *Design* das espécies, a ordem não seria explicada, apenas pressuposta. Assim, as ideias de Kauffman explicariam a ordem pressuposta pelo Darwinismo, sendo, portanto, complementares ao darwinismo. O filósofo concorda com Kauffman que nem todos os tipos de ordem possuem capacidade de evoluir rumo à maior complexidade, é preciso ter exatamente o tipo certo de ordem, com exatamente o tipo certo de liberdade e restrições, rigidez e fluidez, para que ocorra a vida, e por essas pesquisas, isso ocorre na beira do caos.

Um último fator na evolução será necessário ser apresentado por ser fundamental para a teoria de evolução da consciência de Dennett. Trata-se do Efeito Baldwin, descrito pelo naturalista James Mark Baldwin em 1896. Baldwin (1896) chama de Neurogenética as pequenas variações hereditárias que surgem do fato de cada organismo ser único, são certas características variáveis que certo organismo pode usar,

diferenciando-se do que é padrão em sua espécie. As adaptações resultantes da agência consciente, como imitação, vida social, aprendizagem com a mãe, são chamadas por ele de Psicogenéticas. Para Baldwin (1986) a adaptação neurogenética e psicogenética dispensa o apelo para o fator Lamarkista para explicar a hereditariedade social. A inteligência e a capacidade de imitação direcionam o caminho de instintos complexos mesmo seguindo a teoria Neo-Darwinista. Baldwin (1896) utiliza um exemplo para explicar seu fator:

Podemos imaginar criaturas, cujas mãos eram usadas para segurar apenas com o polegar e os demais dedos no mesmo lado do objeto, as quais descobriram, dado o estresse das circunstâncias e as variações que permitiam possíveis adaptações, como usar o polegar do lado oposto dos dedos, como fazemos atualmente. Em seguida, vamos supor que isso foi tão útil que todos os filhos que não aprenderam esta habilidade morreram; a próxima geração seria plástica, inteligente ou imitativa o bastante para assim também fazer.²

Assim, os indivíduos mais plásticos serão preservados para realizar habilidades vantajosas para as quais suas variações o permitirem, e a próxima geração terá uma ênfase justamente nesta direção. O Efeito Baldwin, portanto, é uma forma pela qual decisões feitas por características únicas de um indivíduo podem influenciar subsequentemente no processo de seleção natural do grupo. Isto porque os indivíduos com capacidade de imitar cada vez mais esses companheiros que descobriram habilidades novas serão privilegiados sobre os demais. É provável que estes indivíduos que conseguiram imitar a nova habilidade serão aqueles com disposições genéticas similares ao indivíduo que descobriu a novidade. Como estes provavelmente serão selecionados sobre os demais, o DNA da espécie cada vez mais se aproxima àquele necessário para realizar a habilidade, e com o tempo pode até vir a ser incorporado na forma de um instinto. Entretanto, para Baldwin (1896), a plasticidade é mais importante, pois quanto mais movimentos pré-instalados um animal trazer para o mundo, menos ele será capaz de aprender posteriormente.

Em suma, o naturalista acredita que espécies com a plasticidade necessária para aprender e imitar novas formas de agir durante a vida serão privilegiadas com uma aceleração no processo de seleção natural. Podemos descrever um pequeno exemplo de como a auto-organização na evolução, Efeito Baldwin e a seleção natural podem ocorrer

² We may imagine creatures, whose hands were used for holding only with the thumb and fingers on the same side of the object held, to have first discovered, under stress of circumstances and with variations which permitted the further adaptation, how to make use of the thumb for grasping opposite to the fingers, as we now do. Then let us suppose that this proved of such utility that all the young that did not do it were killed off; the next generation following would be plastic, intelligent, or imitative, enough to do it also.

juntos: processos de auto-organização genético-molecular dão origem a uma novidade fundamental para sobrevivência em certo indivíduo, a qual é imitada por outro de sua espécie, fazendo com que este que imitou tenha uma vantagem adaptativa para ser selecionada, enquanto aqueles que não conseguiram imitar ficam em desvantagem para seleção. Principalmente o Efeito Baldwin será a chave utilizada por Dennett para marcar a origem da consciência.

1.2 A Evolução da Consciência Segundo Dennett

Para Dennett (1991), no começo não havia razão, apenas causas, nada possuía propósito. Isto ocorre até o surgimento dos replicadores simples, os quais também não possuíam interesse propriamente dito, como humanos, mas podemos atribuir certo tipo de interesse a estes. Para que estes replicadores exerçam sua função, seu ambiente precisa possuir condições apropriadas. Quando o comportamento de uma entidade visa anular sua própria decomposição, ela começa agir em termos do que é “bom, ruim e neutro para ela”. Dessa forma, ela cria “interesses”. Quando a entidade começa a ter interesses, o mundo começa a criar razões para esses interesses, mesmo que a entidade não reconheça. Por exemplo, a razão de um agente buscar o calor é porque isto será bom para o funcionamento de seu sistema.

De acordo com Dennett (1991, p. 174), as razões vieram antes das criaturas que as reconhecem. Um dos problemas para os primeiros seres que enfrentaram problemas era reconhecer e agir sobre as razões que o mundo criou a partir dos interesses deles.

Para Dennett (1995), desde o início da agência o preço que acompanha toda ação é de correr o risco de errar. Os primeiros erros foram de cópia. O filósofo considera isto um erro porque existe o preço de errar, diminuição da capacidade de replicar ou até mesmo o fim da linha de reprodução. Antes deste momento não existia a oportunidade de errar, não faria sentido julgar que um processo foi errado; todo erro é subsequente a este primeiro processo. Com o tempo o processo de cópia foi se aperfeiçoando, mas, de toda forma, as pequenas divergências que podem ocorrer do padrão são exatamente as que geram mudança e a novidade na vida. De acordo com Dennett (1995), os erros e acertos são relativos à adaptação do sistema a certo meio, e apenas após o sucesso de tal sistema podemos dizer que um acertou e outro errou, porém até mesmo o que dizemos que errou pode ser um sistema mais adaptável em potencial, ou seja, algum dia pode vir a ser mais apto do que o anterior.

Quando se tem o objetivo de auto-replicação, criar fronteiras é importante. O organismo não pode ter o objetivo de preservar todo o mundo, ele precisa preservar apenas ele, tornando-se egoísta. O egoísmo, portanto, é uma das marcas da vida. Nesta tarefa de criar barreiras é preciso começar a distinguir os invasores dos amigos, e o problema foi resolvido pela criação de detectores de formas. Esta tarefa é essencial para ingestão, excreção, respiração, transpiração e para o sistema imune, ou seja, é essencial para o organismo (DENNETT, 1991).

De acordo com Dennett (1991), a seleção natural não pode nos dizer as fases pelas quais os organismos passaram, mas existem diferenças entre os sistemas criados por ela e por engenheiros inteligentes. Engenheiros preveem problemas e tentam embutir formas de solucioná-los na criação, organizando suas criações com uma função por elemento. Já a seleção natural não possui metas, não tem como se preocupar com os efeitos colaterais de suas criações. Por isso, a maioria de suas criações é “uma bagunça” (como as asas do avestruz que não são uteis para voar); porém isto tem uma vantagem: dois ou mais sistemas inicialmente sem relações funcionais podem interagir e produzir um bônus: funções múltiplas para elementos únicos.

Um dos erros que os teóricos cometem na busca de um *design* do cérebro é acreditar que a consciência funcionaria a partir de elementos com uma função apenas (DENNETT, 1991).

Dennett (1991) resume algumas lições que podemos tirar do estudo da evolução:

1. Existem razões para serem reconhecidas.
2. Onde se tem razões, existem pontos de vista para reconhecê-las e avaliá-las.
3. Todo agente precisa saber distinguir entre “aqui dentro” e “lá fora”.
4. Reconhecimentos são realizados por rotinas mecânicas e “cegas”.
5. Dentro da fronteira não é necessário um executivo central.
6. Na natureza o que importa é como está, não como chegou no ponto em que está, origens não importam (órgãos podem mudar de função).
7. Na natureza, elementos executam diversas funções dentro de um mesmo organismo.

O ponto de vista do observador humano não é o mesmo dos seus antecedentes, mas é um descendente sofisticado dos replicadores que dividiram seus mundos entre

“bom e ruim” (DENNETT, 1991). De toda forma, essa sofisticação precisa ser explicada.

De acordo com Dennett (1991), o propósito fundamental do cérebro é produzir o futuro. Para sobreviver, um organismo possui duas escolhas: se defender e torcer para que ocorra tudo bem, como uma árvore, ou criar um sistema nervoso para guiá-lo pelo mundo. Escolhendo o segundo caminho, ele precisa sempre resolver o seguinte problema: “e agora? O que fazer?”. A chave para o controle é a habilidade de rastrear ou até mesmo antecipar características do ambiente. Assim, todos os cérebros são, em essência, máquinas de antecipação. No início, todos os sinais eram respondidos com completa aceitação ou completa negação, a fuga ou o interesse (*either “scram” or “go for it”*). Esses organismos são capazes apenas de antecipação proximal: comportamento apropriado apenas para o futuro imediato. Nesta etapa não há planejamento; quando sentem dor sabem que há algo de errado, o mecanismo é pré-instalado (*“wired-in”*). Cérebros melhores são aqueles que extraem mais informação, rapidamente, e a utilizam para evitar problemas antes que aconteçam ou para buscar recursos. Para resolver estas tarefas, os organismos tentam extrair leis ou aproximações de leis do mundo. Mecanismos que na maioria das vezes agem de acordo com as regularidades do mundo são premiados.

O próximo passo é a antecipação de curto alcance, a qual também é pré-instalada (*“wired-in”*). Dennett (1991) descreve esta principalmente com um exemplo da habilidade de abaixar quando objetos são arremessados em direção ao organismo. Ela se desenvolve pela observação de regularidades de objetos com movimento em direção aos organismos provocando lesões. Após um tempo o organismo já “sabe” que objetos vindo em sua direção irão danificá-lo, não tendo que esperar até que se lesione efetivamente. Braitenberg (1984) afirma que possuímos reconhecimento automático de simetria para que identifiquemos rostos de predadores nos observando a distância. Este, de acordo com Dennett (1991), seria outro mecanismo de antecipação de curto alcance.

Uma característica importante destes mecanismos de curto alcance é que trocam precisão e verdade por velocidade e economia. Este sistema de reconhecimento de simetria, por exemplo, produz vários falsos alarmes, pois alerta o sistema de qualquer simetria, de árvores por exemplo, que seriam desnecessários, dada a função do sistema. Mas é um preço que os organismos pagam por mecanismos rápidos e “baratos” (DENNETT, 1991).

Sokolov (1975) sugeriu que os organismos possuem um mecanismo chamado de resposta de orientação. De acordo com Dennett (1991), esta se trata de um apelo do sistema nervoso como um todo para a concentração máxima em um determinado estímulo. Suponhamos que o sistema de reconhecimento de simetria dispare com força relevante. Isto desencadearia a resposta de orientação; diversos sistemas nervosos se sintonizam para identificar se há um predador, um amigo, ou alimento por perto. A segunda etapa é a liberação de adrenalina. Se o alarme for falso os sistemas retornam para suas respectivas funções rotineiras.

Dennett (1991, p.180) afirma que esse sistema de interrupção e vigilância aumentada não é propriamente a “consciência ciente” (conscious awareness), mas provavelmente é o precursor na evolução deste estado. Aos poucos este mecanismo foi se tornando mais útil e um hábito, de tal forma que os próprios organismos poderiam escolher quando utilizá-lo, mesmo na ausência de um estímulo alarme. Esta vigilância foi aos poucos se tornando exploração. Especialmente mamíferos e primatas começaram a adquirir informação não para o uso imediato, mas como precaução, para já saber das coisas antes que fosse tarde demais. Assim, estes seres viraram organismos com “fome” de informação. Como comum na evolução, este mecanismo ainda é influenciado pelo mais antigo de ir classificando o mundo entre “bom ou ruim” neste processo de exploração. Dennett (1991), seguindo hipóteses de Kinsbourne, afirma que nos mamíferos este processo foi fortalecido pela divisão de áreas cerebrais entre dorsal e ventral. No processo de exploração o dorsal se especializou em manter o organismo fora de perigo, desviando de paredes e não caindo em buracos. Isso permitiu a área ventral desenvolver técnicas de análise de informação mais precisas, utilizando de “zooms”, tratamento cuidadoso e processamento serial. O neuropsicólogo acredita que esta é a origem dos hemisférios esquerdo e direito, com o direito possuindo funções globais e espaço-temporais e o esquerdo sendo mais concentrado, serial e analítico.

Além dos mecanismos pré-instalados (“*wired-in*”), Dennett (1991) também considera outra classe de mecanismos flexíveis e plásticos capazes de aprendizagem durante a vida do organismo. Para eventos nos quais o futuro normalmente é como o passado, as estratégias pré-instaladas funcionam, porém existem eventos caóticos impossíveis de serem previstos com base no passado. Para lidar com isto, foram necessários organismos que conseguissem se modificar diante das situações que encontram. Essas modificações são chamadas de aprendizagem e desenvolvimento.

Dennett (1991) afirma que o processo de aprendizagem se assemelha ao processo de evolução. Algo fixo no indivíduo precisa fazer o papel do seletor mecânico e outros precisam ter o papel de candidatos para seleção. Diversas teorias de aprendizagem têm esse formato. No caso do condicionamento operante, o reforço é a força seletiva e a resposta aos estímulos os candidatos. Porém o filósofo acha estes mecanismos muito simples para dar conta da complexidade da aprendizagem humana.

São muitas as teorias que propõem entender aprendizagem por seleção natural no cérebro; Dennett (1991) usa um resumo comum à maioria destas, sem entrar em detalhes. Dessa forma, ele afirma que os candidatos para seleção são as diversas estruturas cerebrais que controlam o comportamento, e a força seletora é exercida por processos de remoção de itens indesejáveis geneticamente instalados no sistema nervoso.

Aprender não apenas fornece uma vantagem durante a vida dos indivíduos, mas indiretamente acaba acelerando o próprio processo de evolução por meio do Efeito Baldwin. Esse efeito foi percebido pelo naturalista Mark Baldwin (1896), mas parece não ter recebido muita atenção na época. Dennett (1991) explica este efeito com uma experiência de pensamento. Suponha que um indivíduo entre seu grupo nasça com a predisposição para realizar um feito ótimo para sobrevivência. Pelos meios tradicionais (genéticos) dificilmente conseguirá fazer com que esta característica domine na população da espécie. Porém, nos organismos capazes de aprendizagem, aqueles que tiverem uma estrutura pré-instalada (“*hard-wired*”) semelhante a do indivíduo que sabe o feito, poderão aprender durante suas vidas a realizarem esse mesmo feito. Já os que tiverem estrutura cerebral distante desta formação pré-instalada (“*hard-wired*”) não irão aprender o feito e serão extintos aos poucos. Assim, os indivíduos sobreviventes serão cada vez mais próximos da configuração pré-instalada (“*hard-wired*”) para realizar o feito até o momento em que todos da população da espécie já nasçam com esta mesma configuração que permite realizar o feito ótimo para sobrevivência. Graças ao Efeito Baldwin as espécies podem testar antecipadamente a eficácia de diferentes modelos pela exploração do fenótipo no espaço de suas possibilidades, acelerando o processo de evolução.

De acordo com Dennett (1991), graças aos olhos e a busca constante por informação, o cérebro primata foi bombardeado com informação multimodal, criando assim a necessidade de um controle dos processos de alto-nível. Até o momento podemos supor que a resolução do problema “e agora? O que fazer?” foi realizada pela

decisão entre quatro possibilidades: lutar, fugir, comer ou transar, ou alguma elaboração simples dessas quatro soluções. Mas com a enorme quantidade de informação que aos poucos os sistemas começaram a coletar, surgiu um novo problema, “e agora sobre o que vou pensar?”, ou seja, foi necessário algum processo para organizar a quantidade enorme de voluntários (papel da consciência).

Dennett (1991, p.188) argumenta que não faz sentido supor que já haveria um capitão pronto, pois ele não teria função até dado momento da evolução quando ele se tornou necessário. Assim, os conflitos entre os voluntários precisaria ser resolvido entre eles sem um executivo central, como no caso do sistema imune. Para resolução deste problema, o filósofo segue a linha da competição entre sistemas em um estilo pandemonial. Este será um foco central das discussões desta dissertação, desenvolvido nos próximos capítulos.

Dennett (1991) propõe um primata hipotético para imaginarmos o surgimento da consciência. Este seria capaz de aprender novos truques, constantemente em vigilância e sensível à novidade, mas com atenção de curto prazo e uma tendência de ser capturado por distratores ambientais. Este animal não cria projetos e objetivos de longo prazo. Seria preciso entender a passagem deste tipo de mente, para uma mente mais humana com um “fluxo consciente” capaz de organizar diversas informações e formar pensamentos coerentes.

Dennett (1991) pede para considerarmos os chimpanzés, com os quais possuímos um ancestral comum relativamente recente (seis milhões de anos atrás), e os quais possuem mais parentesco genético com humanos do que com gorilas e orangotangos. O cérebro humano difere do cérebro dos chimpanzés basicamente no tamanho e menos na estrutura. O cérebro era menor mesmo depois da separação das duas linhas evolutivas. Quando ocorreu a era do gelo (dois milhões e meio de anos atrás) a grande encefalização começou e foi essencialmente completada 150.000 anos atrás, antes do desenvolvimento da linguagem, do cozinhar e da agricultura. O cérebro dos primeiros homo sapiens (viveram entre 150.000 anos atrás até a era do gelo mais recente de 10.000 anos atrás) era extremamente complexo e quase indistinguível do nosso em tamanho e formato. Para Dennett (1991) isto é importante, pois o crescimento do cérebro estava praticamente completo antes do desenvolvimento da linguagem, e, portanto, não pode ser uma resposta às complexidades da mente que a linguagem permitiu.

Os mecanismos pré-instalados da linguagem são um incremento recente da evolução. Para Dennett (1991) esta foi acelerada pelo Efeito Baldwin. O nosso cérebro é praticamente o mesmo de 10.000 anos atrás, assim todo o avanço extraordinário dos seres humanos foi provavelmente realizado pela aprendizagem, por algum tipo de incorporação de *softwares* para aumentar seu potencial. Em linhas gerais, podemos dizer que nossos ancestrais aprenderam grandes feitos de comunicação que só estávamos começando a incorporar no sistema pré-instalado (“*hard-wired*”) pelo Efeito Baldwin. Porém, por ter aprendido esses grandes feitos, nós eliminamos boa parte da pressão seletiva. Contemporaneamente não há ameaças fortes o bastante para a humanidade para que a capacidade de comunicação mais eficaz seja selecionada por pressão do ambiente; portanto, funções linguísticas mais aprimoradas dificilmente serão incorporadas por mecanismos pré-instalados. Relembrando que esta incorporação de habilidades aprendidas em instintos pode ocorrer, como explicado acima, pelo próprio Efeito Baldwin.

Para entender como o processo de troca de “*software*” funcionou, Dennett (1991) considera a evolução da linguagem e pede para pensarmos no tempo em que a protolinguagem começou a surgir nos homo sapiens. Esses ancestrais eram onívoros que desenvolveram hábitos de vocalização com propósitos específicos (por exemplo, para informar sobre uma possível ameaça, ou para impor respeito, assim como fazem chimpanzés e gorilas). Esses sons ainda não eram “atos de fala”, para os quais é necessário que a intenção de quem produz o som para atingir certo efeito no outro dependa da apreciação do outro sobre a intenção. Mas podemos supor que esses ancestrais discriminavam entre diferentes vocalizadores e plateias em ocasiões diferentes, utilizando informação sobre o que poderiam acreditar ou querer.

Dennett (1991) especula que o ato de perguntar pode ter estimulado respostas quando outros homínídeos não estavam presentes para respondê-las. Isto ocorreria pela seguinte possibilidade: talvez o sistema que buscava a resposta não tivesse contato com outro sistema do próprio cérebro que poderia resolver o problema. Assim, ao lançar o som para o ambiente, a informação voltaria pela própria audição e seria incorporada em sistemas diferentes, criando assim uma nova rota de comunicação, a qual poderia ser mais eficiente em alguns problemas.

Pelo Efeito Baldwin esse grande feito da auto-estimulação pode ter, aos poucos, se inserido como predisposição genética. Pode ser também aprimorado na medida em que elimina o som em voz alta para uma fala em silêncio consigo mesmo, mantendo a

privacidade do conteúdo. Esse comportamento de falar consigo mesmo pode não ser a melhor forma de unir as estruturas funcionais do cérebro, mas seria um caminho facilmente atingível que faria o bastante para ajudar na sobrevivência; seria lento pois dependeria de órgãos com outras funções iniciais; seria linear como as conversas das quais ele emergiu; dependeria de seu repertório original (se houvesse 50 itens linguísticos que o hominídeo soubesse, ele poderia reportá-los para si próprio). Esse truque não precisaria ser apenas verbal; um hominídeo poderia comparar imagens de duas linhas paralelas e lembrar-se de um rio, e assim lembrar que precisaria de uma ferramenta para atravessar o rio. Posteriormente, poderia desenhar duas linhas no chão sempre que precisasse lembrar-se do rio (DENNETT, 1991).

A transmissão cultural, que informa todos sobre os grandes feitos, pode fazer com que todos consigam sobreviver bem no mundo. Dessa forma, a pressão de seleção para mover grandes feitos para o genoma é extinta ou diminuída. Quando um ser humano nasce, não demora mais do que três anos para que seu cérebro vire um cérebro alemão, inglês ou japonês. Existem mecanismos pré-instalados de aquisição da linguagem que facilitam muito e preparam para a aprendizagem nos primeiros anos de vida. Estes mecanismos pré-instalados foram estabelecidos recentemente provavelmente pelo Efeito Baldwin. Nós seríamos os descendentes dos hominídeos com maior facilidade de se ajustar e imitar dentre os que primeiro aprenderam a falar (DENNETT, 1991).

De acordo com Dennett (1991), após nossos cérebros construírem caminhos de entrada e de saída para os veículos da linguagem, eles são infectados por parasitas. Estes são chamados de memes. Ainda de acordo com Dennett (1991), a seleção natural é um processo biológico, mas podemos abstrair as características relacionais deste processo de tal forma que não se aplique apenas à vida:

1. Variação: uma abundância contínua de elementos distintos.
2. Hereditariedade ou Replicação: os elementos possuem a capacidade de criar cópias ou réplicas de si.
3. Aptidão diferencial: as cópias dos elementos criados variam com o tempo, dependendo das interações entre as características do elemento (aquilo que o diferencia de outros elementos) e características do ambiente sobre o qual ele persiste.

Seguindo esta ideia de processo evolutivo, qualquer elemento replicador, não apenas os genes, possuindo estas três características serão desenvolvidos por seleção

natural. Dawkins (1976) aponta que não precisamos ir a outros planetas para encontrar tais elementos replicadores. O biólogo diz que os novos replicadores, os memes, unidades de ideias, possuem força evolutiva forte o bastante para superar o potencial dos genes.

Estas unidades são as unidades de sentido que se replicam com fecundidade efetiva, ideias complexas que se formam em unidades memoráveis: músicas; frases de efeito; ideias; estilo de roupa; formas de criar potes; dentre outros diversos. Dawkins (1976) insiste que a ideia dos memes não se trata de uma metáfora, mas de um processo literalmente real, como o dos genes, por obedecer às leis da seleção natural. Da mesma forma que os genes de animais precisavam de plantas no ambiente para desenvolver, também os memes precisaram de cérebros humanos para poder se replicar. Os memes, assim como os genes, não ocorrem necessariamente para o bem de algo. Um meme de suicídio pode ser mais difícil de ser espalhado, mas ele pode sobreviver mesmo com os suicídios dos veículos (indivíduos) (desde que seja transmitido para outro veículo). Não há correlação necessária entre a contribuição da aptidão do meme para nossa aptidão e sua sobrevivência.

Genes são carregados por organismos sobre os quais produzem efeitos característicos, e seu destino depende dos organismos. Memes também são carregados por veículos – fotos, livros, frases de efeito, e dependem de seus veículos para sobrevivência. O destino dos memes depende das forças seletivas que agem diretamente nos veículos físicos que o incorporam (DENNETT, 1991).

A memética pode parecer apenas uma releitura do fenômeno cultural. Mas há algumas diferenças importantes. Por esta visão um aspecto cultural pode ter evoluído simplesmente por ser melhor para ele mesmo (ou seja, por ser um bom replicador, não por trazer vantagens para o veículo). Por um lado, não sobreviríamos se não tivéssemos a tendência de escolher os melhores memes para nós. Mas de toda forma, o sistema de escolha não é perfeito, assim alguns memes inúteis ou problemáticos podem persistir (DENNETT, 1991).

De acordo com Dennett (1991), memes dependem de mentes para sobreviver, por enquanto precisamente mentes humanas, pois não nega a possibilidade destes replicadores algum dia fazerem parte da mente de novos animais ou robôs. Mentem possuem limites em recursos e só podem manter uma capacidade limitada de memes, e assim, há uma competição entre memes para entrada no máximo de mentes possíveis. Esta competição é a maior força seletiva que atua sobre os memes. Alguns memes são

“ruins” e outros “bons”; mas possuem em comum o efeito no fenótipo com tendência a desativar as forças seletivas contra eles próprios. A fé tem a tendência de inibir o pensamento crítico que poderia a demolir; as teorias de conspiração não aceitam a falta de evidências como razão para descrença, com a justificação de ser uma conspiração muito forte. Memes possuem outras características de genes, como de fazer parcerias entre si para sobreviver.

A mente humana é um ótimo local de sobrevivência para os memes, e estes a reestruturam para adequar seus habitats para suas necessidades. Todos os cérebros são diferentes no nascimento. Um cérebro chinês difere significativamente de um inglês, e o cérebro de quem sabe ler é diferente de um cérebro analfabeto. O que faz um cérebro humano se diferenciar de outro depende das diferenças microestruturais introduzidas por diversos memes (DENNETT, 1991).

Existe uma persistente tensão entre a “vontade” dos genes e a dos memes, mas na maioria dos casos, seríamos tolos de escolher a dos genes, pois os memes têm maior força em determinar nosso sentimento de quem nós somos. Nosso *self* é também uma construção dos memes (DENNETT, 1991).

Todos os três caminhos de evolução – evolução genética, plasticidade fenotípica (por meio do Efeito Baldwin) e evolução memética – contribuíram para o desenvolvimento da consciência humana. Cada caminho possui sua própria velocidade. A evolução memética é recente, mas apresenta um poder extraordinário, nos dando a possibilidade para uma quarta forma de evolução, por meio da ciência, a engenharia neurocientífica e genética.

Para Dennett (1991), é importante que a consciência seja pensada como *software* por:

1. Ser muito recente para já estar pré-instalada (“*hard-wired*”) na estrutura biológica.
2. Ser em grande parte produto da evolução cultural, que se estabelece em cérebros pelo treinamento na infância.
3. O sucesso em sua instalação ser determinado por microconfigurações do cérebro, significando que suas características funcionalmente importantes são praticamente invisíveis para a neuroanatomia.

É importante ressaltar um ponto. Se Dennett (1991) está propondo que a consciência está relacionada ao falar consigo mesmo e a memes, estaria ele propondo que a consciência é unicamente humana? Parece ser este o caso. Com isso ele parece

fugir um pouco de sua vertente darwinista de aprimoramentos graduais. No capítulo “*Our consciousness, their minds*” (nossa consciência, a mente deles) do livro “*Kinds of Minds*” (Tipos de Mentes), Dennett (1996) argumenta que apenas seres humanos conseguem usufruir de memes de tal forma a construir a consciência. Ele acredita que os demais seres possuem mentes, mas não consciência.

Dennett (1996) cria uma taxonomia de quatro tipos de mentes seguindo uma ordem do surgimento dessas mentes durante a evolução. As primeiras criaturas a surgirem seguiam cegamente suas pré-disposições geneticamente instaladas. Dentre uma mesma espécie, vários fenótipos podem surgir. Por seleção natural os fenótipos mais aptos permanecem, enquanto os menos aptos são eliminados. Por fim, esse processo resulta na multiplicação de genótipos mais aptos para agir em certo ambiente e em certo momento do mundo. Essas criaturas são chamadas de criaturas darwinistas. Os primeiros replicadores e bactérias atuais são bons exemplos destes tipos de criaturas.

Eventualmente criaturas começaram, de forma cega, a tentar diferentes escolhas de ação aleatoriamente, até que algumas dessas ações fossem escolhidas por reforço. Essas criaturas não apenas seguiam instruções genéticas para ação, o meio reforçava iniciativas que foram bem sucedidas durante a vida de um indivíduo da espécie. Isto acelera a adaptabilidade de um ser, porque ele pode aprender comportamentos aptos durante a vida, incorporando novidades de adaptabilidade que seus genes não teriam como “prever”. Dennett (1996) chama criaturas com esse tipo de mente de criaturas skinnerianas. Entretanto, os behavioristas acreditavam que a maioria dos seres possuíam esse tipo de mente, como os pombos e ratos que usavam em seus laboratórios. Dennett (1995) sugere que este é um grande engano, e considera a lesma do mar como um bom exemplo deste tipo de criatura.

Pombos e ratos, para Dennett (1996) seriam criaturas popperianas. Estas criaturas seguem as formas de aprendizagem já descritas, mas além disto, estas criaturas não precisam ficar aleatoriamente testando seus comportamentos até que um eventualmente seja reforçado pelo ambiente durante a vida. Elas podem criar hipóteses plausíveis de como agir no mundo antes de terem seu comportamento reforçado por alguma situação. O reforço continua sendo uma fonte fundamental para moldar o comportamento destas criaturas, mas as primeiras tentativas de comportamento antes do reforço já não são completamente aleatórias. A partir de testes de suas hipóteses de ação em um “ambiente interno” antes de sua primeira tentativa de agir, a criatura cria uma probabilidade maior do que puramente chance de atingir seu objetivo.

O último tipo de criatura consegue utilizar as estratégias das três demais. Mas além de criar hipóteses que são testadas e reforçadas, elas podem dividir suas novidades umas com as outras e criar ferramentas. As novidades interessantes não precisam apenas ser selecionadas e incorporadas nos genes, não precisam nem ser pensadas por cada indivíduo, elas podem ser divididas pela aprendizagem cultural. Uma criatura informa outra que uma ferramenta pode ser útil para sobrevivência, sem que a criatura que aprendeu precise ter isto incorporado em sua estrutura instalada, sem calcular as possibilidades da ferramenta ser útil, e sem precisar ser reforçada. Criaturas com este tipo de mente são chamadas de criaturas gregorianas, bons exemplos destas são chimpanzés e homínídeos.

Seres humanos chegaram a este último tipo de mente, mas através da linguagem e da ampliação do uso de memes conseguiram ir além, organizando suas mentes através de uma propriedade instalada culturalmente, a consciência.

Antes de comentarmos a visão da evolução da consciência de Dennett, primeiramente será apresentada a forma como Baars compreende esta, para que no final do capítulo os comentários a cerca de ambos possam ser realizado a partir de um contraste das visões.

1.3 Funções Adaptativas da Consciência Segundo Baars

Baars (1988) descreve diversas funções adaptativas da consciência, e, portanto, os motivos vantajosos pelos quais a consciência pode ter sido selecionada durante a evolução biológica.

Para Baars (1988) a consciência tem a função de auxiliar o organismo a se situar nos e responder aos diversos contextos distintos a que ele terá de se adaptar. O papel da consciência neste âmbito e de auxiliar no processo de evocação de conteúdos adequados para solucionar problemas, dadas certas condições não completamente previsíveis. Por isso também ela tem o papel de auxiliar na aprendizagem. Conteúdos já conhecidos ou pouco informativos ficam sempre longe do foco da consciência. Quando olhamos para um conteúdo antigo conscientemente é porque estamos encontrando algum tipo de novidade neste conteúdo, nem que seja um momento para “reviver” (reavaliar) memórias e sentimentos.

A consciência também tem a função adaptativa de editar, marcar e depurar conteúdos. Esta não só nos ajuda a aprender, mas também a adaptar o já aprendido a

mudanças. Assim, a consciência consegue localizar possíveis erros e editar o conhecimento já aprendido. Esta pode também depurar (*debug*) processos que se tornaram não adaptativos. Dado todos os processos ocorrendo no cérebro e no mundo, a consciência tem a função de marcar aqueles que serão essenciais para o organismo. Desta possibilidade de marcar, surge a capacidade de recrutar certos processos voluntariamente para exercer uma meta. A consciência também exerce um controle sobre as funções mentais do organismo. Por exemplo, uma meta ao se tornar consciente consegue fazer com que coordenações do corpo como um todo tenha o objetivo de cumprir esta meta como sua finalidade (BAARS, 1988).

Segundo Baars (1988), a consciência tem a função de controlar a probabilidade de um conteúdo se tornar consciente. É possível treinar a atividade de evocar conteúdos específicos e de formas específicas, por exemplo, o uso de certas palavras em certos contextos. O controle se estende para a tomada de decisão na medida em que ficar consciente de possibilidades de ação abre o caminho para escolha de uma dessas, que em seguida poderá influenciar o comportamento. Não apenas para a tomada de decisão em si, mas também para o controle de impulsos. A metacognição consciente nos permite controlar comportamentos impulsivos através da repetição de valores e critérios para ação, por exemplo, crianças impulsivas podem aprender a controlar seu comportamento usando a fala interna. Ainda com esta função, a consciência nos permite avaliar nossos comportamentos com um conceito de *self* próprio, por exemplo, quando queremos muito agir, mas não fazemos por achar que a atitude não corresponde com nós mesmos. Mais especificamente, podemos ter o impulso de agredir alguém com quem estamos tendo problema, mas conscientemente conseguimos impedir esta ação por não nos considerar o tipo de pessoa que agrediria outra. Podemos conscientemente nos treinar a comportar de certa forma, prevenindo hábitos considerados ruins e instalando hábitos considerados positivos. Logo, a consciência está envolvida em uma forma de autoprogramação e manutenção de si mesmo.

A capacidade da consciência de poder relacionar diversas modalidades de conteúdos abre caminho para a formação de metáforas e analogias. Para Baars (1988) essa função possibilita que utilizemos conhecimentos de áreas completamente diferentes como se fossem similares, por exemplo, quando comparamos uma configuração de elétrons com órbitas em torno do um núcleo ao sistema solar. Trata-se de domínios do conhecimento completamente distintos, mas conscientemente conseguimos achar alguns

padrões semelhantes em ambos, assim transferindo nosso conhecimento de um contexto para outro.

1.4 Consciência em Animais

No livro “*A cognitive theory of consciousness*”, no qual Bernard Baars descreve pela primeira vez a Teoria do Espaço de Trabalho Global, a consciência já era entendida como uma adaptação biológica. Mais recentemente, em conjunto com outros pesquisadores (Seth & D. Edelman), foram propostos critérios para determinar a consciência em animais, mostrando a origem antiga da consciência.

O método de trabalho é comparativo. Utilizando o animal humano como base, comparam as estruturas anatômicas, a fisiologia, a cognição e o comportamento destes com de outros animais. Dada a limitação e a dificuldade de atribuir consciência a outros animais, Edelman (1989) e Edelman et al. (2011) propõem uma diferenciação cuidadosa entre consciência primária e secundária. A primária refere-se à presença de uma cena multimodal composta de eventos perceptivos e motores com uma perspectiva em primeira pessoa sobre o mundo. Já a secundária, refere-se à utilização da capacidade de pensamentos de alta-ordem para deixar os conteúdos da consciência primária passarem por uma semântica interpretativa, incluindo um sentido de *self* e a construção de cenas do passado e do futuro. Assim como é para Dennett, a consciência secundária está ligada a capacidades linguísticas.

Os resultados recentes destas pesquisas comparativas foram publicados em forma de uma declaração assinada por pesquisadores renomados em consciência animal. Em suma, a declaração expressa:

A falta de um néocórtex parece não impossibilitar a experiência de estados afetivos. Evidências de fontes distintas indicam que animais não-humanos possuem substratos neuroanatômicos, neuroquímicos e neurofisiológicos de estados conscientes, além da capacidade de exibir comportamento intencional. Consequentemente, o peso das evidências indica que humanos não são os únicos a terem as bases neurológicas da consciência. Animais não-humanos, incluindo mamíferos e aves, e até criaturas como polvos também possuem esses substratos neurais (LOW et al., 2012, tradução nossa)³.

³ The absence of a neocortex does not appear to preclude an organism from experiencing affective states. Convergent evidence indicates that non-human animals have the neuroanatomical, neurochemical, and neurophysiological substrates of conscious states along with the capacity to exhibit intentional behaviors. Consequently, the weight of evidence indicates that humans are not unique in possessing the neurological substrates that generate consciousness. Nonhuman

A primeira vista, pode parecer estranho que estudos comparativos da consciência humana geraram resultados indicando que a ausência de um neocórtex não impossibilita a consciência. Mas revendo os critérios adotados para determinar consciência e um pouco da metodologia e das evidências será possível entender esses resultados.

Seth *et al.* (2005) descreve 17 critérios para determinar a consciência animal, entretanto, não é preciso que o animal preencha todos estes, afinal, seria extremamente difícil encontrar evidências para todos. O primeiro critério é baseado em evidências de eletroencefalograma que diferenciam estados de vigília de estados de sono profundo, estados vegetativos e convulsões. O estado de vigília é associado à atividade rápida, irregular, e de baixa amplitude no núcleo talâmico. Já os estados inconscientes apresentam atividade lenta, regular, de alta amplitude no núcleo talâmico. Todos os mamíferos estudados apresentam esta diferenciação de estados. Esta diferença é altamente interessante por ser tão clara e facilmente verificável em animais. Entretanto, uma pequena dificuldade pode estar em como certificar que estes estados de sono profundo, vegetativos e de convulsões são realmente estados completamente inconscientes. O eletroencefalograma poderia estar verificando formas diferentes de estados conscientes e menos conscientes, mas não necessariamente completamente inconscientes.

O segundo critério é a participação do sistema talâmico cortical. Lesões no cerebelo, na medula espinhal e no resto do corpo em quase nada influenciam estados conscientes. Em contraste, lesões no córtex ou no tálamo geram distúrbios significativos. Lesões no córtex estão associadas a perdas do conteúdo da consciência e de funções específicas, por exemplo, lesões no córtex occipital podem gerar a perda da consciência visual. Já lesões no tronco encefálico, ou mais especificamente no tálamo, geram a completa perda da consciência como um todo. O tronco encefálico parece estar relacionado com a manutenção de estados conscientes, enquanto o córtex parece estar relacionado com o conteúdo e funções específicas, logo, a consciência parece envolver atividade dinâmica do sistema talâmico-cortical. (Seth *et al.*, 2005) Uma chave central desta proposta é que a consciência está envolvida na conectividade reentrante entre áreas de percepção e da memória (Edelman, 1989).

animals, including all mammals and birds, and many other creatures, including octopuses, also possess these neurological substrates.

O terceiro critério para determinar a consciência é a presença de atividade difundida por grandes áreas do cérebro. A atividade de certas regiões por si parece não ser o bastante para gerar a consciência. Atividades novas recrutam funções difundidas por todo o cérebro, enquanto funções automatizadas (normalmente inconscientes) estão relacionadas com atividades de funções localizadas.

A consciência envolve diversos tipos de conteúdos, fenômenos linguísticos do lobo temporal, estímulos visuais captados pelo lobo occipital, fenômenos do sentimento de uma palavra estar “na ponta da língua” resultantes principalmente do lobo frontal. Assim, o quarto critério de Seth *et al.* (2005), exige que a consciência tenha multimodalidade, envolvendo diversas áreas do cérebro.

O quinto critério é a meta informativa da consciência. Seres desenvolvem a consciência como uma forma de lidar com novas informações. A consciência é centrada no presente, a memória sensorial tem a duração de poucos segundos, rapidamente a cena presente é substituída pelo “próximo presente”. Logo, ao invés de se manter no passado, a consciência precisa estar sempre atualizada com a informação mais atual possível. Em suma, o sexto critério diz que a consciência, para ser adaptativa, precisa ter a capacidade de recrutar várias redes neurais o mais rápido possível para ajudar no comportamento e manter-se no presente. O sétimo exige que essas cenas conscientes do presente sejam internamente consistentes. Da necessidade de se manter sempre no presente de uma forma internamente consistente, automaticamente surge a necessidade de esse sistema funcionar em serial (o oitavo critério), uma cena segue diretamente após o fim da última (estas propriedades da consciência serão discutidas no capítulo seguinte).

O nono critério de Seth *et al.* (2005) diz que a consciência tem a função de atar (*bind*) diversos conteúdos. Mesmo em uma área específica como o córtex visual, existem grupos neurais responsáveis pela discriminação de cores, formas, dimensão espacial e temporal. A consciência consegue uma forma adaptativa de unir diversas discriminações realizadas por módulos distintos de processamento.

A experiência consciente é atribuída a um eu (*self*) que a experiencia, em seres humanos essas funções parecem estar associadas ao córtex orbitofrontal, uma área pouco desenvolvida em animais. Entretanto, não apenas humanos possuem um eu (*self*). Além disso, existem áreas semelhantes ao córtex orbitofrontal em animais. O córtex parietal produz mapas egocêntricos que posicionam o corpo de quem percebe relativo a cena visual. Essa capacidade promove a função de planejar a ação do observador em

relação à cena visual em um presente momento, como para alcançar um objeto. Por isso, a atribuição de um eu (*self*) é o décimo critério relacionado à consciência. O décimo primeiro critério diz que fenômenos na consciência podem ser relatados pelo sujeito. Bastante ligada ao eu, está o décimo segundo critério, a perspectiva em primeira pessoa. A consciência está relacionada com uma perspectiva em primeira pessoa sobre o mundo, sendo os conteúdos desta relativamente privados e apresentados na formas de *qualia* (qualidades sensíveis disponíveis para quem observa o mundo). Entretanto, nem todos os conteúdos mentais têm uma sensação, ou um *qualia* associado. O décimo terceiro critério diz respeito a fenômenos de margem, os quais não têm um *qualia* bem definido, como o estado de busca por alguma informação interna, como o fenômeno do conteúdo “na ponta da língua” e sentimentos de familiaridade, e até o próprio sentimento de um eu subjetivo (*self*).

A consciência parece acelerar intensamente o papel da aprendizagem dos organismos no nível fenotípico. O décimo quarto critério de Seth *et al.* (2005) diz que a consciência surge para facilitar a aprendizagem, associada a atividade global de redes talâmico-corticais, em contraste com o relato do já aprendido, que ocorre de forma localizada.

De acordo com Seth *et al.* (2005) a consciência parece ser o resultado filogenético de processos que garantem a estabilidade de nossa percepção em relação às constantes mudanças de nosso corpo e do mundo para garantir decisões efetivas. Logo, o décimo quinto critério diz que a consciência precisa mostrar uma estabilidade (relativamente virtual) do mundo para os organismos tomarem decisões melhores sobre o que fazer no momento seguinte. Dessa facilitação, surge a característica de objetos da consciência terem um caráter de “outro”. Até quando falamos de nós mesmos em terceira pessoa, estamos nos colocando no foco da consciência como um “outro”. Essa capacidade de destacar nós mesmos do mundo e nos referirmos aos objetos desses como outro é o décimo sexto critério. Por fim, o último critério exprime que a consciência está ligada à constante busca de informação e conhecimento (não apenas em momentos de urgência), facilitando a tomada de decisão.

Para atribuir consciência a animais e assim traçar suas origens evolutivas é preciso encontrar essas características da consciência nestes. Alguns desses critérios são mais facilmente testáveis do que outros. Em geral, eles precisam ser testados pela comparação de neuroanatomia, fisiologia, ressonância magnética, eletroencefalograma, experimentos inteligentemente desenvolvidos para verificar a resposta de um animal a

certa situação e pela observação do comportamento em condições naturais. Mesmo com a variação de métodos, muito desses critérios não são facilmente testáveis.

Voltamos à questão proposta no início da seção: como, se o complexo talâmico cortical é essencial para a consciência, poderia a consciência animal não estar relacionada com a presença de um neocórtex? O que ocorre em animais é que funções semelhantes exercidas pelo neocórtex em seres humanos são exercidas por estruturas anatômicas diferentes, mas homólogas, ambas gerando consciência. Edelman e Seth (2009) explicam que o sistema nervoso vertebrado segue um plano corporal bastante conservado que teve início com os primeiros cordatas 500 milhões de anos atrás. Assim, é possível traçar a origem embriológica de diversas estruturas neurais. Os homólogos de estruturas subcorticais de aves são relativamente facilmente identificados. Pesquisas de Reiner (2005) sugerem que a composição e circuitos básicos foram estabelecidos em configurações do núcleo muito antes do surgimento do córtex mamífero. Ainda, segundo Edelman e Seth (2009), existem populações neurais em aves homólogas àquelas em cérebros mamíferos, que nestes últimos são responsáveis por receber input talâmico e projetar para o tronco encefálico e para medula espinhal. Existem, entre mamíferos e aves, grupos de neurônios que são quase idênticos tanto na morfologia quanto à suas propriedades fisiológicas. Homologias estruturais também podem ser identificadas usando tecnologias moleculares e histológicas. Assim, é possível confirmar a presença dos mesmos neurotransmissores, neuropeptídios, e receptores funcionando em áreas homologas em mamíferos e aves.

O desenvolvimento do tálamo e do giro cingulado anterior e posterior algum tempo antes do surgimento dos primeiros mamíferos indica esse momento como o possível início de uma forma da consciência primária, sem a presença de um córtex desenvolvido (neocórtex). Apesar de ser difícil dizer, Edelman et al. (2005), sugerem que a consciência tenha surgido mais de uma vez depois da separação das linhas reptilianas de anapsidas e sinapsidas 300 milhões de anos atrás, as quais levaram posteriormente ao surgimento de aves e mamíferos respectivamente.

Além de homólogos de estruturas neurais, o comportamento e as capacidades cognitivas também parecem indicar consciência primária. Edelman e Seth (2009) mencionam capacidades da memória de aves, uso de ferramentas e construção, ajuste do comportamento para contextos sociais e identificação de competição, e até mesmo um vocabulário léxico desenvolvido utilizado com significado quando treinado (Pepperberg, 2000). Capacidades de memória episódica, de trabalho e de aprendizagem

espacial são demonstradas com estudos de campo e de laboratório (Edelman e Seth, 2009).

Edelman e Seth (2009) consideram as pesquisas de Pepperberg (2000) como possíveis indicadores de certo grau de consciência secundária em aves. As aves de Pepperberg conseguem nomear objetos em tarefas de categorização e fazem discriminações sofisticadas. Por exemplo, Pepperberg (2000) apresentava uma chave amarela grande e uma chave prata menor e perguntava “qual é maior?”, o papagaio respondia “a amarela”. Pepperberg (2000) afirma que os papagaios acertam esse tipo de pergunta com estímulos e perguntas novas, e não apenas tarefas reforçadas pelo método behaviorista. Edelman e Seth (2009) consideram que esses papagaios conseguem fazer um julgamento sobre eventos da consciência primária, identificando, por exemplo, exatamente o que mudou em certa cena perceptiva.

A sugestão de que a consciência pode ter surgido a 300 milhões de anos atrás é baseada nas evidências em animais. Mas isso não significa que novas evidências não possam mostrar que a consciência primária tenha surgido em períodos mais antigos ainda. O que é improvável, é que novas evidências mostrem que a consciência surgiu em períodos mais recentes, pelas razões descritas acima. No mesmo sentido, é possível que a consciência tenha surgido em outras linhas evolutivas, como parece ser o caso dos polvos. Os polvos são do filo Molusca e não seguiram o mesmo caminho evolutivo dos mamíferos, se separando desta linha a 500 milhões de anos atrás (Edelman et al. 2005).

O cérebro do polvo é grande, mas não tem nenhuma semelhança estrutural com o de aves ou mamíferos, mesmo assim exibe funções cognitivas semelhantes. David Edelman e Seth (2009) afirmam que polvos podem fazer discriminações entre objetos diferentes baseados em tamanho, forma e intensidade. Mostram flexibilidade, persistência de trações de memória, e aprendizagem baseada em contextos. Polvos são capazes de encontrar o caminho correto para uma recompensa em labirintos de acrílico e podem retirar objetos de garrafas tampadas. É possível que eles sejam capazes de aprendizagem observacional, ou seja, aprender apenas observando outro ser realizar a atividade, sem precisar ser treinado ou condicionado. Era comum atribuir esta capacidade apenas para animais altamente sociais. Polvos parecem apresentar memória de curto-prazo e longo-prazo, lembrando-se de modificações aleatórias realizadas em labirintos, ajustando seus movimentos. Por fim, polvos parecem gastar certo tempo para tomar a decisão do que fazer ou qual caminho seguir, indicando processamento de informação sobre as possíveis variáveis.

1.5 Considerações Sobre o Capítulo

Ambas as teorias evolutivas da consciência são interessantes e possuem alguns pontos de similaridade e outros de conflito. Para Dennett (1991), a consciência resulta principalmente de um processo de aprendizagem e transmissão cultural através do Efeito Baldwin. É claro que para a consciência emergir, a evolução biológica já precisava ter tomado diversos passos sem a ajuda da imitação e da aprendizagem. Mas Dennett considera que a evolução biológica por si construiu apenas mentes, sendo a consciência o resultado da instalação cultural de hábitos de falar consigo mesmo, de referir-se sobre os eventos do próprio cérebro. Poderíamos dizer assim que para Dennett, apenas a consciência secundária pode ser considerada a consciência genuína. O que Baars, Edelman e Seth chamam de consciência primária seria apenas formas mais evoluídas de mentes. O filósofo reserva o termo consciência para a consciência secundária. Seriam estas diferenças puramente terminológicas?

Acreditamos definitivamente que não. Primeiramente porque as palavras utilizadas para a descrição do mundo já mostram um pouco de nossas crenças, mas principalmente pela diferença entre a continuidade do processo evolutivo e a separação da espécie humana. Para Baars, Edelman e Seth a consciência secundária evolui a partir de novos incrementos da consciência primária seguindo leis darwinistas clássicas. Dennett, ao utilizar o Efeito Baldwin, memes e a cultura para explicar o surgimento da consciência gera um tipo de salto evolutivo, que distancia os seres humanos bastante de outras espécies, fugindo de sua corrente naturalista. Já Baars, Edelman e Seth apesar de considerarem a relação com a linguagem fundamental para a consciência secundária, assumem a possibilidade de formas mais básicas de consciência secundária, por exemplo, nas aves de Pepperberg (2000). Acreditamos que estes pesquisadores mantêm em discussão aberta a possibilidade de outros animais como primatas possuírem consciência secundária. Já Dennett mantém as portas fechadas para essa possibilidade em animais por considerar a linguagem desenvolvida com sintaxe a única capaz de instalar a consciência. Outro diferencial da perspectiva proposta por David Edelman é da consideração da importância dos estados afetivos para que haja consciência, como apresentado na declaração de Cambridge, os quais também não possuem relação com a linguagem. Por fim, Baars (1988) afirma que a percepção (extremamente antiga

evolutive) tem um papel central para a consciência, tendo acesso privilegiado à mesma.

Acreditamos numa terceira possibilidade na qual seria possível até mesmo consciência secundária plena em animais sem linguagem. Acreditamos que podemos aprender uma linguagem apenas se já tivermos condições cognitivas complexas o bastante para isso. Logo a cognição, ou o poder computacional necessário, tem que existir em maior grau anteriormente à linguagem. A linguagem tem um papel essencial de comunicação, e isso amplia nosso conhecimento e funcionamento cognitivo. Mas ao mesmo tempo, já precisávamos ter capacidades de alta-ordem, ou consciência secundária para podermos aprender essa habilidade complexa chamada de linguagem. Por isso, acreditamos não ser impossível encontrar animais como pássaros, polvos e primatas com essa capacidade completamente desenvolvida. Por esta possível perspectiva, a característica marcante dos animais humanos estaria relacionada à transmissão do conhecimento cultural acelerado pela linguagem, mas não à singularidade de sua consciência.

No próximo capítulo será apresentada a Teoria do Espaço de Trabalho Global de Bernard Baars e o Modelo de Esboços Múltiplos de Daniel Dennett, desenvolvidas de forma a corresponder com as respectivas teorias evolutivas da consciência.

CAPÍTULO 2 – ESPAÇO DE TRABALHO GLOBAL E ESBOÇOS MÚLTIPLOS

*“But why should consciousness be
the only thing that can’t be explained?”*
—Daniel Dennett (1991)

Introdução

O título do livro de Dennett, “*Consciousness Explained*”, parece à primeira vista audacioso. Somos induzidos a pensar que o filósofo acredita ter resolvido o problema da consciência em seu livro. Mas, como Daniel Dennett sempre usa de jogos de palavras em seus títulos, uma segunda interpretação se torna mais plausível. O título parece ser uma resposta ao mito de que a consciência é mágica e foge às explicações científicas. Apesar de Dennett não explicar completamente o fenômeno da consciência, ele traça um caminho, argumentando como ela pode ser explicada cientificamente. “*Consciousness Explained*” seria, portanto, um exemplo de como a consciência pode ser, de fato, explicada empiricamente. Essa segunda interpretação parece ser coerente com o fato de ele declarar que sua teoria é só um início, uma base para outras teorias serem construídas.

Neste capítulo serão apresentadas duas teorias empíricas da consciência, a Teoria do Espaço de Trabalho Global de Bernard Baars (1988) e o Modelo de Esboços Múltiplos de Daniel Dennett (1991). As teorias são descritas em livros ricos em conteúdo sobre a Psicologia Cognitiva e a Filosofia da Mente respectivamente, sendo estes extremamente detalhados, servindo não só como uma explanação da teoria, mas como uma apresentação das áreas de estudo. O livro “*A Cognitive Theory of Consciousness*” foca na construção de uma teoria da consciência, a qual se aprimora durante os capítulos. Em cada capítulo, evidências empíricas e argumentos são usados para indicar como uma teoria completa da consciência pode incorporar explicações para funções cognitivas como percepção, atenção, controle voluntário, dentre outras. O autor não apresenta a teoria de forma dissertativa; o livro segue como se estivesse construindo modelos durante os capítulos, acrescentando detalhes em cada etapa.

O livro “*Consciousness Explained*” é dividido em três partes. Trata inicialmente dos problemas e métodos relacionados ao estudo da consciência. Na segunda parte, a partir da crítica à ideia do “Teatro Cartesiano”, de especulações sobre a evolução da consciência e da análise de como a produção da fala ocorre em seres humanos, Dennett, de forma metódica, constrói seu modelo da consciência. O autor segue para uma terceira etapa, na qual aplica sua compreensão de consciência aos problemas clássicos da literatura psicológica e da filosofia da mente, como: representações mentais, visão-cega (*blindsight*), *qualia*, discussões sobre a possibilidade de zumbis fenomenológicos, consciência em seres não humanos e a realidade do *self*.

Os artigos recentes sobre a Teoria do Espaço Global de Trabalho parecem ter tomado uma direção mais científica, com mais evidências empíricas da ciência cognitiva e neurociências a fortalecendo. Ainda, outros modelos adotaram aspectos desta teoria e houve propostas de implementação desta em máquinas (BAARS, 2002; BAARS e FRANKLIN, 2003; CHANGEUX e DEHAENE, 2008; DEHAENE e CHANGEUX, 2011; EDELMAN, GALLY e BAARS, 2011; MADL, BAARS e FRANKLIN, 2011). Os artigos recentes a respeito do Modelo de Esboços Múltiplos de Dennett parecem lidar com aspectos mais filosóficos, sobre a necessidade de um processador central, ou de uma linha de chegada à consciência (CHRISTIE e BARRESI, 2002; SCHNEIDER, 2007; TODD, 2006; 2009).

Não haverá espaço nesta dissertação para a explanação de todas as ideias abarcadas pelas teorias. Haverá, neste capítulo, uma descrição das ideias centrais de ambas as teorias, especialmente aquelas que permitam que no terceiro capítulo seja possível a discussão aprofundada sobre o Teatro Cartesiano, conceito metafórico que diz respeito a problemas sobre o fluxo da consciência, a localização da consciência no cérebro, a linha de chegada para um estado consciente, dentre outros.

A primeira parte deste capítulo é destinada à teoria de consciência de Baars (1988). Na seção 2.1 são descritas as bases elementares sobre as quais o Baars (1988) constrói sua Teoria do Espaço de Trabalho Global, também sua metodologia fundamental. Na seção 2.2 os eixos principais da teoria são propriamente descritos. Uma breve consideração sobre consciência e informação na teoria de Baars é realizada na seção 2.3. Por fim, algumas observações particulares são realizadas sobre esta teoria na seção 2.4.

Na segunda parte deste capítulo uma primeira aproximação à teoria da consciência de Dennett é realizada. A seção 2.5 descreve o método de investigação da consciência endossado pelo filósofo. Na seção 2.6 é realizada uma primeira introdução de como Dennett tenta romper com a visão tradicional do Teatro Cartesiano substituindo esta por um entendimento do ponto de vista espalhado. As seções 2.7 e 2.8 descrevem as afirmações principais do Modelo de Esboços Múltiplos da consciência. Por fim, algumas observações particulares são realizadas sobre esta teoria.

2.1 Noções elementares da Teoria do Espaço de Trabalho Global

Bernard Baars (1988) é um Psicólogo, assim, para bem exercer seu papel, ao criar uma teoria da consciência, deve focar evidências empíricas a respeito deste fenômeno. É exatamente dessa forma que Baars desenvolve sua Teoria do Espaço de Trabalho Global. A Psicologia Cognitiva em 1988 já apresentava um vasto corpo de resultados de experimentos em áreas como percepção, raciocínio, linguagem, atenção, dentre outras. O trabalho de Baars em seu livro foi de realizar uma síntese dessas evidências, mas uma síntese com um olhar inovador, pois nem todos esses experimentos tratavam diretamente da consciência. Um dos méritos de Baars foi o de perceber que estes experimentos já diziam muito sobre a consciência, permitindo-o criar uma teoria que abarcasse muitas dessas evidências.

O método de trabalho desenvolvido por Baars (1988) foi denominado “análise contrastiva”. O procedimento deste método é de manter a consciência como a variável independente e as demais variáveis o mais constante possível. Uma forma de realizar este método é por meio de um contraste entre um estímulo consciente e o mesmo estímulo depois de ter sido habituado. A habituação ocorre quando um estímulo já foi repetido demasiadamente, ao ponto de o indivíduo deslocar sua atenção para outros estímulos, ou seja, perder a consciência deste. Esse fenômeno também foi percebido inicialmente por Sokolov (1963), sendo o oposto da resposta de orientação mencionada no capítulo anterior. A partir da análise contrastiva é possível descobrir a diferença entre o processamento inconsciente e o consciente, pois o estímulo é mantido constante, o organismo é o mesmo, mas em um caso há consciência, enquanto no outro não. Um exemplo ilustrativo é a diferença entre uma lembrança consciente sobre o almoço de hoje e essa mesma lembrança em outro momento, quando está na memória, de forma inconsciente. A análise contrastiva permite comparar o processamento inconsciente com o consciente por um olhar em terceira pessoa, permitindo uma análise científica. Mas será possível definir se um indivíduo está ou não consciente de algo por uma perspectiva em terceira pessoa? Baars (1988) utiliza dois critérios para definir isto.

O primeiro critério para definir se alguém estava consciente de um evento é se o indivíduo é capaz de relatar o evento imediatamente após seu acontecimento. O segundo critério requer que o experimentador seja capaz de verificar se o indivíduo relatou o evento corretamente (de acordo com as convenções do experimento). Ambos os critérios precisam ser preenchidos para que a consciência seja inferida. Evidentemente, nem todos os eventos conscientes podem ser verificados pelo experimentador; logo esse não é um critério absoluto para determinar consciência de algo. Entretanto, apenas pela

consideração dos casos em que é possível haver o relato do indivíduo e a verificação do experimentador já é possível o desenvolvimento de uma boa teoria da consciência. Um exemplo dessa prática seria mostrar ao indivíduo uma fruta. Se ele conseguir identificar corretamente qual fruta observou, o experimentador, que também pode verificar, poderá julgar que o indivíduo estava consciente da fruta. Mas o que dizer de casos nos quais foram muitos objetos apresentados para que o indivíduo relatasse todos? Casos nos quais o indivíduo afirma ter tido consciência dos objetos, mas não se lembra de todos para poder relatar? Sperling (1960) resolveu este problema. Basta apenas fornecer um grupo de itens composto de elementos apresentados anteriormente e elementos não apresentados. Por essa técnica, os indivíduos conseguem lembrar quais elementos já haviam sido apresentados, garantindo que houve consciência destes anteriormente, pelo preenchimento do segundo critério.

Diversas discussões sobre a consciência têm utilizado casos confusos, nos quais é difícil distinguir se há ou não consciência; Dennett (1991) discute bastante esse tipo de caso. Baars (1988) critica esta metodologia, argumentando que toda ciência parte do estudo de casos simples primeiramente, para apenas posteriormente, com teorias já elaboradas, tentar resolver casos difíceis. Por esse motivo o psicólogo resolve utilizar evidências de casos nos quais é fácil determinar se houve ou não consciência de um objeto no experimento. Alguns exemplos de fenômenos claramente conscientes são percepções facilmente relatadas, imagens mentais claras, fala interna voluntária e conteúdo evocado da memória de longo-prazo. Exemplos de fenômenos claramente inconscientes são regras sintáticas usadas durante a fala, conteúdo não evocado da memória de trabalho e habilidades automáticas. Exemplos de casos de difícil determinação são conceitos abstratos, percepções periféricas e itens ativos mas não ensaiados na memória imediata.

Baars (1988) identifica três tipos de representações inconscientes através de estudos tradicionais da psicologia: episódicas, linguísticas e habituais. A primeira diz respeito a memórias não atendidas no momento, mas que devem estar representadas inconscientemente por serem trazidas à consciência se necessário. A segunda se trata de regras linguísticas que dificilmente conseguimos manter consciente. Na leitura, por exemplo, se focarmos na sintaxe e nas próprias letras conscientemente, dificilmente conseguimos absorver a semântica. Dessa forma, essas regras precisam ser representadas inconscientemente. A terceira se refere ao fenômeno da habituação

(mencionado acima), no qual um estímulo perde sua força na consciência com a prática e automaticidade, e até desaparece com o domínio da tarefa.

O inconsciente também realiza processos sobre essas representações. Uma das hipóteses de Baars (1988) e também de Dennett (1991) é que o sistema nervoso é composto de um conjunto de processadores especializados que operam de forma inconsciente. Baars (1988, p.50, tradução nossa), baseando-se em Fodor (1983) e Rozin (1975), define um processador como “uma coleção organizada, relativamente unitária, de processos que trabalham juntos em serviço de uma função particular⁴”. Existem evidências neuropsicológicas e psicológicas para esta hipótese. As principais evidências neuropsicológicas dizem respeito à modularidade do cérebro. A partir de lesões cerebrais e deformações cerebrais resultadas de erros genéticos descobriu-se que existem áreas especializadas no cérebro para funções como produção da fala, compreensão da fala, reconhecimento de faces, visão, noções espaciais, planejamento, emoções, dentre outras. Isso foi confirmado pela perda de funções específicas correlacionadas sempre com a mesma estrutura anatômica deficientes nesses pacientes, enquanto preservando funções gerais.

Acreditamos que o cérebro seja uma rede integrada a ser compreendida de forma sistêmica para possibilitar novas descobertas, por exemplo, a respeito de sua forma de processar, porém é difícil negar a modularidade do órgão, não podemos negar fatos neuropsicológicos para melhor adequar perspectivas teóricas.

De acordo com Baars (1988), as evidências psicológicas para existência de processadores inconscientes vêm basicamente de duas fontes: do estudo da automaticidade de tarefas e de estudos de erros em percepção, memória, linguagem, ação e conhecimento. Ao dirigir um carro pela primeira vez uma pessoa precisa estar consciente de todas as subtarefas a serem realizadas, como passar a marcha, direcionar o volante e monitorar o trânsito de veículos. Com a prática, apenas subtarefas que envolvem mais novidades precisam ser mantidas conscientes (monitorar o trânsito de veículos), enquanto as demais deixam de ser conscientes. Se alguém perguntar para uma pessoa em qual marcha ela estava há dois minutos antes da pergunta, possivelmente ela não lembrará. Estes fatos da automaticidade com a prática da tarefa servem de evidência para a existência de processadores inconscientes que se especializaram na realização da tarefa.

⁴ A processor can be defined as a relatively unitary, organized collection of processes that work together in the service of a particular function.

Para exemplificar as evidências em erros de processamento, será utilizado o caso da linguagem. O processamento da linguagem parece ser decomposto em partes realizadas por sistemas distintos. Por exemplo, parece haver sistemas separados responsáveis por erros na sintaxe, na semântica e na análise de grafemas. Crianças com dificuldades em linguagem podem apresentar erros em apenas uma dessas categorias. Ainda, parece haver sistemas distintos para produção e para compreensão da fala. Novamente, esses erros podem ser associados às estruturas anatômicas específicas no cérebro. No caso da compreensão, a área de Wernicke está associada, e no caso da produção, a área de Broca, ambas localizadas apenas no hemisfério esquerdo.

De acordo com Baars (1988), as pessoas reportam experiências conscientes de duas formas: qualitativas ou não-qualitativas. As primeiras são percepções, imagens mentais, sentimentos, dentre outras. Estas possuem alguma forma de *qualia*, como calor, cor e sabor. Já as experiências não-qualitativas não possuem estes atributos, mas podem ser considerados conscientes por preencherem os dois critérios para consciência mencionados acima. Crenças, expectativas ou intenções, por exemplo, parecem não ter *qualia* associado. Após algum desenvolvimento de sua teoria, Baars (1988) sugere que a consciência possivelmente é qualitativa por natureza, e que abstrações relatadas sem *qualia* podem ter imagens fracas relacionadas não percebidas pelo sujeito, mas admite precisar de mais pesquisas empíricas para verificar esta hipótese.

Baars (1988) utiliza a definição de consciência apresentada primeiramente por Dennett (1978), a saber: a consciência é aquilo a que nós temos acesso. Mais especificamente, o que está consciente é aquilo a que a pessoa possui acesso em um dado momento. A inconsciência, em contraste, é tudo aquilo a qual não temos acesso em dado momento. Neste momento, o leitor provavelmente não possui acesso à localidade da moradia de sua mãe, mas imediatamente após ler esta frase, o acesso é obtido. Note que esta definição de consciência é homuncular por usar a palavra “nós” na própria definição. Porém, para resolver este problema, a estratégia do Funcionalismo homuncular é utilizada tanto por Dennett quanto por Baars. Logo, ao invés de deixar o homúnculo com muito poder e sem detalhamento, ambos os pesquisadores explicam esse “nós” ou *self* através de sistemas funcionais que se adequam às respectivas teorias. É possível notar também que os itens aos quais nós temos acesso em um dado momento são extremamente limitados se comparados aos itens aos quais não temos acesso. Isso evidencia uma das características mais curiosas e interessantes da consciência de acordo com Baars (1988), sua capacidade limitada.

Baars (1988) comenta que o cérebro é um órgão complexo, enorme, responsável por tantas tarefas computacionais e mesmo assim nossa capacidade consciente parece limitada. Existem três fontes de evidências para essa limitação: experimentos sobre atenção seletiva, exercícios de dupla tarefa e testes com a memória imediata.

Em experimentos sobre atenção seletiva o sujeito deve monitorar um fluxo de informação complexo, como um vídeo de um jogo de basquete em câmera acelerada. Nessas condições, as pessoas ficam altamente inconscientes de fluxos de informação alternativos até se apresentados para o mesmo órgão. Em estados de absorção, por exemplo, ao assistir a um filme cativante, as pessoas parecem ignorar o ambiente, como se tornassem parte do próprio filme. Estes estados de atenção seletiva mostram que as pessoas só podem ficar conscientes de um fluxo de eventos denso e coerente por vez. Isso não significa que os outros estímulos do ambiente não sejam processados, pois a pessoa reage a um estímulo significativo fora do fluxo denso de eventos, como por exemplo, ao ouvir seu nome. Assim, percebe-se que um monitoramento do ambiente continua, porém esse monitoramento é inconsciente, tornando-se consciente apenas se um estímulo significativo for identificado. Inicialmente acreditava-se haver um filtro para economizar capacidade computacional, porém se a informação é processada tanto no canal da atenção quanto no canal inconsciente, perde-se o propósito de economizar capacidade computacional (BAARS, 1988).

Em exercícios de dupla tarefa os sujeitos são conduzidos a realizarem duas tarefas ao mesmo tempo. Um exemplo seria reagir o mais rápido possível a um estímulo visual no momento em que começaram a verbalizar uma sentença. O resultado é que ambas as tarefas são prejudicadas pela competição. Entretanto, se uma das tarefas se tornar previsível e automática, esta se torna mais inconsciente, e quanto mais inconsciente uma tarefa ficar, menos ela interfere com a tarefa consciente. Este é um caso de interferência não específica, pois uma tarefa é visual e a outra linguística. O efeito ainda ocorre se ambas as tarefas forem da mesma modalidade. Em todos os casos, pelo menos uma tarefa se automatiza com a prática, se torna inconsciente e, portanto, deixa de interferir com a outra (BAARS, 1988).

A técnica de Sperling (1960), além de solucionar o problema do segundo critério para considerar um evento consciente, também evidencia a memória sensorial, um componente da memória imediata. Ao observar diversas letras em pequenas frações de tempo, apenas algumas podem ser relatadas, porém as demais parecem estar em uma memória sensorial, pois os sujeitos conseguem apontar quais letras haviam sido

observadas. Baars (1988) acrescenta que o componente mais bem estudado da memória imediata é chamado de memória de curto-prazo. Esta memória ensaia itens linguísticos como palavras e números de telefone. Uma das maiores evidências para a capacidade limitada é que esta memória tem um limite extremamente curto para a quantidade de itens não relacionados que pode guardar. Miller (1956) mostrou como, ao ensaiar itens, conseguimos recordar cerca de sete, com uma variação de mais ou menos dois itens, enquanto sem ensaiar, apenas três ou quatro são lembrados em média. Ainda, o limite também se mostra no tempo de duração, sendo cerca de dez segundos sem ensaio.

Baars (1988) comenta uma capacidade interessante da memória de curto-prazo. Ele pede para considerarmos o número “677124910091660128417891”. Esse número ultrapassa nossa capacidade de sete itens. Apesar disso, no exame da sequência ao contrário, identificamos seis datas: 1987, 1482, 1066, 1900, 1942, 1776. Unir (*chunking*) os itens em grupos com significado faz nossa memória de curto-prazo expandir sua utilidade. Ainda, isto sugere uma ligação com a memória de longo-prazo (toda a quantidade de informação que um indivíduo conhece e pode evocar), pois esses anos só podem ser identificados por um conhecimento prévio.

A memória imediata não se iguala à consciência. Estamos conscientes apenas de itens ensaiados na memória de curto-prazo e de itens evocados da memória sensorial. Assim como eventos periféricos do campo visual, esses itens inconscientes podem vir à consciência se forem significativos. Portanto, a memória imediata guarda um pouco mais informação do que está sendo usada em dado momento na consciência (BAARS, 1988).

Baars (1988) considera um contra-argumento à hipótese da capacidade limitada para a consciência. Ao olhar para uma cena rica de um jogo de futebol, nossa experiência pode ser extremamente complexa e detalhada e ainda assim completamente consciente. O psicólogo justifica que isto ocorre porque cada parte da cena ajuda a prever o restante; se diversos eventos não relacionados forem mostrados por experiência visual, novamente o limite da consciência se mostra.

A capacidade limitada é uma característica essencial da consciência que está relacionada com todas as demais, mas, ainda assim, não é a única. Seguindo o método da análise contrastiva, Baars (1988) demonstra as diferenças entre processadores especializados inconscientes e processos conscientes. A tabela a seguir expõe o contraste.

Processos Conscientes	Processadores Inconscientes
Computacionalmente Ineficiente: maior quantidade de erros, baixa velocidade, interferência entre dois ou mais processos.	Altamente eficientes em suas tarefas específicas: poucos erros, alta velocidade, pouca interferência entre processadores.
É multimodal, capaz de relacionar essas diversas modalidades entre elas e com conteúdos inconscientes.	Realizam tarefas específicas, sendo autônomos e relativamente isolados.
Apresentam consistência interna, serialidade e capacidade limitada.	Funcionam em paralelo e juntos possuem grande capacidade.

Tabela 1 – Realizada pelo autor baseada em tabela semelhante em Baars (1988, pg.75).

Ao tentarmos identificar a sintaxe de uma sentença conscientemente (separar verbo, adjetivos, sujeito, predicado), percebemos que realizamos estas tarefas com baixa velocidade, apresentando erros e interferindo com nossa capacidade de compreender a semântica. Apesar disso, em qualquer conversa cotidiana, inconscientemente conseguimos fazer essas diferenciações sintáticas com extrema facilidade e rapidez enquanto realizando outras tarefas. Por isso, Baars (1988) afirma que esse tipo de contraste mostra como processos conscientes são computacionalmente ineficientes, apresentando erros, baixa velocidade e interferência com outras tarefas; enquanto os processadores inconscientes especializados são altamente eficientes em suas tarefas específicas, realizando-as com poucos erros, de forma veloz e não interferindo com o sucesso de outras tarefas.

Baars (1988) afirma que a consciência é multimodal, podendo trabalhar com todos os tipos de percepções sensoriais, e também com imagens mentais, sonhos, fala interna, sentimentos, prazeres e dores. Com a inclusão de conteúdos não qualitativos como crenças, conceitos e intenções, as possibilidades são ainda maiores. Baars (1988, p.77, tradução nossa) afirma que “podemos ganhar um grau de controle consciente

sobre quase qualquer população de neurônios, desde que recebamos retroalimentação consciente imediata da atividade neural⁵”.

Em contraste com essa multimodalidade da consciência, cada processador é especializado em uma dada tarefa; logo é limitado a realizar apenas funções específicas. Foi mencionado anteriormente como evidências neuropsicológicas e das subdivisões do processador de linguagem mostram essa limitação. Uma forma recente de mostrar essa especialização é por um experimento inovador (Nishimoto *et al.*, 2011) no qual apenas as atividades dos neurônios do córtex occipital foram monitoradas enquanto sujeitos assistiam a filmes. Após muito estudo dessas atividades, os cientistas conseguiram reconstruir a imagem do filme apenas por conhecer a atividade desses neurônios. Baseados apenas na imagem reconstruída através da atividade do córtex occipital Nishimoto *et al.* (2011) conseguiram dizer qual filme certo paciente estava assistindo. Se o córtex occipital for lesionado, dificilmente outras áreas do cérebro conseguirão assumir sua função, dada a limitação de função dos processadores inconscientes. Existe a plasticidade cerebral, a capacidade de alguns neurônios exercerem a função de outros após lesão, porém essa capacidade infelizmente é limitada.

Razran (1961) mostrou que na associação de estímulos no estilo Pavloviano, no qual há associação de dois estímulos pela sua ocorrência em conjunção constante (por exemplo, o som de um sino e o momento da alimentação), ambos os estímulos precisam estar conscientes. Se um dos estímulos estiver habituado, a associação não ocorre. Baseado nesta e em outras pesquisas semelhantes, Baars (1988) chega à conclusão de que apenas a consciência tem a capacidade de relacionar estímulos arbitrários, pois se um deles estiver inconsciente isto não acontece.

De acordo com Baars (1988), os processadores inconscientes podem ser incoerentes entre si. Um sistema pode estar oferecendo certas interpretações sobre o mundo enquanto outro está ao mesmo momento discriminando os mesmos eventos de outra forma. Por vezes, apresentamos comportamentos e discursos incoerentes. Uma mesma palavra pode estar representada inconscientemente com mais de um significado. Já a experiência consciente parece ser sempre coerente, apresentando consistência interna. Por exemplo, não conseguimos entender diversos conteúdos semânticos ao mesmo tempo, precisamos focar em um fluxo de linguagem por vez. O mesmo ocorre com a percepção, já que conseguimos ver apenas uma imagem coerente por vez; um

⁵ [...] we can gain a degree of conscious control over virtually any population of neurons, provided that we receive immediate conscious feedback from the neural activity

cubo pode ser enxergado de duas formas distintas, dependendo de qual de seus quadrados está sendo enxergado como “mais distante”. Em ilusões de ótica com figuras nas quais é preciso atenção para chegar ao efeito, apenas um de dois estados é possível em um mesmo momento: a percepção normal da figura ou a percepção com efeito de ilusão de ótica. Outro exemplo na linguagem é a impossibilidade de se pensar em dois sentidos da mesma palavra no mesmo exato momento.

Este último exemplo está relacionado com mais uma característica da consciência, sua serialidade. Baars (1988) afirma que quando utilizamos a consciência para aprender, precisamos dedicar nosso foco para cada etapa do processo por vez. Precisamos organizar toda a tarefa da uma forma serial. Quando alguém aprende a tocar bateria, primeiramente consegue manter apenas um ritmo para todos os quatro membros do corpo. Com a prática e automatização, é possível aplicar até quatro ritmos, um para cada membro do corpo. Para a Teoria do Espaço de Trabalho Global, nem todos os ritmos precisam estar conscientes, alguns são executados inconscientemente, sendo a consciência aplicada para o controle serial de detalhes críticos e de alguns passos principais, como mudança de um refrão. Dennett (1991) chega a afirmar que a consciência depende de uma Máquina de Turing virtual dada essa serialidade (explicada posteriormente). Inclusive, a capacidade de unir sequências de números complexos (como o exemplo das datas mencionado acima) para serem processadas como apenas um item, também é uma característica da Máquina de Turing. Em sua aplicação física, o computador, ao encontrar uma sequência de 32bits (por exemplo) conhecida, ela lê esta como se fosse um item apenas e impulsiona um comando. A incapacidade de realizar várias tarefas conscientes ao mesmo tempo também mostra a serialidade da consciência.

Há um debate sobre como os processadores inconscientes operam, mas parece que alguns processam em serial e outros em paralelo. Por exemplo, Pinker (1994) e Colheart *et al.* (2001) argumentam em favor de funcionamentos seriais para o processamento da linguagem, enquanto Caramazza (1997), a partir de modelos computacionais conexionistas, e Dennett (1991) a partir de argumentação, mostram como a linguagem pode ser processada em paralelo. De toda forma, a estrutura anatômica do cérebro sugere fortemente um processamento em paralelo. Baars (1988) acredita que a forma mais simples de resumir as evidências é afirmando a serialidade da consciência em contraste com o funcionamento em paralelo dos processadores inconscientes.

A última característica citada da consciência é a capacidade limitada, a qual foi discutida acima por meio das evidências em atenção seletiva, exercícios de dupla tarefa e estudos com a memória imediata. Em contraste, os processadores inconscientes juntos apresentam alto potencial. A maior parte de nossas habilidades automáticas e as nossas memórias guardadas a longo-prazo são inconscientes. Também a neurofisiologia indica a enorme quantidade de sinapses sendo ativadas em um mesmo momento, e evidentemente, apenas uma pequena parcela dessas estão relacionadas com a consciência (BAARS, 1988). Se considerarmos a consciência como um subsistema do cérebro, fica claro que todos os outros subsistemas em conjunto terão maior potencialidade.

Com essas noções básicas necessárias para a teoria de Baars (1988), agora será apresentado como o psicólogo utiliza essas evidências e caracterizações dos processamentos inconscientes e conscientes para desenvolver sua Teoria do Espaço de Trabalho Global, uma teoria cognitiva da consciência.

2.2 A Teoria do Espaço de Trabalho Global

A Teoria do Espaço de Trabalho Global de Baars (1988) parte de uma metáfora que incorpora outras quatro hipóteses metafóricas já propostas na literatura: A hipótese da ativação, a hipótese da novidade, a hipótese da ponta do iceberg e a hipótese do teatro.

A hipótese da ativação sugere que a consciência está relacionada a um ápice ou uma forma de ativação de elementos na memória de tal forma a chegar a um limiar. Existem evidências para essa hipótese em estudos de percepção, imagem mental e memória. Quando há estímulos ambientais de baixa intensidade, não há consciência desses, e quando dois estímulos provocam a mesma associação, há mais chance dele ser consciente do que quando um estímulo apenas promove a associação. O alto nível de ativação pode até ser uma condição necessária, mas não é uma condição suficiente para a consciência. Isso se dá pelo fenômeno da habituação e automatização, os quais mostram que perdemos consciência de estímulos repetidos e previsíveis. Logo, o grau de ativação pode também levar à inconsciência, gerando um paradoxo na definição (BAARS, 1988).

O paradoxo pode ser resolvido pela hipótese da novidade, para a qual fenômenos conscientes não estão relacionados apenas com alto grau de ativação, mas também com

a novidade, a quebra de hábitos e com a comparação. A Teoria do Espaço de Trabalho Global de Baars (1988) incorpora essa ideia por meio de uma adequação da teoria de Informação de Shannon e Weaver (1949), apresentando uma proposta inovadora para a relação entre a novidade e a consciência (descrita na próxima seção).

A hipótese da ponta do iceberg sugere que a consciência resulta de uma grande quantidade de processos inconscientes. Essa metáfora está relacionada com as ideias da capacidade limitada e de que temos pouco acesso ao processamento do cérebro. A hipótese reitera o fato de nossa consciência representar apenas uma pequena porção de nossa atividade mental.

A última hipótese é conhecida como a hipótese do teatro e Baars (1988, p.31 tradução nossa) utiliza de uma descrição feita pelo filósofo e historiador Hyppolite Taine (1828-1893) para descrevê-la. Esta hipótese será uma questão central para o restante desta dissertação e, portanto, precisa ser descrita exatamente da forma apresentada:

É possível, portanto, comparar a mente de um homem a um teatro de profundidade indefinida com uma cortina estreita, mas com um palco que se estende após esta. Nesta cortina iluminada há espaço para apenas um ator. Ele entra, gesticula por um momento, e sai; outro chega, e depois outro e assim por diante... No mesmo cenário, em uma parte distante, ou antes das luzes da cortina, evoluções desconhecidas tomam lugar incessantemente entre essa massa de atores de todos os tipos, para preparar as estrelas que se apresentam aos nossos olhos um por vez, como uma lâmpada mágica⁶.

Baars (1988) comenta a passagem apontando como o filósofo incorpora ideias centrais para a consciência. Primeiro, ao dizer da ordem de passagem dos atores, ele faz lembrar a serialidade do processamento consciente. Ao se referir a uma parte distante do palco, ou antes das luzes da cortina, ele incorpora a metáfora da ponta do iceberg, por dizer que há mais acontecendo em nosso processamento do que podemos perceber. A preparação das estrelas antes das luzes da cortina parece se referir a um controle exercido pelos mecanismos inconscientes sobre as experiências conscientes, ideia relacionada com a noção de ‘contexto’ na Teoria do Espaço de Trabalho Global. A

⁶ One can therefore compare the mind of a man to a theatre of indefinite depth whose apron is very narrow but whose stage becomes larger away from the apron. On this lighted apron there is room for one actor only. He enters, gestures for a moment, and leaves; another arrives, then another, and so on ... Among the scenery and on the far-off stage or even before the lights of the apron, unknown evolutions take place incessantly among this crowd of actors of every kind, to furnish the stars who pass before our eyes one by one, as in a magic lantern.

metáfora incorpora ainda a hipótese da ativação, como se os atores precisassem de algum tipo de *momentum* para entrar em cena.

Uma metáfora similar compara a consciência a uma luz de farol iluminando a cada momento um pequeno grupo de elementos no sistema nervoso. Baars (1988) afirma que esta metáfora não acrescenta muito à do teatro, mas acreditamos que ela deixa clara a relação da consciência com a memória de curto-prazo. Esta pode expor a memória de curto-prazo como mais extensa e a consciência focando sua luz em alguns elementos desta memória a cada momento.

Baars (1988) compreende o sistema nervoso enquanto um grupo de processadores especializados e inconscientes trabalhando em paralelo, portanto, tem como premissa a ausência de um processador central para dar ordem ou forçar alguma atividade sobre outros processadores. Cada grupo de processadores pode ter uma função executiva, dependendo do trabalho, e podem trocar seus papéis de forma flexível. Mas o psicólogo argumenta que mesmo sem um executivo verdadeiro, o conjunto de processadores especializados precisa de algum tipo instituição central (*central facility*) para haver a comunicação entre os diversos especialistas. Esta instituição central não confere ao sistema algum tipo de poder absoluto, pois esta é mais semelhante a uma reunião de uma assembleia; em termos cognitivos, se trata de um espaço de trabalho. Um espaço de trabalho (*workspace*) é apenas uma memória, na qual diferentes sistemas operam; a palavra “global” implica que símbolos nessa memória são distribuídos para uma variedade de processadores. Com essa noção, cada processador pode ter variáveis e operações locais, mas também pode chegar a fornecer símbolos globais.

Baars (1988) tenta criar novas metáforas para incorporar essa ideia, junto às demais metáforas citadas e às evidências sobre processadores inconscientes e processos conscientes. O psicólogo menciona diversas vezes a comparação entre o espaço de trabalho global e uma estação de televisão enviando sinais para um país inteiro. Mas a analogia considerada por ele mais apta se trata de um comitê de especialistas formando uma assembleia na qual nenhum elemento em si consegue resolver o problema desejado sozinho. Como cada especialista tem sua própria forma de compreender e de comunicar o problema, há de início um problema de comunicação entre eles. A solução emerge da utilização de um quadro negro para publicar mensagens globais, para que cada especialista possa entender o que os outros estão fazendo. Como Baars (1988, p.87, tradução nossa) explica:

Um passo importante para resolver esse problema de comunicação é pela publicação de uma mensagem global em um grande quadro negro na frente do auditório, para que em princípio qualquer um possa ler a mensagem e reagir. De fato, a mensagem seria lida apenas por especialistas capacitados para entendê-la ou entender partes dela, mas ninguém pode saber, antes do tempo chegar, quem serão esses especialistas, de forma a se tornar necessário que a mensagem seja potencialmente disponível para qualquer um na plateia. A qualquer momento, um número de especialistas pode tentar enviar mensagens globais, mas o quadro negro não pode acomodar todas as mensagens de uma mesma vez – mensagens diferentes serão comumente mutualmente contraditórias. Dessa forma, alguns dos especialistas poderão competir por acesso no quadro negro, e alguns poderão cooperar em um esforço para publicar a mensagem global⁷.

Baars (1988) acrescenta que esse tipo de situação possui vantagens e desvantagens. Um quadro negro não será usado quando o problema é simples, conhecido ou quando uma ação rápida é necessária. Mas é uma forma efetiva de resolver problemas quando há necessidade de cooperação entre fontes distintas de conhecimento, com tempo para e vantagens em decidir sobre as possibilidades.

Essa última metáfora de Baars (1988) pode incorporar as outras quatro anteriores. Apenas especialistas com certo grau de ativação, nesse caso, com informação relevante para ser publicada, terão chances de acesso ao quadro negro. A metáfora pode incorporar a hipótese da novidade, acrescentando a regra de uma prioridade para conteúdos inovadores para serem publicados, já que conteúdos antigos não seriam vantajosos para publicação. Ela está coerente com a metáfora da ponta do iceberg, pois a consciência estaria relacionada ao conteúdo do quadro negro e não a toda informação disponível na assembleia. Finalmente a hipótese do teatro se encaixa perfeitamente à nova hipótese, sendo que o teatro é substituído pelo quadro negro e os atores atrás da cortina pelos especialistas realizando a assembleia.

A linguagem metafórica proposta por Baars (1988) é puramente ilustrativa, sendo uma forma de esclarecer e ajudar o leitor a imaginar a forma de processar desses sistemas; o uso puro de um vocabulário científico pode dificultar a construção dessa imagem clara. No capítulo seguinte será discutido se essas metáforas podem enquadrar a teoria de Baars no conceito de Teatro Cartesiano de Dennett (1991); se elas realmente

⁷ One helpful step to solve this communication problem is to make public a global message on a large blackboard in front of the auditorium, so that in principle anyone can read the message and react. In fact, it would only be read by experts who could understand it or parts of it, but one cannot know ahead of time who those experts are, so that it is necessary to make it potentially available to anyone in the audience. At any time a number of experts may be trying to broadcast global messages, but the blackboard cannot accommodate all of the messages at the same time --- different messages will often be mutually contradictory. So some of the experts may compete for access to the blackboard, and some of them may be cooperating in an effort to broadcast a global message.

acabam revelando pressupostos filosóficos adotados por Baars ou se podem gerar interpretações equivocadas em relação à teoria exposta mais formalmente.

Trocando os membros da assembleia na metáfora por processadores inconscientes especializados e o quadro negro por uma memória de trabalho denominada espaço de trabalho global, Baars (1988) chega a um modelo simples, capaz de abarcar os contrastes realizados entre processadores inconscientes e conscientes. Um espaço de trabalho global realiza uma troca de informação que permite a interação de sistemas inconscientes especializados do sistema nervoso. Assim, é um sistema de difusão (“*broadcasting*”) da informação para o cérebro. Diversos especialistas inconscientes podem competir ou cooperar para obter acesso ao espaço de trabalho global. Após obtenção de tal acesso, podem transmitir suas informações para todos os outros sistemas.

A teoria abarca a afirmação de processos conscientes serem computacionalmente ineficientes, pois computações processadas por um espaço de trabalho global exige ativação de todos os processadores relevantes. Como cada ação precisa de um relativo consenso entre os processadores relevantes, muito mais tempo é gasto em comparação a processos especializados em uma função específica, cujo algoritmo já é conhecido.

Como a mensagem é globalizada, ela precisa atingir todos, ou a maioria dos sistemas inconscientes, utilizando algum tipo de código entendido por todos. Isso incorpora a multimodalidade dos processos conscientes em contraste com a especificidade dos processadores inconscientes. Essa globalização também permite maior relacionamento entre conteúdos, algo difícil para processadores inconscientes com dificuldade de comunicação.

A consistência interna da consciência também é incorporada à teoria. A mensagem presente no espaço de trabalho global em certo momento precisa ser internamente consistente, pois é necessário certo consenso entre os especialistas, ou uma força maior de um grupo que esteja competindo com outro para o acesso. Portanto, existe uma competição ou cooperação para acesso ao espaço de trabalho que precisa globalizar uma mensagem coerente. Da necessidade de mensagens únicas e coerentes, emerge naturalmente a serialidade dos processos conscientes, pois cada mensagem é globalizada uma por vez.

Como a mensagem precisa ser internamente coerente, ela é limitada a poucos itens, pois os itens incompatíveis com a mensagem não podem ser globalizados. O uso

do espaço de trabalho global é, portanto, reservado para solucionar apenas os problemas que não podem ser resolvidos por processadores inconscientes especializados funcionando isoladamente.

Assim o espaço de trabalho global possui três características centrais. Ele promove *disponibilidade global*, de mensagens *internamente consistentes*, que precisam ser *informativas* para o sistema. Baars (1988) menciona mais duas características do espaço de trabalho global, as quais não serão exploradas neste trabalho, a relação deste sistema com o *self* e o acesso privilegiado de percepções.

Baars (1988) sugere um correlato neural desse sistema no cérebro. Atualmente, acreditamos que o psicólogo tenha delegado este papel a biólogos com mais experiência na área e até sugere uma ligação do espaço de trabalho global com os correlatos neurais da consciência descritos por Edelman (Edelman *et al.* 2011). Mas, de toda forma, é interessante apresentar esta primeira hipótese de Baars, pois ela revela um pouco sobre seu entendimento de como a consciência funciona no cérebro.

Baars (1988) sugere que o correlato neural da consciência precisa estar ligado a funções como atenção, habituação, respostas de orientação e o estado de vigília. Este correlato neural precisa receber *inputs* de todo o cérebro e ter *outputs* que possam servir como um meio de globalizar informação. Baars (1988) atribui o funcionamento do espaço de trabalho global a um conjunto de estruturas chamado por ele de Sistema de Ativação Reticular-Talâmico Estendido. Esse sistema tem início na formação reticular, a qual recebe informação (*input*) de todas as maiores estruturas do cérebro (pelas suas ligações com o tálamo), e se estende até o núcleo inespecífico do tálamo. De acordo com Machado (2000), a estimulação deste núcleo modifica os potenciais elétricos de territórios muito grandes do córtex cerebral. O sistema talâmico de projeção difusa exerce um papel de *output* do espaço de trabalho global, através de suas conexões entre o núcleo inespecífico do tálamo com as grandes regiões do córtex. Assim, os especialistas inconscientes distribuídos pelo cérebro competem por acesso ao Sistema de Ativação Reticular-Talâmico Estendido, o qual globaliza a informação privilegiada. A estrutura anatômica sugere também um fluxo de retroalimentação, abrindo a possibilidade para sistemas retroalimentarem seus interesses para o espaço de trabalho global, fortalecendo ou enfraquecendo alguma mensagem em particular.

O último conceito central para a Teoria do Espaço de Trabalho Global é o contexto, um grupo de processadores inconscientes com acesso privilegiado à consciência, um sistema que molda a experiência consciente sendo ele próprio

inconsciente. A palavra contexto normalmente é utilizada para se referir ao ambiente, o qual afeta nossa cognição. Apesar disso, Baars (1988) utiliza a palavra contexto para se referir às estruturas internas inconscientes, pois o contexto do mundo não influencia nossa cognição se não estiver representado de alguma forma internamente; assim, o psicólogo reduz o contexto ambiental a estruturas inconscientes chamadas também de contexto. Essas estruturas cognitivas estão relacionadas a outras tradicionais da ciência cognitiva, como redes semânticas, planos, roteiros, *frames*. Porém, um novo conceito é necessário para se referir às representações inconscientes agindo para influenciar uma experiência consciente. Este conceito pode ser problemático para a teoria de Baars, discutimos isto nos comentários da próxima seção.

Um exemplo simples de um contexto é de um sistema inconsciente realizando previsões em tempo real para compensar nossa relação em constante mudança com a gravidade e com nossa paisagem visual. Quando entramos em um barco, de início o horizonte parece balançar, mas rapidamente percebemos que o barco está movimentando. Após ficarmos um tempo na água, ao voltarmos para a terra-firme, percebemos o mundo como se estivesse balançando novamente, até que nosso sistema comece a acertar as previsões. Assim, o mundo é percebido de forma estável apenas quando as previsões são bem sucedidas. Essas previsões são completamente inconscientes, e enquanto estiverem sendo bem sucedidas elas não mostram sua existência.

O fenômeno do *Priming* evidencia esses contextos no caso da linguagem. O *Priming* acontece quando uma resposta a um estímulo é influenciada pelos estímulos anteriores. Esses efeitos ocorrem em todas as modalidades sensoriais. No caso da linguagem, um bom exemplo é de uma pessoa lidando com frutas. Esta, ao ler a palavra manga, automaticamente lembra-se da fruta manga. Mas se esta pessoa estiver lidando com camisas e lê a palavra manga, ela automaticamente imagina a parte de uma camisa que cobre os braços. Isso evidencia contextos representando conteúdos recentes inconscientemente. Ao ouvir uma música ou uma conversa, ficamos conscientes apenas de cada parte de uma vez, mas se cada parte não estivesse sendo relacionada a um todo inconsciente, não conseguiríamos compreender mudanças sintáticas relevantes na fala ou distinguir mudanças interessantes em uma música, como a entrada de um refrão (BAARS, 1988).

Os contextos também influenciam a forma como entendemos o mundo conceitualmente. Sua existência também é evidenciada nesse âmbito. Uma visão de

mundo influencia diretamente na forma como entendemos os fatos momentâneos e conscientes. Por exemplo, um cristão está muito propenso a enxergar uma imagem de Jesus em uma torrada, nuvem ou janela, porém um indígena olhando para os mesmos objetos saberá que isto não procede. Esse fenômeno se dá por haverem contextos inconscientes influenciando a forma como compreendemos o mundo, baseado em experiências anteriores. O que acreditamos com muita certeza parece ser óbvio e permanece sem exame consciente. Assim, perdemos foco das alternativas às pressuposições estáveis.

Baars (1988) resume os grupos de contextos em quatro: contextos de percepção e de imagens mentais; contextos conceituais; contextos de meta; e contextos de comunicação compartilhados entre falantes da mesma conversa.

2.3 Consciência e Informação

Os contextos explicados anteriormente estavam mais relacionados às expectativas. O contexto mais relacionado às intenções é chamado por Baars (1988) de “contexto de meta”. Assim como o conhecido psicólogo Maslow (1970), Baars (1988) trabalha com uma hierarquia de metas. Em um dado momento, as metas são ordenadas de acordo com sua relevância, sendo que as mais relevantes predominam sobre outras metas. Por exemplo, no momento, a meta de um indivíduo de ler um texto pode estar no topo da hierarquia, mas se sua casa pegar fogo, certamente outra meta irá dominar. O fluxo da consciência pode ser considerado um fluxo de experiências criadas pela dinâmica de diversos contextos de meta, cada um tentando tornar consciente (através da competição pelo acesso ao espaço de trabalho global) aquilo que garantirá progresso em relação à sua meta.

De acordo com Baars (1988), houve discussões sobre se a correspondência (ou não) de expectativas com estímulos geram a consciência. Ora, se houver correspondência completa entre expectativas e estímulos, o organismo irá se habituar e não haverá consciência. Por outro lado, se houver a não correspondência entre expectativas e estímulos, se o estímulo for tão novo que não se relaciona a nenhum contexto de expectativa, ele também não será compreendido e não será levado à consciência, sendo dado como irrelevante.

Baars (1988) propõe uma solução para esse paradoxo através da noção de informação. A informação envolve tanto uma correspondência com o contexto e uma

não correspondência com detalhes do estímulo. Com essa noção é possível dizer que os estímulos precisam ser úteis, no sentido de reduzir incerteza de contextos. Essa noção de redução de incerteza é uma adaptação da teoria de Shannon e Weaver (1949). Baars (1988) diz utilizar essa ideia em um sentido mais amplo. A informação, para Baars (1988), é pensada em termos de redução de incerteza de um conjunto de escolhas dentro de um contexto estável.

Shannon e Weaver (1949) afirmam que a informação não pode ser confundida com significado, pois uma mensagem com a mesma quantidade de informação pode ser altamente significativa ou não. Informação, nesta definição, é uma medida de liberdade de escolha quando se seleciona uma mensagem. Shannon e Weaver (1949) explicam que o conceito de informação não se aplica a mensagens individuais como o conceito de significado, mas para a situação como um todo. Uma quantidade de informação é medida pelo logaritmo de escolhas possíveis na base de 2. Dessa forma, se houver 16 alternativas de mensagens igualmente possíveis de serem escolhidas, então como $16 = 2^4$ de forma que o $\text{Log}_2 16 = 4$, se diz que essa situação é caracterizada por 4 bits de informação. O “2” representa a possibilidade de respostas numa base binária, o “16” representa o somatório de escolhas possíveis e o “4” representa a quantidade de informação. Na seleção de uma mensagem, se houver muitas escolhas a serem tomadas para reduzir a incerteza, haverá mais informação (bits), mais passos necessários para resolver o processo de seleção. Se não houver nenhuma escolha a ser feita, haverá a redundância, a habituação no caso da consciência. Se houver inúmeras possibilidades de escolha, não haverá como progredir e o problema não terá solução. Baars (1988) tenta burlar a não relação entre o conceito shanoniano de informação com o significado considerando que informação significativa se trata da redução de incerteza dentro de um contexto de meta. Se não houver incerteza a ser reduzida dentro de um contexto relevante, o estímulo se habitua e permanece inconsciente. Se houver tanta incerteza que não encaixa em nenhum contexto, não há como fazer escolhas de mensagens, o estímulo também não se tornará consciente. Dessa forma, um estímulo se torna consciente apenas se trazer novidade produtiva para resolução de incerteza de um contexto.

A consciência ocorre no estágio de aprendizagem no qual há um contexto relativamente estável para entender um estímulo, mas ainda existem incertezas para serem reduzidas dentro deste contexto. O contexto pode processar automaticamente e inconscientemente para influenciar as experiências conscientes seguintes. Como no

exemplo do barco mencionado anteriormente. Se o contexto estiver conseguindo prever a relação de nossos movimentos com os estímulos visuais, ou seja, se não houver incerteza a ser reduzida nesse contexto, a previsão irá funcionar de forma inconsciente. Na medida em que entramos no barco e nosso contexto perceptivo não consegue fazer previsões efetivas da relação entre o movimento do barco e os estímulos visuais, ocorre a incerteza no contexto de percepção relacionado a esta função, logo há consciência desses detalhes até que eles sejam corrigidos e habituados novamente. O acesso de um processador ao espaço de trabalho global é buscado como um apelo de certo contexto para que todos os demais processadores do cérebro ajudem-no a reduzir sua incerteza. Por isso, o acesso ao espaço de trabalho global está relacionado à redução de incerteza.

2.4 Considerações Sobre a Teoria do Espaço de Trabalho Global

A Teoria do Espaço de Trabalho Global é bastante rica e coerente. Mas, também é possível tecermos críticas a esta. Para dizer qual conteúdo estará na consciência em certo momento, Baars depende da noção de hierarquia de metas, mas não há o estabelecimento de um critério de relevância para determinar a hierarquia de metas. Como a consciência está ligada a noção de hierarquia de metas, Baars começa sua explicação de forma detalhada, mas acaba deixando muito sem responder. O conteúdo que estará na consciência depende de uma meta relevante, mas como discernir qual meta é a mais relevante? Baars teria que dizer como uma meta ganha maior relevância do que outra, e quando o fator é interno, está relacionado com uma dinâmica entre emoções, comportamentos sociais e desenvolvimento da personalidade. Delegar a função para emoções, para comportamentos sociais e personalidade seria como criar homúnculos problemáticos para a teoria (por não haver explicação desses conceitos na mesma). Uma possível solução para esse aspecto da teoria de Baars (1988) seria uma união com teorias motivacionais, as quais estabelecem critérios para hierarquias de metas, como a do próprio Maslow (1970). Ainda, ao tentar relacionar novidade e significado com a teoria matemática da comunicação, Baars (1988), possivelmente sem saber, enfrenta uma tradição filosófica extensa que afirma que isto é uma impossibilidade.

Possivelmente há um problema com a própria ideia de contexto que Baars descreve. Por ser um “parente próximo” do conceito de *frame*, o contexto também carrega as consequências do “*Frame problem*” (PYLYSHYN, 1987). Não é possível

que fatores internos sejam os principais responsáveis para determinar quais conteúdos serão os próximos requisitados pela consciência. Quando estamos em um estádio de futebol existe muita informação visual para ser captada e precisamos fazer um esforço para focar nos movimentos da bola e dos jogadores. Esta situação vai escolher para entrar na consciência o resultado de processadores especializados em forçar atenção visual em coisas distantes. Já se estivermos assistindo uma palestra, a situação vai escolher a ativação de interpretadores semânticos e de pensamentos sobre o conteúdo da apresentação. Assim, é primariamente o mundo, ou a mente situada neste, que guia a sequência de eventos conscientes.

Baars (1988) dá uma ênfase para a serialidade da consciência, mas admite a possibilidade desta não ser absoluta. Em frações de segundo, admite que pode haver certo paralelismo dos conteúdos. Mas acredita que a forma mais simples de resumir as evidências é dizendo que a consciência trabalha em serial enquanto a inconsciência em paralelo. Mas com isso, Baars parece estar igualando a consciência demasiadamente à atenção, no sentido de que podemos estar conscientes apenas daquilo que estamos prestando atenção. É uma possibilidade, entretanto, a consciência parece ter uma segunda “camada” para eventos que estamos cientes (*aware*), mas sem prestar atenção. Baars diria que inconscientemente contextos estão acompanhando esses fenômenos de fronteira. Mas essa saída parece estabelecer uma dicotomia muito forte entre consciência e inconsciência, talvez resultado de seu próprio método de estudo, a análise contrastiva.

Apesar de alguns problemas, como estes brevemente mencionados acima, a teoria de Baars (1988) é extremamente interessante e segue atualmente como uma das principais teorias cognitivas da consciência. Postulando apenas três tipos de sistemas (o espaço de trabalho global, processadores inconscientes e contextos), Baars consegue incorporar todo tipo de fenômeno consciente em um simples modelo funcional de processamento de informação. A seguir serão expostas as visões de Dennett (1991) sobre a consciência, iniciando com sua metodologia.

2.5 A Heterofenomenologia

De acordo com Dennett (1991), pode ser comum considerar as evidências vindas da fenomenologia como auto-evidentes, por ser uma forma do mundo se apresentar aos sentidos. Assim, acreditava-se que poderíamos por introspecção contemplar os mistérios

da consciência e dividir essas experiências com o outro, com a ideia de que todos compartilham deste mesmo mundo fenomenológico. Dennett (1991) argumenta que talvez estejamos enganados em atribuir um infalibilismo rígido em relação à nossa capacidade de introspecção. O filósofo acredita que o maior erro desta doutrina está em achar que a introspecção é uma forma de observação e de constatação; de acordo com ele, a introspecção é mais uma forma de teorização sobre a realidade, e não um método de auto-evidência. Os efeitos de ilusão de ótica e efeitos resultados de lesões neuropsicológicas nos levantam dúvidas quanto aos resultados da fenomenologia. Dennett (1991) acredita em um acesso privilegiado à nossa própria fenomenologia, mas esse acesso está longe de ter um poder para constatar verdades.

Dennett (1991) argumenta que como a ciência funciona sempre por uma perspectiva de terceira pessoa, uma ciência da mente também precisa ser feita por essa perspectiva. No Behaviorismo e também em algumas pesquisas da neurociência, nas tentativas de excluir o método introspectivo se excluía também a mente, por ser impossível estudar esta por uma perspectiva de terceira pessoa. Atualmente, esta tese é defendida principalmente por Nagel (1974).

Dennett (1991) mostra como nem tudo para ser estudado na ciência precisa ser atingível por observação direta. Se o objeto inobservado interferir em outros processos observados, sua postulação pode ser realizada. O filósofo usa o exemplo de buracos negros e genes. Um exemplo recente de um objeto inferido indiretamente é a matéria negra, a qual por definição não pode ser observada, mas pode ser inferida por seus efeitos gravitacionais em planetas. A mente é um objeto ainda mais facilmente inferido do que a matéria negra, pois além dos efeitos comportamentais e cerebrais, suas consequências fazem parte de nosso cotidiano (diferentemente da matéria negra). Então, afirmar que a mente não pode ser estudada por uma perspectiva em terceira pessoa é falso. A solução da Neurociência cognitiva para esse problema é a de inferir a mente relacionando-a a eventos cerebrais e relatos verbais. Dennett (1991) desenvolveu um método semelhante, denominado Heterofenomenologia.

O estudo da consciência é um dos poucos estudos científicos que exige comandos (dos experimentadores) e relatos (dos sujeitos) verbais. Uma forma de garantir a fidedignidade do relato é pela observação de múltiplos experimentadores e gravadores. Dennett (1991) propõe analisar a informação verbal dos sujeitos da experiência com uma postura intencional, ou seja, tratar o falante como um agente racional, que possui crenças, desejos e outros estados mentais que possuem

intencionalidade. Deve-se ter a cautela experimental, para ter certeza de que o sujeito realmente fala sobre seus estados internos e não responde por outros motivos, como por achar ser a melhor resposta, ou uma resposta focada na audiência e não em si mesmo. Atribuir intencionalidade ao indivíduo é importante por permitir a inferência de motivos para certas respostas e ações, as quais ficariam incompreensíveis pela desconsideração da intencionalidade.

Por meio desses relatos verbais, Dennett (1991) acredita estar atingindo o mundo heterofenomenológico do sujeito. Um mundo a ser entendido como uma ficção útil, pois diz sobre o que o sujeito acredita. A interpretação de relatos fictícios é possível e chega a bons resultados. O filósofo afirma que explorar mundos fictícios não é inútil; pode-se aprender sobre uma história, sobre um texto escrito, sobre moral, sobre o autor ou até mesmo sobre o mundo real. Se excluirmos nossa opinião pessoal e formos neutros com a ficção, podemos extrair fatos incontestáveis sobre o mundo fictício. É possível conhecer verdades sobre o mundo fictício sem mesmo ter lido o texto no qual esse foi escrito; algo semelhante ocorre quando alguém é informado sobre o final de um filme que está prestes a assistir.

Pela análise da ficção percebe-se o quanto é possível saber sobre o que está representado, sem ter a mínima ideia de como tal coisa foi representada. O experimentador pode ter uma ideia do mundo heterofenomenológico do sujeito da experiência pela análise de seu relato enquanto uma ficção. Ainda, dois pesquisadores podem concordar com fatos sobre esse mundo heterofenomenológico e discordarem de como este é processado no cérebro. O mundo heterofenomenológico é para Dennett (1991) uma postulação teórica confirmável intersubjetivamente, tendo o mesmo status metafísico da Londres de Sherlock Holmes.

Nesta etapa do método é preciso acreditar na palavra do sujeito da experiência. Mas, acreditar apenas em relação ao que está representado para o sujeito, pois o método heterofenomenológico não confronta nem aceita como inteiramente verdadeira as asserções dos sujeitos, ele mantém uma neutralidade construtiva na esperança de construir uma descrição definitiva do mundo de acordo com os indivíduos. Desta forma, um heterofenomenólogo é comparável ao antropólogo (DENNETT, 1991).

Quando antropólogos estudam a crença de um Deus indígena, eles adotam uma postura agnóstica e descrevem a crença destes como um constructo lógico. Eles tentam tornar o constructo bem coerente, acabando com as intrigas de opiniões sobre as crenças. Mas o constructo pode ser incoerente no fim, afinal sabem que se trata apenas

do conjunto de crenças de um povo indígena. Essa tática da neutralidade é apenas temporária, necessária no caminho de elaborar e confirmar uma teoria empírica que em princípio poderia deixar de lado o relato dos sujeitos. Dennett (1991) acredita, portanto, em uma possível prevalência de métodos empíricos neurocientíficos sobre o próprio relato do sujeito sobre sua consciência.

Um exemplo pode ser dado a partir da pesquisa de Nishimoto *et al.*, 2011 mencionada anteriormente. Suponhamos que esses pesquisadores, após conseguirem dizer quais filmes certo indivíduo estava assistindo apenas pela reconstrução da imagem a partir da atividade do córtex visual (sem haver contato prévio dos pesquisadores com o filme), encontrem uma resposta incompatível do sujeito com a imagem reconstruída. Digamos que a imagem mostre um filme “x” e o sujeito relata assistir um filme “y”. Garantindo que não houve falhas no processo, quem terá a palavra final nesse caso serão os pesquisadores, baseados na imagem reconstruída. A tarefa deles será apenas de descobrir por que houve essa não correspondência entre a imagem e o relato. Nesse caso, o mais plausível seria inferir que o sujeito estava mentindo, mostrando também a importância de considerá-lo como um sujeito intencional.

Para exemplificar seu método, Dennett (1991) conta a história de Shakey, um robô desenvolvido na Califórnia no final dos anos 60, o qual, apesar de não fazer nada de especial, abriu caminho para reflexões filosóficas. Shakey era uma caixa em cima de rodas com uma câmera e uma ligação via rádio a seu processador. Ele ficava em uma sala com objetos geométricos, e ao digitar em seu computador o comando “empurre uma caixa” o robô realizava essa ação. Shakey funcionava da seguinte forma: a sua câmera recebia uma imagem que gerava um conjunto de 0s e 1s de acordo com os píxeis que representava em uma tela. Foram instalados padrões de 0s e 1s que correspondiam às linhas essenciais de figuras geométricas, de tal forma a permiti-lo “enxergar” e se dirigir em direção a elas de acordo com comandos humanos. Shakey utilizava de subprocessos sem inteligência, simples e mecânicos, de tal forma que nenhuma parte do computador entendia o que estava fazendo, e não havia mistério sobre como cada etapa era realizada. A organização esperta de mecanismos simples pode substituir a necessidade de um homúnculo cartesiano.

Dennett (1991) coloca uma questão. Suponhamos que programaram Shakey com repostas automáticas (digitadas pelo programador). O que ele deveria responder ao ser perguntado como consegue distinguir caixas de pirâmides?

1. “Eu analiso cada sequência de 10 mil dígitos de 0s e 1s procurando padrões”.
2. “Eu encontro limites de luz e escuridão e desenho linhas brancas em minha mente, procuro os vértices e assim identifico o objeto geométrico”.
3. “Não sei, algumas coisas parecem caixas, isso vem na minha cabeça como intuição”.

De acordo com Dennett (1991), todas as repostas são verdadeiras, são descrições do processamento de informações em diferentes níveis. Suponhamos que não fosse possível saber como Shakey realmente identifica os objetos e que ele pensaria de acordo com a resposta 2. Se ele fosse como uma pessoa, poderia dizer que sabe exatamente o que está fazendo, imaginando linhas e identificando os vértices para reconhecer os objetos, e assim não teria como sabermos melhor do que ele o que está sendo feito. Teríamos que aceitar sua descrição numa postura heterofenomenológica. Poderíamos até mesmo programá-lo para achar que identificava os objetos de forma completamente diferente de como realmente faz, e ainda assim ele acharia que sua forma era a correta. Por existir essa possibilidade em nós, devemos tratar a heterofenomenologia como ficção. Para Dennett (1991), da mesma forma que nós acreditamos termos *Qualia*, Shakey ou outros robôs mais complexos também poderiam acreditar. Não haveria uma diferença entre uma máquina e um humano no que diz respeito a experienciar fenômenos.

Em suma, o método é neutro em relação à realidade de itens fenomenológicos. A heterofenomenologia precisa supor que as pessoas possuem um mundo heterofenomenológico, uma postulação menos problemáticas do que algumas da física. O método descreve o mundo heterofenomenológico do sujeito que contem objetos intencionais. A melhor resposta para a natureza desse mundo é “nenhuma”, assim como um mundo fictício. Um heterofenomenólogo estuda mundos fictícios da mesma forma que antropólogos estudam deuses e mundos fictícios de culturas variadas. As histórias não são loucuras, mas ficções de um teórica dedicado; apesar disso, podem vir a ser substituídas com os avanços das explicações da ciência empírica. Os sujeitos não ganham a palavra final sobre o que de fato acontece neles próprios, mas ganham a palavra final sobre o que parece estar acontecendo pelo ponto de vista deles. Após descrever o mundo heterofenomenológico do sujeito, os teóricos podem focar a questão do que explica a existência de todos os detalhes desse mundo.

2.6 O Ponto de vista do Observador e o Teatro Cartesiano

Dennett (1991) acredita ser uma característica central de seres dotados de consciência ter um ponto de vista no mundo. Em situações cotidianas, as quais ocorrem em uma temporalidade natural para o ser humano, é comum localizar esse ponto de vista no próprio sujeito como um todo. Assim, um sujeito dentro de um trem observa o mesmo de forma estacionária, enquanto um indivíduo fora do trem o observa em movimento. Nesse caso, um ponto de vista está no sujeito dentro do trem, enquanto o outro está no sujeito fora do trem. Dennett (1991) questiona o que acontece ao tentarmos localizar um ponto mais específico no qual esse ponto de vista se situa.

Seria natural tentarmos localizar esse ponto de vista no cérebro. Descartes (1664) localizou este ponto de vista na glândula pineal, um órgão do cérebro que não possui um correspondente simétrico do lado esquerdo ou direito, é único e localizado no centro. Este foi um erro de Descartes, pois esta glândula em nada está relacionada com a consciência ou com a visão. Dennett (1991) argumenta que o cérebro já é o local no qual o ponto de vista está situado; não existe outro lugar mais específico para localizá-lo, e isto tem implicações contra intuitivas. O ponto de vista está espalhado pelo cérebro.

O conceito do Teatro Cartesiano na obra de Dennett (1991) está relacionado a diversos significados. No terceiro capítulo desta dissertação é feita uma análise desses significados. Para uma primeira introdução à teoria de consciência do filósofo, utilizaremos o significado principal e simplificado do conceito: O Teatro Cartesiano é um modelo tradicional de consciência que postula um lugar no cérebro/mente onde toda informação se junta e a consciência acontece, formando um teatro da consciência. O termo “Teatro Cartesiano” pode se referir tanto ao modelo de consciência quanto ao próprio teatro localizado no cérebro. Para Dennett (1991), esta visão implica que haja uma linha de chegada dividindo eventos inconscientes e conscientes. Assim, neste modelo tradicional de consciência, a informação é processada por áreas específicas do cérebro, e depois é enviada para um centro unificador de informação, no qual ocorre a consciência. Apesar de os psicólogos e neurocientistas terem abandonado o dualismo cartesiano, o filósofo acredita que estes não abandonaram a ideia de um Teatro Cartesiano no cérebro, um lugar onde a informação é reunida, uma linha de chegada para a ocorrência da consciência.

Para ilustrar as consequências contra intuitivas do abandono da linha de chegada no cérebro para a ocorrência da consciência, Dennett (1991) considera um experimento de pensamento em uma temporalidade não cotidiana, em frações de segundos. Suponha que alguém viu uma moça de cabelo curto e de óculos. Uma semana depois esta pessoa viu uma moça semelhante, mas de cabelo comprido e sem óculos. Após ver a moça de cabelo comprido perfeitamente, a memória desta pessoa sofre de uma revisão, na qual ela mistura propriedades das duas moças, fazendo o sujeito se lembrar de ter visto uma moça de cabelo comprido e de óculos. Ele estaria lembrando vividamente uma experiência que nunca ocorreu. Dennett chama essas revisões pós-experienciais da memória de *Revisões Orwellianas*. Mas, poderia ter acontecido desta pessoa ter alucinado no momento da experiência, assim ele teria realmente visto a moça com cabelo comprido e de óculos. Posteriormente ele apenas recordaria corretamente essa alucinação, afinal se trata de uma experiência de alucinação perceptiva, mas que ocorreu. A esta revisão pré-experiencial, Dennett (1991) dá o nome de *Revisão Stalinista*.

Em uma escala de tempo cotidiana, as revisões Orwellianas e Stalinistas funcionam bem. Podemos dizer sem dúvida que há uma diferença clara entre eventos que ocorreram antes da experiência e alterações da memória pós-experiência. Apesar disso, em uma temporalidade de frações de segundo, o ponto de vista está espalhado por todo o cérebro e não faz sentido (nessa temporalidade) marcar uma diferença entre as duas formas de revisão. Se houver uma linha de chegada para a consciência, a revisão teria que ser Orwelliana ou Stalinesca, pois o conteúdo que chegou primeiro à consciência seria “moça de cabelo comprido sem óculos” ou “moça de cabelo comprido com óculos”. Dennett (1991) afirma que a ordem temporal de discriminações não pode ser o que fixa a ordem subjetiva da experiência.

Como não há uma linha de chegada para um local onde a consciência ocorre, em uma fração de tempo tão curta, não há diferença entre eventos pré-experienciais e revisões de memórias, pois não há local algum onde a consciência em si está localizada. Essa é uma das consequências contra intuitivas que a ausência de uma linha de chegada no cérebro gera, que de acordo com Dennett (1991), não havia sido identificada por materialistas da contemporaneidade. Para abarcar essas consequências e ir contra a ideia de um Teatro Cartesiano o filósofo desenvolve seu Modelo de Esboços Múltiplos.

2.7 Esboços Múltiplos

De acordo com o Modelo de Esboços Múltiplos, todos os tipos de percepções, pensamentos e atividades mentais são realizados no cérebro por processos paralelos de interpretação e elaboração de *inputs* sensoriais. Toda informação que entra no sistema nervoso sofre contínuas “revisões editoriais”. O processo editorial ocorre durante frações de segundos: neste tempo, várias adições, incorporações, emendas e substituições podem ocorrer. Assim, não experienciamos a percepção crua, mas o produto de muitas interpretações e revisões. Estas conclusões vêm de estudos da percepção. Dennett (1991) cita um efeito especial que ocorre naturalmente, denominado de efeito McGurk. Em resumo, trata-se de um efeito no qual a audição é influenciada pelo movimento do lábio de quem fala. Em filmes dublados, alguns sons podem ser literalmente escutados diferentemente da real sequência de áudio por uma influência dos movimentos originais dos lábios dos atores. O filósofo interpreta essa evidência como uma vitória dos processamentos dos olhos sobre o dos ouvidos numa batalha editorial.

Essas fixações de conteúdo são distribuídas espacialmente e temporalmente no cérebro e podem ser localizados com precisão no espaço-tempo, porém o início delas não marca o início da consciência de seu conteúdo. Fica sempre em aberto se um conteúdo se tornará ou não consciente. É confuso perguntar quando ficam conscientes. Essas discriminações de conteúdo distribuído produzem, com o tempo, algo como um fluxo narrativo ou sequência, sujeito a contínuas edições por diversos processos distribuídos pelo cérebro, continuando indefinitivamente para o futuro. Esse fluxo de conteúdo é como uma narrativa por causa de sua multiplicidade. A qualquer momento no tempo, existem esboços múltiplos de fragmentos de vários estágios da edição em diversos lugares no cérebro (DENNETT, 1991)

Uma das principais novidades do Modelo dos Esboços Múltiplos é que a discriminação só precisa ser realizada uma vez, ou seja, não precisa ser enviada para outro discriminador mestre. Ainda, o Modelo de Esboços Múltiplos evita o erro tentador de supor ser necessária uma narrativa final (“publicada”) correspondente ao fluxo da consciência do sujeito (DENNETT, 1991).

Dennett (1991) cria uma analogia. Com a chegada da internet, de processadores de palavra e correio eletrônico, pode acontecer de diversos esboços de um mesmo artigo estarem em circulação. O autor pode estar fazendo diversas revisões em resposta a comentários de revisores feitos por correio eletrônico. Fixar um momento de publicação, e considerar um dos esboços como o artigo principal se torna arbitrário, e

nenhuma versão pode ser considerada a final, já que nenhuma contém todas as partes. De forma semelhante, procurar um momento no processamento do cérebro como o momento da consciência seria uma escolha arbitrária. Pode-se sempre “traçar uma linha” no fluxo de processamento cerebral, mas não existem diferenças funcionais para motivar uma declaração de um local onde todos os estágios são ajustes pré-conscientes e os processos subsequentes são contaminações da memória pós-experiência.

Voltando às revisões Orwellianas e Stalinistas no experimento de pensamento, podemos supor que ambas as visões possuem a mesma teoria do que acontece no cérebro; concordam sobre quando o conteúdo no cérebro entra em caminhos causais, apenas discordam se um dado momento ou local deve ser classificado como pré-experiência ou pós-experiência. Assim sendo, só haveria uma diferença verbal entre as duas teorias; ambas dizem a mesma coisa, com a diferença apenas em onde fazem a divisão - nas palavras de Dennett - “mitológica”. Esta divisão seria um momento no tempo e um lugar no espaço cujo sujeito não pode ajudar os teóricos a localizar, e que é indiferente para o resto da teoria. Trata-se de uma “diferença que não faz diferença”. O que importa para Dennett (1991), é que, em ambos os casos, no fim do evento, o sujeito teve a crença de ter visto uma moça de cabelo comprido e óculos, as duas situações são a mesma. Em frações de segundo, durante o acontecimento do evento, neurônios carregam informações contraditórias sobre os óculos ou o cabelo da moça, e nenhuma parte específica tem a informação final; logo, não há como dizer se um erro, nessa temporalidade, seria uma revisão de memória ou uma revisão pré-experiencial, os dois casos são exatamente o mesmo.

Dennett (1991, p.132) afirma que o Modelo de Esboços Múltiplos pode ser considerado um operacionalismo de primeira pessoa, pois ele nega a possibilidade em princípio da consciência de um estímulo na falta da crença do sujeito nessa consciência. Isto porque não há nenhum lugar aonde a consciência foi realmente mostrada. Ainda, utilizando o método heterofenomenológico, lembramos que o relato da consciência do sujeito é uma ficção útil, que só faz sentido enquanto um conto de como o sujeito acredita ter entendido a situação (DENNETT,1991).

Se não é a discriminação em um dado local do cérebro que marca o início da consciência de um conteúdo, como Dennett (1991) pretende explicar esta relação temporal? O filósofo aponta uma diferença entre o tempo daquilo que representa (veículo) e o tempo representado (conteúdo). Como uma carta de aniversário que chega atrasada, o conteúdo representa a felicidade de certo dia, mas o tempo do veículo é

outro. Neste exemplo, isso pode ser corrigido pelo indivíduo que olha a data da carta, mas o cérebro não tem informação sobre o momento de saída do estímulo, apenas de chegada. De toda forma, a diferença do tempo do veículo e do conteúdo é importante para a análise.

Todo evento no cérebro tem começo e fim, e aquilo que representa também tem começo e fim. Mas não há motivos para supor que o começo daquilo que representa, corresponde ao começo do que é representado. É possível haver uma foto enorme de objetos microscópicos; esse princípio se aplica também para o tempo. De acordo com Dennett (1991), o que importa para o cérebro é que ele controle eventos com a informação de que A aconteceu antes de B, não importa se a informação que A aconteceu seja reconhecida antes ou depois da informação de que B aconteceu.

Dennett (1991) ilustra a diferença do tempo do que é representado e do tempo daquilo que representa com um exemplo do exército britânico na guerra de 1812 e a batalha de Nova Orleans. Este exemplo ilustra, no nível macro, fenômenos que ocorrem no nível micro no cérebro. Antes da internet, telefone e rádio existirem, comunicar com um exército espalhado pelo mundo inteiro estando em Londres não era tarefa fácil. Durante a batalha de Nova Orleans, a guerra já havia acabado há 15 dias com o acordo de paz, mas a notícia só chegou após a batalha. O exército estava disperso pelo mundo, e cada grupo recebeu a notícia em uma época diferente. Se compararmos o exército espalhado pelo mundo com os neurônios espalhados pelo cérebro chegamos a uma ideia interessante, pois se no nível macro faz sentido dizer que a sensação de tempo é coerente, no nível micro esta não é mais, pois não há momento nenhum nos eventos do cérebro em que caiba a pergunta “do que está consciente agora?”; seria como perguntar “quando o Império Britânico foi informado sobre a paz na guerra de 1812?”; não houve um momento em que todos os membros do império soubessem, e ainda, alguns agiram como se houvesse guerra quando a decisão de paz já havia sido tomada. Da mesma forma, como a consciência está distribuída no cérebro, em questões de milissegundos não há um momento preciso, exato, em que uma pessoa está consciente de um evento.

Dennett (1991) pode propor que a consciência não é um fluxo coerente de eventos, e que na realidade está espalhada pelo cérebro, mas o filósofo ainda precisa explicar por que a consciência *parece* ser um fluxo coerente. Para dar conta dessa necessidade, a Máquina Joyceana foi postulada.

2.8 Máquina Joyceana

A fala é uma troca de informação que nos permite conhecer um pouco sobre os eventos conscientes de outro sujeito; ela expressa julgamentos sobre vários conteúdos da consciência deste. Para Dennett (1991), na realidade o que o sujeito faz na introspecção é a mesma coisa. O sujeito faz um julgamento sobre os vários conteúdos da consciência, os quais ele próprio pode interpretar, agir sobre e lembrar. Tanto na fala quanto na introspecção, há uma fixação da interpretação do que o sujeito experienciou que fornece pontos na sequência subjetiva. Dessa forma, a consciência também está relacionada ao falar consigo mesmo.

Como apresentado no capítulo anterior, para Dennett (1991) o hábito de falar consigo pode ter surgido como uma necessidade durante a evolução humana; pode não ser a melhor forma de unir as estruturas funcionais do cérebro, mas seria um caminho facilmente atingível que faria o bastante para ajudar na sobrevivência. Essa fala consigo serviria como uma retroalimentação entre as áreas do cérebro, realizando releituras seriais coerentes e contínuas dos eventos espalhados e incoerentes. Essas releituras geram certa organização por promoverem uma narrativa coerente, mesmo que não completamente realista, do que aconteceu. Por serem tão eficazes, essas descobertas se tornaram parte da cultura, a qual foi disponibilizada – através de memes – para todos.

As mudanças meméticas causadas no cérebro são tão significantes que fazem cada um de nós sermos individuais, e é o que faz um cérebro chinês ser diferente de um cérebro inglês. Mas essas mudanças são praticamente invisíveis para a tecnologia neurocientífica, e assim precisam ser estudadas em outro nível. Dennett (1991) aposta na compreensão deste nível em termos de *softwares*. Assim, o filósofo propõe que da dinâmica paralela dos neurônios surge um padrão informacional instanciando uma máquina virtual processando em serial. Esta máquina virtual implementada no cérebro recebeu o nome de Máquina Joyceana. De acordo com Teixeira (2008), esse nome foi derivado da obra de James Joyce, o autor de *Ulisses*, uma história baseada no fluxo do pensamento de um personagem.

A hipótese é que a consciência humana é um complexo de efeitos meméticos no cérebro que podem ser compreendidos enquanto operações de uma máquina virtual do estilo da máquina desenvolvida por Von Neumann implementada na arquitetura paralela do cérebro, o qual não foi desenvolvido para tais atividades. Os poderes desta máquina virtual aumentam o poder do *hardware* orgânico no qual ela processa (DENNETT, 1991).

Dennett (1991) acrescenta que sabemos da semelhança de nossas mentes com máquinas de Von Neumann, pois afinal elas serviram de inspiração para sua criação. Programadores têm dificuldades de programar em paralelo, enquanto que programar de forma serial é intuitivo, pois é assim que realizamos nossas próprias tarefas. Contudo, ao programar um computador em paralelo não há como ter intuições sobre como realizar uma tarefa, pois nossas consciências funcionam de forma serial.

Existe uma grande diferença entre o processamento em paralelo e o processamento serial. Este é muitas vezes um problema colocado contra a inteligência artificial: como imitar um cérebro se este funciona paralelamente e os computadores funcionam de forma serial? Uma máquina virtual pode copiar exatamente as funções de uma máquina que funciona em paralelo, apenas será absurdamente mais devagar. Ainda, não há como um usuário saber qual *hardware* uma máquina está usando para simular uma máquina virtual; sendo simulada, ela se torna extremamente semelhante (DENNETT, 1991). Esta simulação é feita pela acelerada execução do comportamento de cada nó dentro de uma rede neural de forma serial. No nível do programa, há uma simulação de atividade em paralelo, mas, na realidade, no nível do *hardware* há um processamento serial realizando cada detalhe passo a passo.

Uma máquina virtual é, portanto, um conjunto temporário de regularidades estruturadas impostas ao *hardware* por meio de um programa: uma receita estruturada de diversas instruções que possibilitam ao *hardware* ter conjuntos de hábitos e disposições para reagir (DENNETT, 1991, p.216).

Para Dennett (1991), a implementação da consciência humana é um processo contrário ao tradicional na computação, ou seja, se trata de um *hardware* processando em paralelo e simulando máquinas virtuais que processam em serial. Um programa no cérebro é resultado de diversas plasticidades ajustáveis que incorporam diferentes micro hábitos.

Para rodar programas, no caso de um sistema com processamento serial, o conjunto de 0s e 1s são registrados separadamente em registros na memória e são acessados um conjunto por vez no registrador de instruções. No caso de uma máquina paralela, esta tarefa é realizada por bilhões de padrões de forças de ligação entre os neurônios, os quais juntos fornecem ao *hardware* novos conjuntos de micro hábitos, um novo conjunto de regularidades condicionais do comportamento (DENNETT, 1991).

Os computadores tradicionais são padronizados para serem compatíveis aos programas. Isso faz com que a “aprendizagem⁸” seja instantânea, pois o processamento é o mesmo para todos os computadores de uma mesma marca. Já para os cérebros, o processo de aprendizagem não tem como ocorrer assim, pois cada cérebro, apesar de muito semelhante aos demais, tem um nível de individualidade.

Existem problemas de controle próprio resultantes da ação de especialistas funcionando em paralelo. De acordo com Dennett (1991), uma das tarefas fundamentais da Máquina Joyceana é resolver disputas, fazer com que transições entre regimes ocorram de forma sutil e inibir a ação conjuntas de forças contraditórias, privilegiando as forças que funcionam juntas. Como também mostrado por Baars (1988), tarefas conhecidas pelo sujeito podem ser executadas de forma rotineira sem a convocação de forças extras e, portanto, inconscientemente, mas quando tarefas são difíceis, elas requerem concentração e ajuda de ensaios e manipulações de si, como falar em voz alta. Essas técnicas de controle de si promovem a possibilidade de uma busca por elementos de uma forma única para os seres humanos, desligando a busca automática e escolhendo elementos específicos para buscar.

De acordo com Dennett (1991), podemos trabalhar com eventos futuros graças à possibilidade do pensamento hipotético, a utilização de hábitos de recordar para si e o ensaio dos desfechos das condutas que adotamos. Criamos uma memória da rota pela qual chegamos onde estamos (memória episódica).

Na visão de Dennett (1991), os fatos ocorridos em esboços múltiplos são rapidamente perdidos. Graças à linguagem, podemos descrever para nós mesmos o que ocorre no nosso processamento na forma de um julgamento. A evolução não tinha em mãos uma memória RAM potente para trabalhar e, portanto utilizou de vários truques para aumentar o potencial da memória, principalmente o ensaio, que pode ser ajudado por rimas e frases fáceis de serem lembradas. Neste caso, a ajuda para a memória RAM viria pela utilização de recursos auditivos já presentes. Ainda, os movimentos corporais usados para o pensamento podem não ser apenas efeitos colaterais, mas estratégias de ensaio para expandir a memória RAM ou traços de estratégias antigas. Por exemplo, desviar o foco da visão para o chão ou para o céu pode ser uma forma de evitar contato visual com estímulos interessantes, liberando recursos computacionais para o pensamento em questão. O filósofo afirma que os hábitos de auto-estimulação

⁸ A capacidade do computador de exercer as funções desejadas, as quais não poderiam ser exercidas antes da instalação.

aprimoram os poderes da memória fazendo com que os humanos sejam os melhores neste aspecto. Ainda, diz que o efeito de globalização citado por Baars (1988) é possível pela auto-estimulação.

A importância dessa máquina é de realizar um milagre político, criar um líder virtual entre o grupo de processadores sem elevar um deles para ditadura a longo termo. Assim, o comando é distribuído primeiro a uma coligação e depois a outra, com a transição de uma para outra sendo realizada de forma não caótica, visando sequências coerentes e com propósito, não sendo assim uma confusa e desorganizada luta por poder (DENNETT, 1991). Essa organização parece não ocorrer em temporalidades de milissegundos, durante, por exemplo, uma discriminação de estímulo visual. O trabalho da Máquina Joyceana parece ocorrer mais em uma temporalidade cotidiana. Essas organizações são responsáveis por tarefas como planejamento de ações, resolução de problemas complexos, inferências lógicas com muitas premissas, organização de tarefas por etapas, dentre outras capacidades humanas realizadas em longo prazo.

Parece correto afirmar que esta máquina cria o mundo heterofenomenológico do sujeito, o qual é apenas uma ficção sobre o que realmente aconteceu no cérebro que foi realizado por esboços múltiplos. A ideia da máquina virtual explica por que a consciência *parece ser* um Teatro Cartesiano. Assim, em certo momento, na temporalidade micro não há como dizer se um fenômeno é consciente ou não, o que importa é se posteriormente ele foi lembrado pela máquina virtual (a qual não está localizada em nenhum lugar específico do cérebro).

2.9 Considerações Sobre a Teoria de Dennett

Dennett (1991) admite que sua teoria não seja completa, mas ela chega a ser de certa forma homuncular. Como uma máquina virtual serial pode obter, em dado momento, tanto conhecimento e poder sobre esboços de narrativas que estão espalhados por todo o cérebro não é explicado. Como ela é implementada também não é dito. Além disso, a causalidade virtual não é discutida por Dennett, tornando confusa a forma pela qual a Máquina Joyceana poderia fugir de um epifenomenalismo. Concluímos que para solucionar o problema de como a consciência se relaciona com o cérebro (problema que é a fonte de todas as críticas realizadas por ele), o filósofo acaba postulando uma máquina virtual sem explicar a relação desta com o funcionamento do cérebro. Logo, ele postula uma entidade não objetiva sem descrever muito sua relação com os fatos

observáveis. Considerando seu funcionalismo homuncular, a Máquina Joyceana teria que ser explicada em termos de agentes cada vez menores, com o objetivo de aos poucos passar do domínio do virtual para o biológico, e do biológico para o físico. Entretanto, como isso deve ser feito com sua máquina virtual não é nada claro. Dennett (1991, p.402) diz que o epifenomenalismo é um problema por definição, pois se algo não tem efeito algum no mundo, este algo não poderia ser percebido por um indivíduo em primeiro lugar. Apesar disto, o filósofo não trata em nenhum momento de sua causação virtual. Sem maior explicação de como essa poderia ocorrer, a teoria fica vaga, pois se a máquina virtual for epifenomenal, ele entra em uma contradição, e se ela tiver o poder causal ela fica sem explicação de como isso pode ocorrer.

Há também uma complicação no que diz respeito ao método da heterofenomenologia e as afirmações sobre o Modelo dos Esboços Múltiplos. Como o método é descrito no mesmo livro do modelo, entende-se que este deve ser um método coerente com as propostas do modelo. Quando está descrevendo seu modelo, Dennett (1991, p.132) afirma que o Modelo de Esboços Múltiplos pode ser considerado um operacionalismo de primeira pessoa, pois ele nega a possibilidade em princípio da consciência de um estímulo na falta da crença do sujeito nessa consciência. Nos exemplos mencionados por Dennett, o sujeito acreditava em certa coisa, e se o cientista tivesse evidências e razões, poderia duvidar do relato do sujeito. Neste caso o método e o modelo funcionam em conjunto. Entretanto, se considerarmos o contrário, o caso no qual o sujeito não acredita que certo fenômeno foi consciente, mas os neurocientistas têm razões para acreditar que estava, o método e a afirmação sobre o modelo entram em conflito. O método aprovaria esta atitude do neurocientista dizendo que ele não precisa do sujeito para determinar a consciência, enquanto o modelo diria ser impossível. Isto ficou claro no exemplo citado baseado nas pesquisas de Nishimoto *et al.* (2011). Pelo método, os neurocientistas teriam razão para dizer que o filme assistido foi “x” mesmo que o sujeito relatasse “y”. Pelo modelo, não seria correto afirmar que o sujeito ficou consciente de “x”, pois ele não acredita ter assistido a este filme.

O Modelo de Esboços Múltiplos é uma alternativa a ser considerada ao Teatro Cartesiano. No próximo capítulo este modelo será comparado à Teoria do Espaço de Trabalho Global de Bernard Baars. Até agora foram discutidos principalmente experimentos de pensamentos de Dennett. Para sermos justos, ainda será discutida, no próximo capítulo, a interpretação de Dennett do fenômeno de Metacontraste. Também

será discutido se a Teoria do Espaço de Trabalho Global defende um modelo de consciência no estilo Teatro Cartesiano.

CAPÍTULO 3 – O TEATRO CARTESIANO

*“If you wonder if I am attacking a straw man,
it will be reassuring to discover some experts
tying themselves in knots because
they are genuine Cartesian materialists
in spite of themselves”*

—Daniel Dennett (1991)

3.1 Um contraste entre a Teoria do Espaço de Trabalho Global e o Modelo de Esboços Múltiplos

O modelo de consciência de Dennett (1991), como o próprio enfatizou em seu livro, é apenas um esboço, um caminho para se construir uma teoria da consciência. Dado esse problema, teria como o filósofo incorporar outras teorias ao seu modelo para fortalecê-lo? Dennett (2001) propôs algo semelhante. O filósofo propõe uma compatibilidade ou pelo menos certa semelhança de seu modelo à Teoria do Espaço de Trabalho Global. Mas mesmo em seu livro, Dennett (1991, p.257, tradução e itálico nosso) já afirmava:

Em seu livro *A Cognitive Theory of Consciousness* (1988), o psicólogo Bernard Baars sumariza o que ele considera um “início de um consenso” de que a consciência é realizada por uma sociedade distribuída de especialistas equipados com uma memória de trabalho, chamada de espaço de trabalho global, cujo conteúdo pode ser difundido para o sistema como um todo. Como ele nota, uma variedade de teóricos, apesar das enormes diferenças em perspectivas, treinamento, e aspiração, estão gravitando em direção a esta visão de como a consciência deve estar no cérebro. É uma versão desse consenso emergente que eu tenho introduzido cautelosamente, ignorando algumas características e enfatizando outras – características as quais penso ser negligenciadas ou subestimadas, e outras as quais penso ser particularmente cruciais para superar os mistérios conceituais que ainda restam⁹.

Em alguns aspectos, estas teorias são compatíveis. Ambas enfatizam um paralelismo do funcionamento inconsciente, fazendo com que poderes localizados sejam espalhados por redes neurais no cérebro. Os dois modelos delegam um poder organizacional para a consciência que funciona em serial e de forma limitada ajuda na orientação de especialistas para resolução de problemas. Apesar disso, parece haver afirmações centrais incompatíveis entre essas teorias.

Schneider (2007) aponta três pontos principais de incompatibilidade grave entre as teorias. O primeiro diz respeito ao conteúdo da consciência. Para Dennett (1991), o conteúdo da consciência depende de uma busca, realizada por sujeitos dotados de linguagem, dentre os eventos paralelos sendo realizados pelos esboços múltiplos. Para

⁹ In his book *A Cognitive Theory of Consciousness* (1988), the psychologist Bernard Baars summarizes what he sees as a "gathering consensus" that consciousness is accomplished by a distributed society of specialists that is equipped with a working memory, called a global workspace, whose contents can be broadcast to the system as a whole" (p. 42). As he notes, a variety of theorists, in spite of enormous differences in perspective, training, and aspiration, are gravitating toward this shared vision of how consciousness must sit in the brain. It is a version of that emerging consensus that I have been gingerly introducing, ignoring some features and emphasizing others — features that I think are either overlooked or underestimated, and that I think are particularly crucial for breaking through the conceptual mysteries that still remain.

Baars (1988), nem todos os conteúdos da consciência são buscados de forma voluntária, apesar de podermos realizar desta forma também. Esta distinção talvez seja apenas um problema resultante da pouca precisão de Dennett (1991) em diferenciar entre tipos de consciência em esboços múltiplos e consciência com a implementação da Máquina Joyceana. Esta distinção poderia ser resultado apenas da terminologia de Dennett. Se considerarmos que os processos em esboços múltiplos do cérebro são processos conscientes, a necessidade de buscas por conteúdos termina. Adicionando uma máquina virtual para a consciência, buscas são necessárias. Entretanto, parece ser mais do que meramente terminológico, quando no capítulo “Nossa consciência, a mente deles” (*Our consciousness, their minds*) Dennett (1996) claramente considera o acréscimo da máquina virtual como a única forma legítima de consciência.

O segundo ponto de incompatibilidade grave apontado por Schneider (2007) se refere à ligação que o espaço de trabalho global tem com um sistema central, ou seja, um sistema não modular, no qual conteúdos de diferentes modalidades são reunidos e o planejamento ocorre. Esta ideia está ligada à noção de um sistema central no estilo de Fodor (1983), a qual Dennett (1991) se opõe. Concordamos que este é um ponto de tensão. Entretanto, talvez a tensão diminua se concluirmos que o planejamento na realidade não é realizado pelo espaço de trabalho global e sim pelo cérebro todo, sendo apenas divulgado (“*broadcast*”) pelo espaço de trabalho global (posição que acreditamos ser mais coerente com as palavras de Baars [1988]). Mas mesmo assim, ainda há um sistema central semelhante àquele proposto por Fodor, por ser um sistema não modular capaz de se relacionar com todos os demais sistemas.

De toda forma, acreditamos ser o terceiro o ponto de maior incompatibilidade apontado por Schneider (2007). Este se refere à noção da linha de chegada para a consciência (como explicado no capítulo anterior), sendo o aspecto de maior crítica de Dennett (1991) aos modelos tradicionais de consciência. Para a teoria de Baars (1988), parece haver uma linha de chegada clara para a consciência, a saber: a entrada de um conteúdo no espaço de trabalho global. A mensagem precisa ser globalizada para fazer parte da consciência. Dessa forma, em um primeiro momento parece ser fácil concluir que a Teoria do Espaço de Trabalho Global é uma teoria no estilo “Teatro Cartesiano”, sendo incompatível com a teoria de Dennett. Como esse aspecto é o mais problemático, ele será discutido em detalhe na continuidade do capítulo.

Schneider (2007) parece ter identificado bem os principais pontos de conflito entre as teorias. Além desses três, encontramos em nossa análise mais quatro pontos de

conflito. O primeiro diz respeito à evolução da consciência. Como Baars (1988) relaciona o papel da consciência a uma estrutura anatômica antiga em termos evolutivos (o Sistema de Ativação Reticular-Talâmico Estendido), o psicólogo entende a consciência (no sentido completo) como uma função comum a muitos dos animais, pelo menos a todos os mamíferos. Já Dennett (1991) diz que a consciência é muito recente para ter evoluído biologicamente e delega esse papel para a evolução cultural – a difusão e instalação de memes. Dennett abre uma lacuna entre o animais humanos e outros animais, a qual não ocorre na teoria de Baars (1988).

Dennett (1991) chega a igualar a função dos hábitos de auto-estimulação ao papel da difusão de informação do espaço de trabalho global. Mas parece ser incerta a questão dessa igualdade do papel funcional destas duas formas de globalização de informação. Quando Dennett (1991, p.197, tradução nossa) discute a evolução da consciência (apresentada no primeiro capítulo desta dissertação), ele afirma que esses hábitos de auto-estimulação permitem “a criação de novos caminhos de comunicação interna¹⁰”. Essa afirmação parece ser um pouco mais fraca do que a globalização de informação do psicólogo. Para Baars (1988, p.100, tradução nossa), “uma verdadeira variável global no sistema nervoso pode ser distribuída em princípio para todos os níveis de todos os processadores, talvez para o nível de células individuais¹¹”. Para Dennett (1991) a criação de novos caminhos se trata apenas de novas rotas para comunicação interna. Assim, apesar de ter delegado à Máquina Joyceana um poder de globalização, não fica claro como a hipótese da criação de novos caminhos internos pode se igualar a globalização. Esta incompatibilidade é pouco problemática.

Uma incompatibilidade mais forte, e relacionada àquelas mencionadas por Schneider (2007) diz respeito à integração de informação. A consciência para Baars (1988) possui um papel central de integrar informações de especialistas espalhados pelo cérebro. Já Dennett (1991) afirma que o ponto de vista do indivíduo (consciência) é espalhado por todo o córtex. Poderíamos pensar que a Máquina Joyceana exerce tal papel integrador, mas isso não parece ser coerente com o resto da teoria. Seria mais coerente afirmar que esta faz releituras utilizando de diversos momentos das múltiplas narrativas geradas no cérebro. Afinal, para Dennett, não há um local onde toda informação é reunida para formar a experiência consciente.

¹⁰ [...] New paths of internal communication [...]

¹¹ A truly global variable in the nervous system might be distributed in principle to all levels of all processors, perhaps down to the level of single cells.

Este ponto de incompatibilidade leva a um outro. Seria correto afirmar que os especialistas inconscientes cooperam e competem pelo mesmo motivo em ambas as teorias? Novamente, não parece ser o caso. No caso de Baars (1988), a competição e cooperação ocorrem para o acesso ao espaço de trabalho global para formar a experiência consciente e influenciar o trabalho dos demais especialistas do cérebro. A competição e cooperação no caso do Dennett (1991) não é para que o conteúdo em si atinja algum tipo de estado privilegiado, mas se trata de uma competição mais direta para influenciar o comportamento, o vencedor da competição já causa o comportamento, ao invés de apenas informar os demais especialistas sobre sua proposta de ação.

No total, identificamos sete pontos principais de incompatibilidade entre a Teoria do Espaço de Trabalho Global e o Modelo de Esboços Múltiplos. Dentre os mais problemáticos estão aqueles que relacionam a teoria de Baars (1988) à noção do Teatro Cartesiano, justamente o ponto de crítica do livro inteiro (*Consciousness Explained*) de Dennett. Mas antes de verificar se a Teoria do Espaço de Trabalho Global postula um Teatro Cartesiano, é necessário primeiramente esclarecer os significados relacionados a este conceito.

3.2 O Conceito do Teatro Cartesiano

Dennett (1991) foi acusado de estar atacando um espantalho, ou seja, estar atacando uma visão da consciência que ninguém adota atualmente (BLOCK, 1993, 1995; TYE, 1993; BAARS, 1997). Dennett (1991), no próprio livro (antes mesmo das críticas), já tenta explicitar por que não ataca um espantalho. Concordamos que Dennett de fato não ataque um espantalho; acreditamos que essa falsa crítica se resulta de uma falta de cautela do leitor para com o livro de Dennett. Entretanto, como os críticos são filósofos especializados, possivelmente a culpa principal seja do próprio Dennett, por não definir com clareza o que está chamando de Teatro Cartesiano. Talvez o filósofo acreditasse ser um conceito comum da filosofia. Mas mesmo se isso for verdade, não justifica a quantidade de significados associados a esse termo, combatidos como se estivessem todos em uma mesma corrente inseparável. De toda forma, muitos neurocientistas não entenderam ou desconsideraram as críticas de Dennett, pois a maioria ainda postula modelos que parecem se encaixar em sua descrição do Teatro Cartesiano.

O filósofo se remete ao Teatro Cartesiano em diversos momentos do livro, trazendo alguma novidade sobre o que está realmente combatendo em cada momento. Dennett (1991, p.107, tradução nossa) sugere uma primeira definição:

O Teatro Cartesiano é uma figura metafórica de como a experiência consciente deve estar posicionada no cérebro. No primeiro momento parece ser uma extrapolação inocente do familiar fato incontestável de que para intervalos de tempo cotidianos e macroscópicos, podemos, de fato, ordenar eventos entre duas categorias “ainda não observados” e “já observados”¹².

Ainda, Dennett (1991, p.107, tradução nossa) afirma que irá combater o Materialismo Cartesiano descrito como:

[...] a visão que você chega quando descarta o dualismo cartesiano mas falha em descartar a imagem de um teatro central (mas imaterial) no qual “tudo se unifica.” A glândula pineal seria um candidato para tal Teatro Cartesiano, mas outros foram sugeridos – o cingulado anterior, a formação reticular, diversos locais do lobo frontal. O Materialismo Cartesiano é a visão de que há uma linha de chegada ou uma fronteira em algum lugar no cérebro, marcando um lugar no qual a ordem de chegada é igual à ordem de apresentação na experiência porque o que acontece lá é aquilo do qual se está consciente¹³.

Essas primeiras indicações dizem respeito a quatro pontos iniciais a serem combatidos. Primeiramente, a consciência é localizada em um lugar específico do cérebro. Segundo, há uma linha de chegada no cérebro, marcando um momento necessário para a experiência consciente acontecer. Terceiro, há um local no cérebro, relacionado à consciência que unifica a informação do cérebro. E quarto, existe a possibilidade de marcar um momento específico no tempo o qual defina que antes houve pré-experiência e depois deste houve pós-experiência (não se remete ao cérebro nesse caso). Concordamos que, de certa forma, todos esses sentidos estão relacionados, mas acreditamos que uma causa de confusão é que não necessariamente todos os pontos precisam ser adotados simultaneamente; talvez não haja uma relação de necessidade

¹² The Cartesian Theater is a metaphorical picture of how conscious experience must sit in the brain. It seems at first to be an innocent extrapolation of the familiar and undeniable fact that for everyday, macroscopic time intervals, we can indeed order events into the two categories "not yet observed" and "already observed."

¹³ [...] The view you arrive at when you discard Descartes's dualism but fail to discard the imagery of a central (but material) Theater where "it all comes together." The pineal gland would be one candidate for such a Cartesian Theater, but there are others that have been suggested — the anterior cingulate, the reticular formation, various places in the frontal lobes. Cartesian materialism is the view that there is a crucial finish line or boundary somewhere in the brain, marking a place where the order of arrival equals the order of "presentation" in experience because what happens there is what you are conscious of.

entre todos. Talvez haja como afirmar, por exemplo, que essa estrutura anatômica relacionada à consciência não integre informação, ou que integra informação, mas a experiência não está ali localizada. Ainda, se postularmos um centro funcional no modelo cognitivo que está espalhado pelo cérebro (como por exemplo, um centro funcional que pode estar em uma determinada rede neuronal ou em outra, em um circuito ou em outro, dependendo do momento) isto estaria ou não sendo combatido por Dennett? Se ele combatesse apenas o Materialismo Cartesiano, como descrito acima, isso não seria um problema. Mas o filósofo parece combater diversas outras ideias relacionadas a um Teatro Cartesiano. O fato de não adotar um centro funcional Fodoriano em sua teoria nos faz entender que talvez um centro funcional no modelo cognitivo, com localidade dinâmica no cérebro, não seja uma boa ideia. De toda forma, isto seria um Teatro Cartesiano?

Um teórico poderia adotar uma linha de chegada no cérebro para ocorrência da experiência consciente e não dizer que este local é um centro com todo poder executivo. E esse, parece ser outro ponto relacionado ao Teatro Cartesiano, como Dennett (1991, p.106, tradução nossa) explica: “[...] até o mais sofisticado materialista atualmente esquece que quando o *res cogitans* fantasmagórico é descartado, não há mais um papel para um portão centralizado, ou qualquer centro funcional no cérebro¹⁴.” Isto está relacionado também à – dita – novidade do modelo de esboços múltiplos, que um estímulo precisa apenas ser discriminado uma vez, não precisa ser – como nos modelos na forma de Teatro Cartesiano – enviado para outro centro para uma discriminação final.

Dennett não ataca apenas o Materialismo Cartesiano, um local no cérebro para a consciência. Em outro momento do livro, Dennett (1991, p.166-167, tradução nossa) faz uma declaração talvez negligenciada por quem o acusa de estar atacando um espantalho:

Parece haver um modelo alternativo para o início da consciência que foge do cérebro centralizado cartesiano e ainda permite a temporalidade absoluta. Não poderia a consciência ser uma questão não de chegada a certo ponto, mas sim uma questão de uma representação excedendo algum limite de ativação por todo o córtex ou a maior parte deste? Em tal modelo, um elemento de conteúdo se torna consciente em certo tempo *t*, não por entrar em um sistema funcionalmente definido e localizado anatomicamente, mas

¹⁴ [...] even the most sophisticated materialists today often forget that once Descartes's ghostly *res cogitans* is discarded, there is no longer a role for a centralized gateway.

por mudar de estado bem onde está: adquirindo alguma propriedade ou por ter a intensidade de uma de suas propriedades elevadas sobre algum limite¹⁵.

Dennett (1991) chama esta ideia de um modelo por mudança de estado, o qual a primeira vista parece se distinguir do Teatro Cartesiano, mas que na realidade é um disfarce. Nesse momento, o filósofo revela sua crítica mais inovadora, de que a ligação com o Teatro Cartesiano está principalmente em afirmar que certas discriminações em um tempo “T” no cérebro estão correlacionadas com certas experiências no mesmo tempo “T”, uma afirmação que qualquer teórico da identidade precisa adotar, cuja fuga só é possível pelo forte Funcionalismo do filósofo. Como Dennett (1991, p.166, tradução nossa) deixa claro:

A ideia de que a consciência é um modo de ação do cérebro ao invés de um subsistema do cérebro é bastante adotada (veja, e.g., Kinsbourne, 1980; Neumann, 1990; Crick e Koch, 1990). Ainda, tais mudanças de modo de ação podem ser presumidamente cronometradas por observadores externos, providenciando, em princípio, uma sequência única e determinada de conteúdos obtendo o modo especial. Mas isso ainda é o Teatro Cartesiano se for afirmado que o real (absoluto) momento de tais mudanças de modo de ação é definitivo da sequência subjetiva. A imagem é um pouco diferente, mas as implicações são as mesmas. Conferir a propriedade especial relacionada à consciência em um momento é apenas metade do problema; discriminar que a propriedade foi conferida naquele momento é outro, e apesar de observadores científicos com seus instrumentos poderem fazer isto com acurácia em microssegundos, como o cérebro pode realizar isso?¹⁶

Como Dennett (1991) citou, pesquisadores reconhecidos adotam uma versão disfarçada do Teatro Cartesiano (como definido pelo filósofo). Então, talvez o ponto principal de combate contra o Teatro Cartesiano seja na correspondência entre uma discriminação do cérebro em uma fração de segundo e a experiência nesse mesmo

¹⁵ There is, it seems, an alternative model for the onset of consciousness that avoids the of Descartes's centered brain while permitting absolute timing. Couldn't consciousness be a matter not of arrival at a point but rather a matter of a representation exceeding some threshold of activation over the whole cortex or large parts thereof? On this model, an element of content becomes conscious at some time t, not by entering some functionally defined and anatomically located system, but by changing state right where it is: by acquiring some property or by having the intensity of one of its properties boosted above some threshold.

¹⁶ The idea that consciousness is a mode of action of the brain rather than a subsystem of the brain has much to recommend it (see, e.g., Kinsbourne, 1980; Neumann, 1990; Crick and Koch, 1990). Moreover, such mode shifts can presumably be timed by outside observers, providing, in principle, a unique and determinate sequence of contents attaining the special mode. But this is still the Cartesian Theater if it is claimed that the real ("absolute") timing of such mode shifts is definitive of subjective sequence. The imagery is slightly different, but the implications are the same. Confering the special property that makes for consciousness at an instant is only half the problem; discriminating that the property has been conferred at that time is the other, and although scientific observers with their instruments may be able to do this with microsecond accuracy, how is the brain to do this?

momento. Nesse caso, dificilmente poderíamos afirmar que Dennett ataca um espantalho.

Mas se seguirmos rigorosamente a ideia de Dennett não chegaríamos à noção de que qualquer correlato neural da consciência já implica em um Teatro Cartesiano? Possivelmente. Em seu livro *Consciousness Explained*, o filósofo não deixou isso claro. Porém, Dennett (2009, p.234, tradução nossa) chega a afirmar que:

Outra área de ingenuidade filosófica na ciência cognitiva da consciência diz respeito à busca dos Correlatos Neurais da Consciência (CNC). Parece ser óbvio para diversos cientistas buscando resolver o mistério da consciência que é necessário haver um CNC, condições necessárias e suficientes, caracterizadas em termos de atividade neural localizável, para a experiência consciente. Como, de fato, não poderia haver um, se o materialismo é verdadeiro? [...] Não precisa haver nada misterioso ou anti-científico na observação de que uma busca dos CNC é provavelmente sem propósito. Se os processos que elevam conteúdos para a consciência são como processos que elevam divergências de linhagem em eventos de especiação, estaríamos procurando no lugar errado e no momento errado para os fundamentos da consciência focando em processos cerebrais que ocorreram no tempo da fixação de conteúdo – no tempo, em outras palavras, no qual o conteúdo relevante acabou de fazer uma diferença para a vida cognitiva do sujeito¹⁷.

Após essa afirmação fica fácil perceber que Dennett (1991, 2009) não ataca um espantalho, pois a maioria dos neurocientistas buscam estes correlatos neurais da consciência. O filósofo provavelmente acredita na possibilidade de não haver tais correlatos por considerar a consciência como uma questão de teorizar sobre os eventos do cérebro. Logo, temos a impressão de um Teatro Cartesiano pela atividade de uma máquina virtual (não localizada em um lugar específico) falando sobre discriminações do cérebro após as verdadeiras discriminações da consciência acontecerem por esboços múltiplos. Nesse ponto, o Funcionalismo de Dennett fica evidente e faz toda a diferença; a dimensão virtual parece de certa forma adquirir uma ontologia não redutível a estados cerebrais.

¹⁷ Another area of philosophical naivete' in the cognitive science of consciousness concerns the quest for the Neural Correlate of Consciousness (NCC). It has seemed obvious to quite a few scientists aspiring to solve the mystery of consciousness that there has to be an NCC, the necessary and sufficient conditions, characterized in terms of locatable neural activity, for conscious experiences. How indeed could there not be one, if materialism is true? [...] There need not be anything mysterious or anti-scientific about the observation that the quest for the NCC is probably a wild goose chase. If the processes that elevate contents to consciousness are like the processes that elevate lineage divergences into speciation events, we would be looking in the wrong place at the wrong time for the hallmarks of consciousness by homing in on the brain processes that occurred at the time of content fixation—at the time, in other words, when the relevant content just began to make a difference in the subject's cognitive life.

Todas as implicações do conceito do Teatro Cartesiano de Dennett discutidas até o momento foram relacionadas a uma localização da consciência no cérebro ou uma linha de chegada para a consciência. Entretanto, existem mais implicações concernentes à centralização do poder em diversas áreas da atividade humana, implicações da postulação de diversos homúnculos relacionados à plateia que assiste o Teatro Cartesiano: O observador central, o intencionador central, o planejador central, o compreendedor central, o falante central, e o próprio *self*. Todos esses homúnculos se referem à centralização do poder para uma função específica, como compreender uma frase ou planejar o que verdadeiramente queremos falar. Porém, nada impede que um teórico concorde que a fala, a visão e a compreensão sejam realizadas por sistemas de processamento em paralelo, nos quais não há um chefe, e ainda assim postule um centro executivo para o planejamento ou a formação do *self*. Nesse caso, tal teórico estaria adotando características do Teatro Cartesiano para algumas funções, enquanto para outras não.

Em suma, Dennett (1991) não combate apenas a ideia do Materialismo Cartesiano, mas diversas implicações que se seguem (de acordo com o mesmo) da metáfora do Teatro Cartesiano. A seguir organizamos uma lista com todos os requisitos para uma teoria não se enquadrar em um Teatro Cartesiano (pela visão de Dennett):

1. A consciência não pode ser localizada em nenhum lugar específico do cérebro.
2. Não pode haver uma linha de chegada no cérebro, marcando um momento necessário para experiência consciente acontecer.
3. Não pode haver um local no cérebro, relacionado à consciência que unifique a informação do cérebro ou toda a experiência consciente.
4. Não podemos marcar um momento específico no tempo para discernir entre pré-experiência e pós-experiência (não se remete ao cérebro nesse caso).
5. Não pode haver um módulo central executivo como proposto por Fodor.
6. Na percepção, um estímulo não precisa ser discriminado primeiramente para depois ser enviado para um centro executivo.
7. Não precisa haver um modo de ação no cérebro que corresponda à consciência.

8. Uma discriminação em tempo “T” no cérebro não pode corresponder a um evento no mesmo tempo “T” na consciência.
9. Não precisa haver correlatos neurais para a consciência.
10. Não podem existir homúnculos para lidar com funções específicas como linguagem, visão, planejamento e a construção do *self*.

No total, identificamos pelo menos 10 requisitos que, de acordo com Dennett, teóricos da consciência precisam preencher para fugir do Teatro Cartesiano. Entretanto, os mais importantes parecem ser o de número 2 e 8, a saber: “não pode haver uma linha de chegada no cérebro, marcando um momento necessário para experiência consciente acontecer”, e “uma discriminação em tempo ‘T’ no cérebro não pode corresponder a um evento no mesmo tempo ‘T’ na consciência”. Como argumentado por Dennett (1991, p.107, tradução nossa):

Se o “ponto” de vista do observador precisa ser espalhado por grande parte do cérebro do observador, o senso subjetivo de sequência e simultaneidade do observador precisa ser determinado por outra coisa que não a “ordem de chegada,” já que a ordem de chegada é incompletamente definida até que o destino seja especificado¹⁸.

Apesar de não englobar todos os outros, os requisitos de número 2 e 8 incorporam os pontos centrais da parte da crítica de Dennett que não está direcionada a um espantalho. Acreditamos que as confusões, a respeito da crítica de Dennett, acontece por nem todos entenderem que o ataque era maior do que parecia ser e não concordarem que todas as implicações do Teatro Cartesiano precisam ser descartadas.

3.3 A Teoria do Espaço de Trabalho Global é um Teatro Cartesiano?

Baars (1997), em seu artigo “*In The Theatre of Consciousness*”, volta a enfatizar a semelhança de sua proposta do espaço de trabalho global com a metáfora do teatro. O psicólogo compara o espaço de trabalho global ao palco, a atenção ao foco de luz no palco, a plateia aos processadores inconscientes e pessoas nos bastidores aos contextos

¹⁸ If the "point" of view of the observer must be smeared over a rather large volume in the observer's brain, the observer's own subjective sense of sequence and simultaneity must be determined by something other than "order of arrival," since order of arrival is incompletely defined until the relevant destination is specified.

influenciando os atores (conteúdo da experiência consciente). Como Baars (1997, p.301, tradução nossa) descreve:

um palco, um foco de luz da atenção brilhando no palco, atores representando o conteúdo da experiência consciente, uma plateia, e algumas pessoas invisíveis nos bastidores, exercendo grande influência sobre o que se torna visível no palco. O palco recebe informação sensorial e abstrata, mas apenas os eventos no foco de luz são completamente conscientes. [...] O foco de luz seleciona os atores mais significantes do palco, e quando esta acende, suas mensagens são distribuídas por uma plateia composta de rotinas e fontes de conhecimento inconsciente¹⁹.

Podemos concluir com essas palavras que Baars concorda que a Teoria do Espaço de Trabalho Global se encaixa em um modelo do Teatro Cartesiano da consciência? Não tão rapidamente. Baars (1997, p.229-300, tradução nossa) ainda no mesmo artigo afirma justamente o contrário:

Daniel Dennett e Marcel Kinsbourne criticaram uma forma de modelo de teatro, o Teatro Cartesiano, no qual a experiência consciente se junta em um mesmo local do cérebro, da mesma forma que René Descartes pensava que a consciência poderia estar localizada na pequena glândula pineal. Descartes estava procurando por um ponto sem dimensão no qual a alma individual pode se ligar ao cérebro. Dennett e Kinsbourne afirmam que o Teatro Cartesiano não pode funcionar, e acredito que estão certos. Não faz sentido algum. Não há um ponto único no cérebro onde tudo se junta. Mas ninguém na ciência atualmente sugere um Teatro Cartesiano. Certamente nenhum dos modelos cognitivos de teatro sendo propostos desde 1950 sofre desses defeitos. Telas de cinema também não se unificam em um único ponto sem dimensão. Teatros funcionam bem no mundo real e são metáforas boas para a experiência humana. Se tal metáfora se tornar enganosa em algum momento, precisaremos simplesmente abandoná-la e procurar algo melhor²⁰.

¹⁹ [...] a stage, an attentional spotlight shining on the stage, actors to represent the contents of conscious experience, an audience, and a few invisible people behind the scenes, who exercise great influence on whatever becomes visible on stage. The stage receives sensory and abstract information, but only events in the spotlight shining on the stage are completely conscious. [...] The spotlight selects the most significant actors on stage, and once lit up, their messages are distributed to an audience consisting of all the unconscious routines and knowledge sources – the vast array of unconscious tools we use to adapt to the world.

²⁰ Daniel Dennett and Marcel Kinsbourne have criticized one conceivable theatre model, the ‘Cartesian Theatre’ in which conscious experience ‘comes together in a single point in the brain’, much as René Descartes thought consciousness might be located in the tiny pineal gland. Descartes was looking for just *one* dimensionless point where the singular soul might connect with the brain. Dennett and Kinsbourne claim that the Cartesian Theatre cannot work, and I believe they are right. It makes no sense. There is no single point in the brain where ‘it all comes together’. But no one in science today suggests a Cartesian Theatre. Certainly none of the cognitive theatre models that have been proposed since the 1950s suffer from these defects. Nor do movie theatres converge on a single dimensionless point. Theatres work just fine in the real world, and provide helpful metaphors for exploring human experience. If such a metaphor becomes misleading at some point, we should simply walk away from it and look for something better.

Como exemplificado por esta passagem de Baars (1997), as críticas de Dennett raramente são completamente compreendidas ou levadas em consideração; o psicólogo afirma claramente que Dennett ataca um espantalho. Entretanto, como descrito na seção anterior, este não parece ser o caso. A fonte de confusão provavelmente está em haver múltiplos requisitos (listados na seção anterior) para uma teoria fugir do Teatro Cartesiano. A fala de Baars (1997) se refere apenas a dois requisitos, dentre os 10 listados anteriormente: (1) a consciência não pode ser localizada em nenhum lugar específico do cérebro; (3) não pode haver um local no cérebro, relacionado à consciência que unifique a informação do cérebro. Dessa forma, restam ainda oito requisitos que poderiam não estar sendo preenchidos, possivelmente classificando a Teoria do Espaço de Trabalho global como um Teatro Cartesiano. É importante notar que não foi dado qualquer critério para dizer quantos desses requisitos precisam ser preenchidos para a teoria se tornar um Teatro Cartesiano. Mas o mais relevante é poder discernir quais requisitos normalmente são preenchidos por teorias cognitivas da consciência (no caso deste trabalho será apenas a Teoria do Espaço de Trabalho Global) e se esses requisitos não preenchidos são realmente um problema para uma teoria cognitiva da consciência (discutido na próxima seção).

Como discutido no capítulo anterior, Baars (1988) pretende unificar, a partir de sua teoria, quatro hipóteses tradicionais sobre a consciência: A hipótese da ativação, a hipótese da novidade, a hipótese da ponta do iceberg e a hipótese do teatro. Com exceção da hipótese da novidade, todas as demais, de acordo com Dennett (1991), estão ligadas a noção de um Teatro Cartesiano.

No caso da hipótese da ponta do Iceberg, Dennett (1991, p.278, tradução nossa) explica:

Ray Jackendoff (1987) argumenta, no mesmo espírito, que os mais altos níveis de análise realizados pelo cérebro, referindo-se aos mais abstratos, não são acessíveis à experiência, apesar de tornarem a experiência possível por tornarem-nas significativas. Sua análise promove uma vacina útil para outra encarnação do Teatro Cartesiano como um “ápice” ou “a ponta do Iceberg”²¹.

Possivelmente Baars (1988) incorpore apenas alguns aspectos não problemáticos dessa metáfora da ponta do iceberg, como, por exemplo: os processos conscientes são

²¹ Ray Jackendoff (1987) argues in the same spirit that the highest levels of analysis performed by the brain, by which he means the most abstract, are not accessible in experience, even though they make experience possible, by making it meaningful. His analysis thus provides a useful antidote to yet another incarnation of the Cartesian Theater as the "summit" or "the tip of the iceberg."

apenas uma pequena porção acessível a nós, apenas a ponta do iceberg em um mar de processadores inconscientes. Nesse caso ele não estaria delegando uma superioridade a esta ponta do iceberg, mas apenas usando a metáfora para dizer do maior potencial da inconsciência sobre a consciência.

A hipótese da ativação também foi criticada por Dennett (1991, p.166). Esta crítica estava presente em sua problematização dos modelos por modo de ação na seção anterior. Se a consciência está relacionada a um ápice de ativação e o teórico reivindicar que este ápice no momento “T” corresponde a um evento no momento “T” na consciência, o modelo ainda será considerado como um Teatro Cartesiano. Outro problema para esta hipótese é de não explicar como o cérebro, como um todo, sabe que o limite em diversas áreas foi excedido. Dennett (1991, p.166, tradução nossa) argumenta: “as simultaneidades objetivas e a sequencia de eventos espalhados pelo córtex não possuem relevância funcional a não ser que possam ser detectados de forma precisa por mecanismos no cérebro²²”.

A argumentação de Dennett (1991) não apresenta indícios da possibilidade de haver uma metáfora do teatro, que não esteja ligada à imagem do Teatro Cartesiano como a que Baars (1997) tentou explicar em seu artigo. Logo, acreditamos ser seguro concluir que toda hipótese de teatro da consciência tenha alguma ligação aos problemas apontados por Dennett (1991). Se ainda restarem dúvidas de que Dennett não ataca um espantalho, a seguinte declaração de Baars (1997, p.307, tradução nossa) certamente resolve o problema: “De fato, todos os nossos modelos unificados do funcionamento mental atualmente são metáforas de teatro, é essencialmente tudo que temos²³”.

Existe a possibilidade de essas confusões estarem relacionadas ao pobre uso dessas metáforas por um dos teóricos, ou pela natureza incerta do conteúdo semântico de metáforas em geral. Assim, será preciso comparar a teoria de Baars exposta mais formalmente aos 10 requisitos listados anteriormente.

Baars (1988) concorda com o primeiro: “A consciência não pode ser localizada em nenhum lugar específico do cérebro”, mas ao mesmo tempo contradiz o requisito de número 9 “não precisa haver correlatos neurais para a consciência”. Baars (1988) acredita que informações relacionadas à consciência precisam passar pelo Sistema de Ativação Reticular-Talâmico Estendido para serem globalizadas para o resto do cérebro.

²² The objective simultaneities and sequences of events spread across the broad field of the cortex are of no functional relevance unless they can also be accurately detected by mechanisms in the brain.

²³ As it happens, all of our unified models of mental functioning today are theatre metaphors; it is essentially all we have.

Este parece ser um problema, na visão do filósofo, como Dennett (2001, p. 222, tradução nossa) enfatiza: “se, como Jack e Shallice ... apontam, a neuroanatomia funcional de Baars foi superada, isso mostra um progresso realizado nos anos subsequentes²⁴”. Talvez Baars concorde com isto atualmente e por isso tenha adotado a neuroanatomia funcional de Edelman *et al.* (2011).

O terceiro requisito diz que “não pode haver um local no cérebro, relacionado à consciência que unifique a informação do cérebro ou toda a experiência consciente.” Se estivermos nos referindo a um local único onde toda informação do cérebro é reunida, Baars (1997) afirmou que discorda. Mesmo que sua neuroanatomia sugerisse uma passagem pelo Sistema de Ativação Reticular-Talâmico Estendido, isso não precisa ser para toda informação do cérebro, mas ainda seria para toda informação consciente. Porém, se ainda restava algum problema no sentido de unificação de informação em um só local, esse foi superado pela adoção da neuroanatomia funcional de Edelman *et al.* (2011). A unificação da experiência consciente precisa acontecer no espaço de trabalho global e para Edelman *et al.* (2011) este corresponde a um núcleo dinâmico, que não tem localização fixa no cérebro. Como Baars (1988, p.32, tradução nossa) mostra, ao considerar teses aristotélicas:

Uma versão da metáfora do teatro teve bastante influencia no pensamento ocidental e oriental; a noção de um “sentido comum,” um domínio no qual todos os sentidos especiais encontram e trocam informação. [...] Existe uma semelhança notável entre as conclusões de Aristóteles e os argumentos [...] deste livro²⁵.

Os sentidos, e qualquer tipo de informação, precisam ser unificados e relacionados no espaço de trabalho global para juntos formarem uma experiência consciente única, coerente e integrada. Baars (1988, p.350, tradução nossa) explica:

Se fossemos forçados a escolher uma função principal da consciência, seria a habilidade de sistemas conscientes combinarem uma variedade de fontes de informação para definir uma experiência única e coerente²⁶.

²⁴ If, as Jack and Shallice (this volume) point out, Baars' functional neuroanatomy has been superceded, this shows some of the progress we've made in the intervening years.

²⁵ One version of the Theater Metaphor has had great influence in Western and Eastern thought; that is the notion of a "common sense," a domain in which all the special senses meet and share information. [...] There is a remarkable resemblance between Aristotle's conclusions and the arguments made in Chapters 2 and 3 of this book.

²⁶ If we were forced to choose one premier function of consciousness, it would be the ability of the consciousness system to combine a variety of knowledge sources in order to define a single, coherent experience.

Portanto, Baars assume a necessidade de integração da informação processada de modo distribuído no cérebro, mas não assume que tal integração seja realizada por uma estrutura fixa. Deste modo, a Teoria do Espaço de Trabalho Global não contradiz o terceiro requisito.

O segundo requisito “Não pode haver uma linha de chegada no cérebro, marcando um momento necessário para experiência consciente acontecer” parece ser contradito pela ideia de que a linha de chegada é o momento em que o Sistema de Ativação Reticular-Talâmico Estendido difunde a informação por todo o cérebro. Esse aspecto também pode ser superado pelo abandono da neurofisiologia de Baars (1988). Mas considerando o requisito de número 4 “não podemos marcar um momento específico no tempo o qual difere entre pós-experiência e pós-experiência” parece haver uma linha de chegada necessária para consciência na teoria de Baars, o momento da globalização. Eventos conscientes são claramente definidos. Como Schneider (2007, p.319, tradução nossa) identificou: “[...] existe um sentido definitivo no qual certos estados mentais estão *na* consciência, enquanto outros não estão: estados são conscientes quando estão no espaço de trabalho global²⁷”. Essa é uma questão central também por não preencher os requisitos 7 “não precisa haver um modo de ação no cérebro que corresponda à consciência” e 8 “uma discriminação em tempo “T” no cérebro não pode corresponder a um evento no mesmo tempo “T” na consciência”. Se Baars (1988) afirmar que a globalização em certo tempo “T” corresponde ao mesmo evento no tempo “T” subjetivo do indivíduo, sua teoria ainda é um Teatro Cartesiano. Isso acontece porque por esse modelo ainda há um sentido em que uma interpretação de experimentos da consciência em microssegundos possa ser Orwelliana ou Stalinista. No experimento de pensamento considerado no capítulo anterior, a revisão teria que ser Orwelliana ou Stalinesca, pois o conteúdo globalizado precisa ser “moça de cabelo comprido sem óculos” ou “moça de cabelo comprido com óculos”.

Por esses aspectos, poderíamos concluir que a Teoria do Espaço de Trabalho Global postula um Teatro Cartesiano. Entretanto, Baars (1988, p.131, tradução nossa) menciona um efeito denominado “efeito bola de neve” (*snowballing effect*), o qual pode fazer a teoria ser um pouco mais semelhante à de Dennett (1991):

²⁷ [...] there is a definite sense in which certain mental states are *in* consciousness, while others are not: states are conscious when they are in the global workspace

A neurofisiologia sugere que a difusão de informação pode não ser um evento instantâneo, mas um recrutamento de ativação em forma de “bola de neve”, apoiados por vários sistemas em retroalimentação. Por exemplo, o trabalho de Libet indica que pode haver uma demora de meio segundo para uma atividade cortical se tornar consciente. Isso é muito mais tempo do que uma mensagem tomaria para se difundir pelo cérebro, e sugere um fluxo circulando entre áreas corticais e subcorticais, crescendo em si mesmo até exceder um limite. Dessa forma, não podemos levar a metáfora de difusão tão literalmente: uma acumulação relativamente lenta resultaria no mesmo fim funcional²⁸.

Apesar dessa passagem aparentemente se assemelhar mais às ideias de Dennett, o efeito bola de neve resulta novamente em um ápice (excedendo certo limite) do processamento relacionado à experiência consciente. Poderíamos ainda dizer se certo fenômeno foi uma alucinação ou uma revisão de memória, nos baseando na mensagem que foi globalizada após a ativação exceder o limite.

Os requisitos 5 e 6 respectivamente dizem que na teoria “não pode haver um módulo central executivo como proposto por Fodor” e “na percepção, um estímulo não precisa ser discriminado primeiramente para depois ser enviado para um centro executivo”. Para Baars (1988) os estímulos são primeiramente discriminados pelos especialistas inconscientes e só após uma detecção de relevância destes são enviados para uma instituição central (*central facility*) para serem difundidos. Entretanto, o espaço de trabalho global não é um mestre executivo, o qual realiza todo o trabalho. As funções executivas são realizadas por vários especialistas inconscientes com o auxílio do espaço de trabalho global. A globalização serve justamente para que muitos dos sistemas inconscientes possam tentar resolver o problema. Assim, nesse ponto existe tanto uma semelhança quanto uma discordância com as ideias de Dennett (1991), pois existe um sistema central, ainda que este sistema seja dinâmico, como em Edelman *et al.* (2011). De toda forma, o problema da dupla discriminação do estímulo ainda permaneceria. Porém, uma forma de fugir deste problema seria afirmando que não há uma nova discriminação do estímulo no espaço de trabalho global, mas apenas uma difusão deste para sistemas com os quais ainda não teve contato.

O décimo e último requisito para uma teoria não ser um Teatro Cartesiano é que “não podem existir homúnculos para lidar com funções específicas como linguagem,

²⁸ The neurophysiology suggests that broadcasting may not be an instantaneous event, but a “snowballing” recruitment of global activation, supported by many systems, that may feed back on each other. For example, Libet’s work indicates that for cortical activity to become conscious may take as long as a half second (Libet, 1978; 1981). This is much longer than a single broadcast message would take, and suggests a circulating flow between cortical and sub-cortical areas, building upon itself until it reaches a threshold. Thus we must not take the broadcasting metaphor too literally: a relatively slow accumulation would accomplish much the same functional end.

visão, planejamento e a construção do *self*". Neste caso, não parece haver um problema para a teoria de Baars (1988), pois os homúnculos são dissolvidos em sistemas de redes neurais processando em paralelo, seguindo a solução do Funcionalismo Homuncular de Dennett.

Dennett (1991) diz propor uma versão do consenso explicado por Baars mencionando características que deveriam mudar. Entretanto, exatamente como uma união dessas teorias poderia ocorrer fica pouquíssimo claro. Talvez uma alternativa para uma junção seria a possibilidade dessas globalizações não corresponderem diretamente à consciência em certo momento, e a globalização de mensagens vagas (em frações de segundos) tornar impossível diferir entre revisões Stalinistas e Orwellianas. Ainda assim haveria correlatos neurais da consciência e um centro funcional relacionado à esta. Da forma que a Teoria do Espaço de Trabalho Global de Baars (1988) é apresentada, se assemelha um pouco mais ao modelo do Teatro Cartesiano do que ao Modelo de Esboços Múltiplos. Mas isso realmente é um problema? Na próxima seção será discutida a real força das críticas de Dennett.

3.4 Há Realmente um Problema em Postular um Teatro Cartesiano?

Como apresentado, o conceito do Teatro Cartesiano está relacionado a vários requisitos. O que parece ser essencial para Dennett, e também mais controverso, está ligado à afirmação de que uma discriminação em tempo "T" no cérebro não pode corresponder a um evento no mesmo tempo "T" na consciência, não podendo haver uma linha de chegada para o início da consciência. Isso se dá porque a adoção desta tese implica a possibilidade de discernir entre interpretações Stalinistas e Orwellianos em frações de segundos. Dennett (1991) acredita mostrar empiricamente que isto é impossível. Nessa seção essa questão será comentada.

A descrição do capítulo anterior foi de certa forma injusta com Dennett, pois diversas evidências empíricas foram apresentadas para sustentar as ideias de Baars, enquanto apenas experimentos de pensamento foram explicados para corroborar a teoria de Dennett. Entretanto, isso foi proposital, pelo fato das interpretações de Dennett sobre experimentos empíricos serem complicadas e controversas. Dennett (1991) trabalha com diversas evidências durante todo o livro, mas uma das principais trata o experimento de Metacontraste, exposto em Dennett (1991) e Dennett e Kinsbourne

(1992). Este fenômeno foi primeiro estudado por Fehrer e Raab (1964) e posteriormente por Breitmeyer (1984).

Da forma descrita por Dennett (1991), o experimento ocorre com a apresentação de um primeiro estímulo, um disco, com a duração de 30mseg, seguido imediatamente de um segundo estímulo, um anel cujo diâmetro de seu círculo interior é semelhante ao diâmetro do disco (Figura 1). O sujeito do experimento relatava apenas ter experienciado o segundo estímulo (o anel), porém se perguntado para adivinhar quantos estímulos foram apresentados, conseguia dizer “dois” com uma precisão maior do que se estivessem respondendo aleatoriamente.

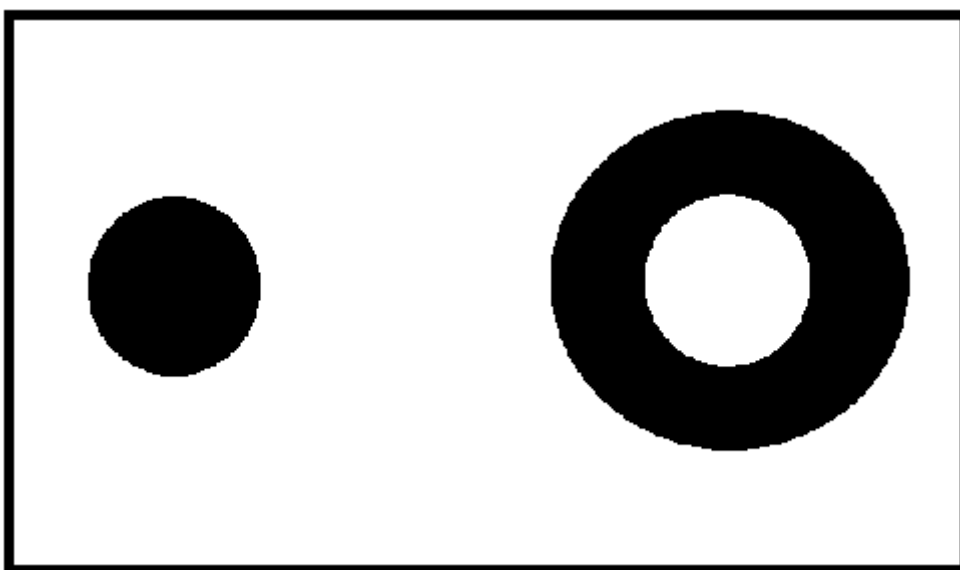


Figura 1 – O disco (esquerda) é apresentado primeiramente, seguido de um anel (direita), cujo diâmetro de seu círculo interior é semelhante ao diâmetro do disco (imagem desenvolvida pelos autores baseado no experimento como mencionado por Dennett [1991^a, p.141]).

Dennett (1991) afirma que a interpretação tradicional desse fenômeno é Stalinista: o segundo estímulo previne a experiência do primeiro de alguma forma. Pois, se o disco for apresentado individualmente ele é percebido normalmente. Como os sujeitos acertavam com frequência que foram apresentados dois estímulos, o teórico Stalinista pode concluir que o anel previne a chegada do disco à consciência, mas o disco ainda assim é discriminado inconscientemente, influenciando a resposta. Nas palavras de Dennett (1991, p.142, tradução nossa) “o primeiro estímulo nunca chega ao palco da consciência [...]”²⁹.

²⁹ The first stimulus never plays on the stage of consciousness [...].

Apesar de a interpretação tradicional ser Stalinista, uma alternativa Orwelliana também é inteiramente plausível. Por essa interpretação, sujeitos ficam conscientes do primeiro estímulo, mas a memória dessa experiência consciente é quase completamente substituída pelo segundo estímulo. Assim estes experienciam apenas o segundo estímulo, porém conseguem adivinhar que dois estímulos foram apresentados por causa de uma fraca memória que ainda resta do primeiro estímulo.

Com esse experimento (e claro, vários outros), Dennett (1991) pretende mostrar que não há como discernir empiricamente entre interpretações Stalinistas e Orwellianas (duas versões do Teatro Cartesiano) em fenômenos ocorrendo em frações de segundos. Essa impossibilidade é utilizada como principal argumento para a adoção de seu Modelo de Esboços Múltiplos.

A interpretação do fenômeno pelo Modelo de Esboços Múltiplos ocorre da seguinte forma. Em frações de segundos o cérebro pode tomar decisões que simplifiquem os fenômenos observados. O círculo do disco rapidamente se transforma no círculo interno do anel. Sem novas informações mostrando que havia o disco, o cérebro chega à conclusão de que havia apenas o anel. Algumas narrativas de quando o disco foi discriminado podem influenciar a resposta final. A novidade está na não necessidade de dizer se o disco chegou ou não à consciência, por não haver um Teatro Cartesiano, um local onde o estímulo precisa chegar, marcando o momento da consciência. Como Dennett (1991, p.142, tradução nossa) explica:

Deveríamos insistir que o disco foi experienciado porque se não houvesse a intervenção do anel o disco seria relatado? Isso seria cometer o erro de supor que poderíamos “congelar” o filme no Teatro Cartesiano para garantir que a imagem do disco realmente chegou ao Teatro antes que esta memória fosse enfraquecida por eventos subsequentes³⁰.

O argumento de Dennett é válido. Se não há como discernir entre os dois modelos do Teatro Cartesiano pelos experimentos de sua época, seu Modelo de Esboços Múltiplos se torna a opção mais plausível. Entretanto, Dennett (1991, pg.168) por acreditar ser um erro conceitual, ele conclui ser impossível haver um experimento que possa discernir entre as interpretações Stalinistas e Orwellianas, cometendo uma petição de princípio. O motivo é novamente o de não haver uma linha de chegada para a

³⁰ Should we insist that the disc was experienced because if the ring hadn't intervened the disc would have been reported? That would be to make the mistake of supposing we could "freeze-frame" the film in the Cartesian Theater and make sure that the disc frame really did make it into the Theater before the memory of it was obliterated by later events.

consciência. Em suma, o argumento pode ser colocado da seguinte forma: Nunca haverá como discernir entre interpretações Stalinistas e Orwellianas da consciência porque não há uma linha de chegada na consciência para marcar diferenças entre pré-experiência e pós-experiência. Logo, o Modelo dos Esboços Múltiplos é verdadeiro (não há uma linha de chegada para a consciência). A conclusão “não há uma linha de chegada para consciência” implícita em seu modelo é também uma premissa do argumento sobre a impossibilidade de discernir entre interpretações Stalinistas e Orwellianas mesmo com experimentos futuros. Dessa forma, Dennett (1991) pode apenas dizer que o Modelo de Esboços Múltiplos é mais plausível com os resultados experimentais de sua época. Isso implica a possibilidade de surgirem novas evidências a favor de um dos dois modelos do Teatro Cartesiano. E se de fato houver tais evidências, seu principal argumento contra o Teatro Cartesiano e a favor do Modelo de Esboços Múltiplos perde sua validade, a saber: como não há diferenças entre pré-experiência (Stalinista) e pós-experiência (Orwelliana) em frações de segundo, a consciência deve ocorrer por múltiplos esboços, não havendo como marcar uma linha de chegada para a consciência.

Os projetos de Todd (2006, 2009), O’Brien e Opie (1999), Breitmeyer *et al.* (2007) e Christie e Barresi (2002) pretendem mostrar, justamente, a partir de novos experimentos, evidências que possam diferenciar entre interpretações Stalinistas e Orwellianas, para tornar o argumento principal a favor do Modelo Esboços Múltiplos falso. Todd (2006, 2009) e Breitmeyer *et al.* (2007) concluem para o fenômeno do Metacontraste que a interpretação Stalinista é a única verdadeira. Christie e Barresi (2002) mostram como ambos os tipos de interpretação são verdadeiros. Dependendo da tarefa realizada pelo sujeito, um fenômeno pode ter uma interpretação Stalinista e outro Orwelliana. Já O’Brien e Opie (1999) argumentam que por uma visão conexionista não há problemas para a adoção de um Materialismo Cartesiano.

Como, nesta seção do trabalho, o fenômeno do Metacontraste está sendo focado, sendo um dos principais experimentos utilizados por Dennett (1991) e Dennett e Kinsbourne (1992) a favor do Modelo de Esboços Múltiplos, uma breve descrição das interpretações Stalinistas de Todd (2009) do fenômeno será apresentada.

No início do século XXI houve uma retomada nas pesquisas sobre o fenômeno de Metacontraste, por Breitmeyer e O Gmen (2000, 2006) e Breitmeyer *et al.* (2004). Todd (2006, 2009) baseado nessas novas pesquisas propôs uma crítica a Dennett e Kinsbourne (1992). Todd (2009) explica como a mudança de tempo entre a apresentação dos dois estímulos, e as cores utilizadas nos estímulos podem fazer “uma

diferença que faz a diferença” (em alusão à dita impossibilidade de discernir entre pré-experiência e pós-experiência, de acordo com Dennett, “uma diferença que não faz diferença”).

De acordo com Todd (2009), a variação de tempo entre a apresentação dos dois estímulos pode fazer com que os sujeitos variem seu relato de experiência do disco. Em um tempo otimizado entre a apresentação dos dois estímulos os sujeitos não relatam terem experienciado o disco (condição 1). Já em algumas variações do tempo entre a apresentação dos estímulos, os sujeitos conseguem relatar a experiência do disco (condição 2).

Nos experimentos (BREITMEYER & O GMEN, 2000, 2006; BREITMEYER *et al.* 2004) analisados por Todd (2009), foram utilizadas três cores nos estímulos: verde, azul e branco. O sujeito teria que responder o mais rápido possível identificando a cor do estímulo (cronometrando o tempo de reação). Na condição 1, quando ambos os estímulos eram verdes, ocorria um efeito de *priming* positivo, ou seja, o tempo de reação diminuía. Quando ambos os estímulos eram azuis, ocorria novamente um efeito de *priming* positivo, ou seja, o tempo de reação diminuía. Quando um estímulo era verde e o outro azul ocorria um efeito de *priming* negativo, ou seja, o tempo de reação aumentava. O disco branco teve o mesmo efeito do disco verde, ou seja, *priming* positivo para o anel verde e negativo para o azul. Na condição 2 (quando sujeitos relatavam ver o disco), os discos verde e azul tinham o mesmo efeito relatado para condição 1. Entretanto, o disco branco não teve nenhum efeito significativo nos tempos de reação para nenhum dos anéis.

Ainda de acordo com Todd (2009), Breitmeyer, Ro e Singhal (2004) analisaram a tela utilizada no experimento e descobriram que o disco branco era composto de 68% fósforo verde, mas apenas 12% fósforo azul e 20% do fósforo vermelho. Logo, o comprimento de onda do estímulo branco e do verde era similar, enquanto o do branco e azul era diferente. Assim, os efeitos de *priming* da condição 1 foram correlacionados com o comprimento de onda (discriminação realizada pela área v1 do córtex visual), enquanto na condição 2 os efeitos de *priming* estavam correlacionados com a percepção da cor (realizada pela área v4 do córtex visual). Na condição 1 o disco branco teve um efeito de *priming* relacionado ao comprimento de onda e na condição 2 não houve efeito de *priming* pelo branco ser uma cor neutra.

De acordo com Dennett (1991), e Dennett e Kinsbourne (1992), eventos que participam da experiência consciente precisam ter um papel funcional. Relembremos

que tanto pelos critérios de Baars (1988) quanto os de Dennett (1991) podemos inferir sobre a consciência quando o sujeito relata ter tido consciência do estímulo e o experimentador pode confirmar a acurácia do relato. Evidentemente, então, na condição 2, na qual houve o relato do disco branco, podemos concluir que o sujeito estava consciente deste disco. Como uma interpretação Orwelliana necessita que o disco seja primeiro conscientemente percebido e os efeitos subsequentes sejam por causa desta discriminação consciente, Todd (2009) argumenta coerentemente, que para a interpretação Orwelliana funcionar nesse experimento, as condições 1 e 2 teriam que dar o mesmo resultado funcional. Por essa interpretação, em ambos os casos, o estímulo precisa passar pela consciência. Mas não é isso que acontece: na condição 1 o disco branco tem o efeito de *priming* positivo com o disco verde e negativo com azul, e na condição 2 este mesmo não promove mudanças significativas no tempo de reação. As condições 1 e 2 não tiveram o mesmo efeito funcional, logo, não pode ter havido consciência do disco em ambas. O filósofo acaba por realizar uma análise contrastiva (no estilo de Baars) que pode eliminar a interpretação Orwelliana no caso do Metacontrast.

Pela interpretação Stalinista, na condição 1, o disco branco não chegou à consciência, seu comprimento de onda foi percebido inconscientemente e influenciou o tempo de reação. Na condição 2, o disco branco chegou à consciência e sua cor não resultou em um *priming* de outras cores. De acordo com Dennett (1991) e Dennett e Kinsbourne (1992), o Modelo de Esboços Múltiplos é mais favorável por não haver diferenças funcionais entre processamentos inconscientes e conscientes nestas pesquisas com temporalidade micro e por não haver diferenças entre as previsões das interpretações Stalinistas e Orwellianas para estes fenômenos. Todd (2009) mostrou que pode haver diferenças funcionais entre processamentos inconscientes e conscientes em frações de segundos, e que apenas a interpretação Stalinista pode prevê-las corretamente.

Concordamos com as conclusões de Todd (2009) de que uma prova empírica capaz de discernir entre interpretações Stalinistas e Orwellianas é capaz de enfraquecer consideravelmente a proposta do Modelo de Esboços Múltiplos. Dennett (1991) não utiliza apenas o Metacontrast como a única fonte de evidência, mas a maioria das outras evidências também diz respeito à incapacidade de distinção de uma interpretação Stalinista de uma Orwelliana.

3.5 Considerações sobre o capítulo

Foram diversas as críticas de Dennett (1991) ao Teatro Cartesiano, e pudemos identificar 10 requisitos para uma teoria não se tornar um Teatro Cartesiano (modelo problemático, de acordo com o filósofo). Algumas críticas de Dennett podem ser consideradas não muito originais, como a questão da descentralização do poder. Entretanto, identificamos outras críticas sérias direcionadas à maioria das teorias cognitivas da consciência, inclusive a Teoria do Espaço de Trabalho Global. De toda forma, uma análise rigorosa destas críticas bem direcionadas, complementada pelas análises de Todd (2009), revelou uma possível falsidade das mesmas. Na lógica clássica um exemplo contrário a uma afirmação universal é o bastante para falseá-la. Logo, a afirmação “para todo evento em milissegundos não há uma diferença entre interpretações Stalinistas e Orwellianas” seria falseada pela análise de Todd (2009).

Unifiquemos os 10 requisitos listados para uma teoria não se enquadrar nas críticas ao Teatro Cartesiano em dois, aqueles relacionados à linha de chegada no cérebro e aqueles relacionados à centralização do poder. Aqueles relacionados à centralização do poder estariam, no caso de Baars (e para a maioria de outros teóricos da consciência), atacando um espantalho. Já aqueles relacionados à linha de chegada seriam contra a teoria de Baars, mas não seriam problemas graves, como mostrado por Todd (2009). Talvez o restante das críticas de Dennett possa ser direcionado a teorias que ainda não descentralizaram todo o poder executivo, incluindo a exclusão de um módulo central Fodoriano e uma não necessidade da unificação da experiência consciente em lugares específicos do cérebro. Entretanto, estes apelos, apesar de relevantes, dificilmente são exclusivos do autor, sendo críticas já tradicionais na ciência cognitiva antes da publicação de *“Consciousness Explained”*. Além disso, a Teoria do Espaço de Trabalho global postula uma instituição central (*“central facility”*), como um sistema dinâmico que promove a difusão (*“broadcasting”*) da informação processada em paralelo, e não como um centro executivo fixo. Mesmo assim, algumas críticas de Dennett, como a não necessidade de correlatos neurais para consciência permanecem e precisam ser investigadas se procedem.

A teoria do espaço de trabalho global como um modo de ação do cérebro inteiro, como proposto em Edelman (2011), é bastante coerente. Entretanto, mesmo se a crítica de que uma discriminação em tempo “T” no cérebro não pode corresponder a um evento no mesmo tempo “T” na consciência for falsa, resta ainda o problema de não

explicar como o cérebro como um todo sabe que o limite do modo de ação em diversas áreas foi excedido. Acreditamos ser esta uma questão empírica que novos incrementos a teorias neurobiológicas da consciência possam superar, como proposto por Pereira Jr. e Furlan (2010). Estes autores entendem que a rede astrogliial faz a leitura da atividade neural, gerando ondas de íons de cálcio de longa escala apenas quando a atividade neural extrapola um limiar de sincronização oscilatório. Embora não seja conveniente entrar aqui em detalhes fisiológicos deste modelo, o citamos como exemplo da possibilidade de existência de um “teatro cerebral” da consciência, que se distingue em diversos aspectos do Teatro Cartesiano criticado por Dennett.

CONSIDERAÇÕES FINAIS

No primeiro capítulo sobre a evolução da consciência chegamos à conclusão de que existem certos pontos de semelhanças e alguns de conflitos entre a visão de Dennett e a visão que segue a Teoria do Espaço de Trabalho Global. Enquanto Dennett foca a cultura e a proliferação de memes para entender a origem da consciência, a Teoria do Espaço de Trabalho Global entende esta como uma adaptação biológica. Apesar de distinguir entre consciência primária e secundária, e dar valor na linguagem para a consciência secundária, a visão relacionada a teoria de Baars parece aceitar mais facilmente a possibilidade de outros animais terem aspectos da consciência secundária. Já Dennett por postular que a consciência é instalada justamente pela linguagem, nega a possibilidade de consciência propriamente dita em animais. Acreditamos que a linguagem tem pouca relação com a consciência, sendo esta segunda um produto emergente de fenômenos de interação entre redes neurais e astrogliais.

No segundo capítulo, discutimos pressupostos da Teoria do Espaço de Trabalho Global e do Modelo dos Esboços Múltiplos. Identificamos que para dizer qual conteúdo estará na consciência em certo momento, Baars depende da noção de hierarquia de metas, mas não há o estabelecimento de um critério de relevância para determinar a hierarquia de metas. A ideia de contexto que ele desenvolve também compartilha dos mesmos problemas de conceitos da ciência cognitiva tradicional, o “*frame problem*”. A consciência é também demasiadamente igualada à atenção nesta teoria. Sobre o Modelo dos Esboços Múltiplos, chegamos à conclusão de que para resolver o problema de como a consciência se localiza no cérebro, Dennett utiliza de uma máquina virtual, sendo que a relação desta com o cérebro é o pouquíssimo explorada, tornando difícil entender o avanço desta ideia em relação a outras teorias de como a consciência se relaciona com o cérebro. Uma possível incoerência entre o método proposto por Dennett e seu Modelo de Esboços Múltiplos também foi apontado como uma dificuldade da teoria.

No terceiro capítulo conseguimos identificar, pelas palavras de Dennett, 10 requisitos para um teórico fugir do Teatro Cartesiano. Alguns desses não foram muito originais, como aqueles que se referem à descentralização do poder. Entretanto, identificamos críticas sérias que poderiam atingir a Teoria do Espaço de Trabalho Global. De toda forma, uma análise rigorosa destas críticas parece falseá-las.

Concluimos o seguinte para cada um dos 10 requisitos:

O número 1 – “A consciência não pode ser localizada em nenhum lugar específico do cérebro” – está principalmente atacando um espantalho, pois a ideia de uma estrutura cerebral apenas para o local da consciência já foi superada.

O número 2 – “Não pode haver um local no cérebro, relacionado à consciência que unifique a informação do cérebro ou toda a experiência consciente.” – pode ser falseado por Todd (2009).

Se o número 3 – “Não pode haver um local no cérebro, relacionado à consciência que unifique a informação do cérebro ou toda a experiência consciente” – estiver se referindo a algum local do cérebro no qual toda informação é reunida, está atacando um espantalho. Pode estar dirigida a Teoria do Espaço de Trabalho Global se for uma crítica válida sobre a integração da experiência.

O número 4 – “Não podemos marcar um momento específico no tempo para discernir entre pré-experiência e pós-experiência” – é também falseado por Todd (2009).

O número 5 – “Não pode haver um módulo central executivo como proposto por Fodor” – é uma verdadeira incompatibilidade entre as teorias que não foi resolvida. Pois apesar do espaço de trabalho global não ter um papel executivo tradicional, ele ainda é um sistema central no modelo.

Sobre o número 6 – “Na percepção, um estímulo não precisa ser discriminado primeiramente para depois ser enviado para um centro executivo.” – não fica claro se a Teoria do Espaço de Trabalho Global afirma que há dupla discriminação, mas é um fato que os estímulos são primeiramente discriminados inconscientemente para posteriormente requisitar o acesso ao espaço de trabalho global.

O número 7 – “Não precisa haver um modo de ação no cérebro que corresponda à consciência” – em parte é falseado por Todd (2009), mas existem problemas do reconhecimento que o cérebro como um todo entrou nesse modo de ação.

O número 8 – “Uma discriminação em tempo ‘T’ no cérebro não pode corresponder a um evento no mesmo tempo ‘T’ na consciência” – é falseado por Todd (2009).

O número 9 – “Não precisa haver correlatos neurais para a consciência” – é uma afirmação muito pesada direcionada a maioria dos neurocientistas da cognição, a qual pode ser verdadeira, mas por enquanto é uma suposição.

Finalmente, o número 10 – “não podem existir homúnculos para lidar com funções específicas como linguagem, visão, planejamento e a construção do *self*” – na maior parte dos casos ataca um espantalho.

Apesar de Dennett acreditar que sua teoria pode ser considerada uma versão do consenso mencionado por Baars, a Teoria do Espaço de Trabalho Global encaixa em algumas das críticas de Dennett sobre o Teatro Cartesiano, tornando pouco claro como estas teorias poderiam ser compatíveis. Algumas dessas críticas são falseadas por Todd. Algumas precisam de evidências, por enquanto não são sérios problemas para Teoria do Espaço de Trabalho Global, apenas diferenças desta com o Modelo dos Esboços Múltiplos. Já para o Modelo dos Esboços Múltiplos em si, as análises de Todd são graves problemas.

Seriam as propostas do espaço de trabalho global como um modo de ação da atividade cortico-talâmica ou a proposta de leituras de redes astrogliais da atividade neural para formação de um espaço de trabalho global verdadeiros Teatros Cartesianos? Como Rockwell (1997, p.335) descreve “O teatro se torna uma arena política, onde palestras são feitas para mover a sociedade da mente para ação, ao invés de meramente entreter uma audiência passiva”. Talvez se tivéssemos que seguir as palavras de Dennett rigorosamente (de uma forma que provavelmente nem ele faz), estas ideias seriam Teatro Cartesianos (por violar alguns dos 10 requisitos). Entretanto, estas ideias já são tão diferentes das ideias originais de Descartes que não deveriam ser chamadas de Cartesianas.

Este trabalho identificou vários problemas que podem ser explorados em futuros trabalhos. Por questões práticas é difícil saber como a consciência pode ter surgido durante a evolução. O contraste das duas visões apresentadas neste trabalho levanta o problema da relação da linguagem com a consciência; também, a época na história da terra no qual os primeiros tipos de consciência podem ter surgido e quais animais possuem graus de consciência semelhantes à consciência humana ou até mesmo iguais.

Identificamos um problema sobre a noção de metas e a Teoria do Espaço de Trabalho Global, e no segundo capítulo sugerimos uma ligação desta teoria com as ideias de Maslow (1970) sobre metas. Entretanto, essas ideias já são antigas, seria necessário um estudo profundo, dedicado a encontrar, ou até mesmo desenvolver, uma teoria de motivação, personalidade e metas que possa fielmente embasar a seleção dos conteúdos conscientes na Teoria do Espaço de Trabalho Global. Isto se considerarmos que as causas da seleção são internas. Seria interessante explorar uma forma de utilizar

ideias da cognição incorporada e situada para solucionar certos problemas que Baars pode ter enfrentado por adotar apenas a ciência cognitiva tradicional.

Baars consegue construir uma relação bastante interessante entre consciência e a teoria matemática da comunicação. Entretanto, estes aspectos são pouco explorados e apesar de importância destes, não foram desenvolvidos pelo autor desde a publicação do livro em 1988. Ideias sobre como o significado pode entrar nessa relação entre informação e consciência precisam ser elaboradas e podem gerar novidades interessantes para a filosofia da mente.

Baars (1988) dá uma ênfase para a serialidade da consciência, mas admite a possibilidade desta não ser absoluta. De toda forma, acredita ser mais simples resumir as evidências dizendo que a consciência trabalha em serial enquanto a inconsciência em paralelo. Por sua metodologia de começar com fenômenos mais facilmente exploráveis, fenômenos na fronteira entre consciência e inconsciência ficaram pouco explorados. É necessário discutir como a teoria de Baars pode trabalhar a consciência de uma forma menos dicotômica com a inconsciência.

Baars trata a consciência como um aglomerado integrado das diversas modalidades perceptivas. Mas esta poderia ser apenas uma integração aparente pelo fato das diversas percepções acontecerem ao mesmo tempo. Será que a consciência é mesmo responsável pela integração dos diversos sentidos? Isto precisa ser trabalhado e questionado.

Dennett propõe que a consciência é um fenômeno virtual que pode não ter correlatos neurais fixos com a consciência. Se isso for verdade, qual seria a ontologia da virtualidade e como poderia ocorrer a causação do virtual com o físico? Dennett não entra em detalhes e seria preciso um trabalho dedicado em filosofia da mente para desenvolver este tema. Outra questão se refere à validade desta proposta de Dennett. Será que realmente não precisa haver correlatos neurais da consciência, tornando a busca realizada pelos neurocientistas desnecessárias? Seria preciso desenvolver esta ideia melhor.

Finalmente, tratamos as críticas de Todd (2009) como se fossem coerentes e realmente atingissem o núcleo do Modelo dos Esboços Múltiplos. Entretanto, as possibilidades de resposta às críticas de Todd precisam ser pensadas. Talvez exista uma forma da teoria de Dennett não ser afetada pelas evidências com metacontraste. Ainda, talvez existam outras maneiras de interpretar as novas evidências com metacontraste de

forma a ser favorável a Dennett. Não encontramos estas respostas, mas podem ser possíveis com mais estudo.

Acreditamos ter chegado a resultados satisfatórios com esta pesquisa, promovendo um contraste entre as duas teorias cognitivas da consciência capaz de mostrar vantagens e falhas em ambas as teorias, resolvendo algumas dúvidas, como a relação das críticas de Dennett sobre o Teatro Cartesiano e a Teoria do Espaço de Trabalho Global e levantando questões para pesquisas futuras.

REFERÊNCIAS

ANDERSON, J. R. *The architecture of cognition*. Cambridge: Harvard University Press, 1983.

ATKINSON, Rita L.; ATKINSON, Richard C.; SMITH, Edward E.; BEM, Daryl J; NOLEN-HOFKSEMA, Susan. *Introdução à Psicologia de Hilgard*. Porto Alegre: Artmed, 2002.

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

BAARS, B.J. *In the Theater of Consciousness: Global Workspace Theory*, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, v.4, n. 4, p.292–309, 1997.

BAARS, B.J. The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences*, v.6, n.1, p. 47–52, jan. 2002.

BAARS, B.J; Franklin, S. How conscious experience and working memory interact. *Trends in Cognitive Sciences*, v.7, n.4, p.166–172, abr. 2003.

BALDWIN, M. A New Factor in Evolution. *The American Naturalist*, Vol.30, No.355, p.536–553, 1896.

BODEN, M. *Mind as Machine: A History of Cognitive Science*. v.1. Oxford: Oxford University Press, 2006.

BLOCK, N. Book review of Dennett's *Consciousness Explained*. *Journal of Philosophy*, v.90, p.181–193, 1993.

BLOCK, N. On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, v.18, p. 227–287, 1995.

BRAITENBERG, V. *Vehicles: Experiments in Synthetic Psychology*. Cambridge: MIT Press, Bradford Book, 1984.

BREITMEYER, B. C. *Visual Masking*. Oxford: Oxford University Press, 1984.

BREITMEYER, B. G.; O GMEN, H. Recent Models and Findings in Visual Backward Masking: A comparison, review, and update, *Perception & Psychophysics*, v.62, p. 1572–95, 2000.

BREITMEYER, B. G.; O GMEN, H. *Visual Masking: time slices through conscious and unconscious vision*, Oxford: Oxford University Press, 2006.

BREITMEYER, B. G.; O GMEN, H.; CHEN, J. Unconscious Priming by Color and Form: different processes and levels, *Consciousness and Cognition*, v.13, p. 138–57, 2004.

BREITMEYER, B. G.; RO, T.; SINGHAL, N. S. Unconscious Color Priming Occurs at Stimulus-Not Percept-Dependent Levels of Processing, *Psychological Science*, v.15, p. 198–202, 2004.

BREITMEYER, B. G.; RO, T.; O GMEN, H.; TODD, S. Unconscious, Stimulus-Dependent Priming and Conscious, Percept-Dependent Priming with Chromatic Stimuli', *Perception & Psychophysics*, v.69, p. 550–7, 2007.

CARAMAZZA, A. How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, v.14, p.177–208, 1997.

CHANGEUX, J.P.; DEHAENE, S. The neuronal workspace model: Conscious processing and learning. In: J. Byrne and R. Menzel (eds). *Learning Theory and Behavior*. Volume 1 of Learning and Memory: A Comprehensive Reference. Oxford: Elsevier, 2008.

CHRISTIE, J.; BARRESI, J. Using Illusory Line Motion to Differentiate Misrepresentation (Stalinesque) and Misremembering (Orwellian) Accounts of Consciousness. *Consciousness and Cognition*, v.11, p.347–365, 2002.

COLTHEART, M, RASTLE, K, PERRY, C. DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, v.108, p.204-12, 2001.

CRICK, F.; KOCH, C. Towards a Neurobiological Theory of Consciousness. *Seminars in the Neurosciences*, v.2, 1990. p. 263-275.

DAMASIO, A. *Descartes' error: Emotion, reason, and the human brain*. New York: Avon Books, 1994.

DARWIN, C. On the origins of species by means of natural selection and the preservation of favored races in struggles for life. Londres, 1859.

DAVIS, P.; KENYON, D.H. *Of Pandas and People: The Central Question of Biological Origins*. Texas: Houghton Publishing Company, 1989.

DAWKINS, R. *The Selfish Gene*. Oxford: Oxford University Press, 1976.

DAWKINS, R. Universal Darwinism. In: BENDALL, D. S. (ed.) *Evolution from Molecules to Men*. Cambridge: Cambridge University Press, p. 403-425, 1983.

DAWKINS, R. *The Blind Watchmaker*. New York: Norton, 1986.

DAWKINS, R. *O Relojoeiro Cego*. São Paulo: Companhia das Letras, 2001 [1986].

DEBRUN, M. A Idéia de Auto-Organização. In: DEBRUN, M., GONZALEZ, M.E.Q.,

- PESSOA Jr, O. (orgs.) *Auto-Organização: estudos interdisciplinares*. Campinas: CLE/UNICAMP, p. 3-23, 1996 (Coleção CLE. v. 18)
- DEHAENE, S.; CHANGEUX, J.P.; Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, n.70, p.200-227, abr. 2011.
- DENNETT, D.C. *Brainstorms*. New York: Bradford Books, 1978.
- DENNETT, D.C. *Consciousness explained*. New York: Black Bay Books, 1991.
- DENNETT, D.C. *Darwin's Dangerous Idea: evolution and the meanings of life*. New York: Penguin Books, 1995.
- DENNETT, D.C. *Kinds of minds*. New York: Basic Books, 1996.
- DENNETT D.C. The part of cognitive science that is philosophy. *Topics in Cognitive Science 1*, p.231-236, 2009.
- DENNETT, D.C. Are we explaining consciousness yet? *Cognition*, v.79, p. 221–37, 2001.
- DENNETT, D.C; KINSBOURNE, M. Time and the Observer: the where and when of consciousness in the brain, *Behavioral and Brain Sciences*, v.15, p. 183–243, 1992.
- DERRY, Sharon J.; STEINKUEHLER, Constance A. Learning and instruction, Cognitive and Situated theories of. In: NADEL, Lynn (Ed). *Encyclopedia of Cognitive Science*. New York: John Wiley & Sons, p.1-6, 2005.
- DESCARTES, R. *Treatise on Man*. Paris, 1664.
- DOBZHANSKY, T. Nothing in Biology Makes Sense Except in the Light of Evolution. *The American Biology Teacher*, v.35, p.125–129, 1973
- EDELMAN, D., BAARS, B.; SETH, A. Identifying hallmarks of consciousness in non-mammalian species. *Consciousness and Cognition*, v.14, p.169–187, 2005.
- EDELMAN, G. *Neural Darwinism: theory of neuronal group selections*. New York: Basic Books, 1987.
- EDELMAN, G. *The Remembered Present*. New York: Basic Books, 1989.
- EDELMAN, G.M.; GALLY, J.A.; BAARS, B.J. Biology of Consciousness. *Frontiers in Psychology*. v.2, n.4, 2011.
- EDELMAN, D.; SETH, A. Animal consciousness: a synthetic approach. *Trends in Neurosciences*, v.32, n.9, 2009.
- ELIASMITH, Chris; BETCHEL, William. Symbolic versus subsymbolic. In: NADEL, Lynn (Ed). *Encyclopedia of Cognitive Science*. New York: John Wiley & Sons, p. 1-8, 2005.

EYSENCK, M. W.; KEANE M. T. *Cognitive Psychology: A Students Handbook*. 4.ed. Hove: Psychology Press, 2000.

FEHRER, E.; RAAB, D. Reaction Time to Stimuli Masked by Metacontrast. *Journal of Experimental Psychology*, v.63, p. 143—147, 1962.

FODOR J. *The Modularity of Mind*. Cambridge: MIT Press; 1983.

GREEN, D. *Cognitive Science: An Introduction*. Oxford: Blackwell, 1996.

HAYDEN, E. Life is Complicated. *Nature*. v.464|1, p.664-667, 2010.

LOW, P.; PANKSEPP, J.; REISS, D.; EDELMAN, D.; VAN SWINDEREN, B.; KOCK, C. The Cambridge Declaration on Consciousness. Cambridge, Francis Crick Memorial Conference on Consciousness in Human and non-Human Animals, 2012.

LYCAN, W.G. *Mind and Cognition: A Reader*. Oxford: Blackwell Publishers, 1990.

KAUFFMAN, S. Antichaos and adaptation. *Scientific American*, ago. 1991.

KAUFFMAN, S. *The origins of order*. New York: Oxford University Press, 1993.

KOSSLYN, Stephen M.; THOMPSON, William L.; GANIS, Giorgio. *The Case for Mental Imagery*. Oxford: Oxford University Press, 2006.

MACHADO, A. *Neuroanatomia Funcional*. 2ed. São Paulo: Atheneu, 2000.

MASLOW, A. *Motivation and Personality*. 2ed. New York: Harper & Row, 1970.

MADL, T.; BAARS, B.J.; FRANKLIN, S. The timing of the Cognitive Cycle. *Plos one*, v.6, n.4, 2011.

MASLIN, K.T. *An Introduction to the Philosophy of Mind*. Malden: Blackwell Publishers Inc., 2001.

MILLER G. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, v.63, p,81- 97, 1965.

NAGEL, T. What is it like to be a bat? *Philosophical Review*. v.4, p.435-450, 1974.

NEISSER, U. *Cognitive Psychology*. New Jersey: Prentice-Hall, 1967.

NEWELL, A.; SIMON, H. A. *Computer Science as Empirical Inquiry: Symbols and Search*, "Communications of the ACM", v.19, n.3, p.113-126, 1976.

NISHIMOTO, S.; VU, A.; NASELARIS, T.; BENJAMINI, Y.; YU, B.; GALLANT, J. Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, v.21, p.1641-1646, 2011.

O'BRIEN, G.; OPIE, J. A Defence of Cartesian Materialism. *Philosophy and Phenomenological Research*, v.59, p.939-63, 1999.

PAIVIO, A. *Mental Representations: A Dual Coding Approach*. Oxford: Oxford University Press, 1990.

PEPPERBERG, I.; LYNN, S. Possible Levels of Animal Consciousness with Reference to Grey Parrots (*Psittacus erithacus*). *American Zoologist*, v.40, p.893-901, 2000.

PEREIRA JR., A.; ALMADA, L. Conceptual Spaces and Consciousness: integrating cognitive and affective processes. *International Journal of Machine Consciousness*, v.3, n.1, p.127-143, 2011.

PEREIRA JR., A.; FURLAN, F. Astrocytes and human cognition, *Progress in Neurobiology*, v.92, p.405–420, 2010.

PEREIRA JÚNIOR, A.; PALEARI, L.M.; COSTA, F.A.P.L.; GUIMARÃES, R.C.; Evolução Biológica e Auto-Organização: apresentando, discutindo e exemplificando uma proposta teórica. In: I.M.L. D'Ottaviano, M.E.Q. Gonzales & G.M. Souza (orgs.). *Auto-Organização: estudos interdisciplinares*. Coleção CLE, v.39, 2004.

PEREIRA JÚNIOR, A.; PEREIRA, M.A.O. Teoria da Auto-Organização: uma introdução e possível aplicação nas ciências da Saúde. *Revista Simbio-Logias*, v.3, n.5, dez. 2010.

PINKER, S. *The Language Instinct*. Morrow, 1994.

PINKER, Steven. *How the mind works*. New York: W. W. Norton & Company, 1997.

PUTNAM, H. The Nature of Mental States. In W.H. Capitan & D.D. Merrill (eds.) *Art, Mind, and Religion*. Pittsburgh: Pittsburgh University press, 1967.

PYLYSHYN, Z. The Robot's dilemma: the frame problem in artificial intelligence. Cambridge: New Jersey: Ablex, 1987.

RAZRAN, G. The observable unconscious and inferrable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychological Review*, v. 68, p.81-147, 1961.

REINER, A.; Yamamoto, K.; Karten, H. Organization and evolution of the avian forebrain. *The Anatomical Record A: Discoveries in Molecular and Cellular Evolutionary Biology*, v.287, p.1080–1102, 2005.

ROCKWELL, T. Global Workspace or Pandemonium? *Journal of Consciousness Studies*, v.4, n.4, p.334–337, 1997.

ROZIN, P. The evolution of intelligence and access to the cognitive unconscious. In J. Sprague & A. Epstein (Eds.), *Progress in Psychobiology and Physiological Psychology*. New York: Academic Press, 1976.

RUSE, M. *Levando Darwin a sério*. Belo Horizonte: Itatiaia, 1995.

RYLE, G. *The concept of mind*. Hutchinson: Pinguim Books, 1949.

SCHMITZ, R.; SCHULTZ, M.; LEWSEY, M.; O'MALLEY, R.; URICH, M.; LIBIGER, O.; SCHORK, N.; ECKER, J. Transgenerational Epigenetic Instability is a Source for Novel Methylation Variants. *Science*, v.21, Out. 2011.

SCHNEIDER, S. Daniel Dennett on the Nature of Consciousness. In: M. Velmans and S. Schneider (eds), *The Blackwell Companion to Consciousness*, Oxford: Blackwell, p. 313–24, 2007.

SEARLE, J. *O Mistério da Consciência: e discussões com Daniel C. Dennett e David J. Chalmers*. São Paulo: Paz e Terra, 1998.

SETH, A.; BAARS, B.; EDELMAN, D. Criteria for consciousness in humans and other mammals. *Consciousness and Cognition*, v.14, p.119–139, 2005.

SHANNON C., WEAVER, W. *The Mathematical Theory of Communication*. Illinois: University of Illinois Press, 1949.

SPERLING, G. *The information available in brief visual presentations*. Psychological Monographs, v.74, n.11, 1960.

SOKOLOV, E. N. *Perception and the conditioned reflex*. New York: MacMillan, 1963.

SOKOLOV, E.N. The Neuronal Mechanisms of the Orienting Reflex. In E.N. Sokolov & O.S. Vinogradova (Eds.) *Neuronal Mechanisms of the Orienting Reflex*. New York: Lawrence Erlbaum, 1975.

TEIXEIRA, J.F. *A mente segundo Dennett*. São Paulo: Perspectiva, 2008.

TODD, S. J. Unmasking Multiple Drafts, *Philosophical Psychology*, v.19, p. 477–94, 2006.

TODD, S. J. A Difference That Makes a Difference: Passing through Dennett's Stalinesque/Orwellian Impasse. *British Journal of Philosophy of Science*, v.60, p.497-520, 2009.

TYE, M. Reflections on Dennett and consciousness. *Philosophy and Phenomenological Research*, v.53, p.893-898, 1993.

VON ECKARDT, B. Mental Representation. In: WILSON, A. Robert; KEIL Frank C (Eds). *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge: The MIT press, 1999. p.527-529