

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE FILOSOFIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

SIMULANDO DENNETT:
Ferramentas e construções de um naturalista

Diego Caleiro

São Paulo
2014

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE FILOSOFIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

SIMULANDO DENNETT:
Ferramentas e construções de um naturalista

Diego Caleiro

Dissertação apresentada ao Programa de Pós-Graduação em Filosofia, do Departamento de Filosofia da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, sob orientação do Prof. Osvaldo Pessoa Jr., para a obtenção do título de Mestre em Filosofia.

São Paulo
2014

Agradecimentos

Todo trabalho intelectual é precedido por uma longuíssima cadeia de eventos passados sem os quais ele não seria possível. Gostaria de destacar alguns desses eventos que me parecem subvalorizados. Primeiramente agradeço é claro a Stanislav Petrov por não destruir a Terra em 26 de setembro de 1983, e permitir a continuidade de tudo que tem valor no universo observável. Agradeço a meus pais por uma combinação de valores, recursos, confiança e modo de ver o mundo sem o qual esse texto não seria possível. Agradeço a Henry Oldenburg pela invenção do sistema Peer Review, que fomentou o rápido desenvolvimento da filosofia no mundo anglófono e deu origem ao pensamento dennettiano. Dois eventos foram de igual importância para que o conteúdo desse trabalho fosse tão distinto do que normalmente é feito em humanas na USP, o surgimento da internet ao longo de minha infância, e a permissividade de Osvaldo Pessoa Jr., meu orientador. Obrigado por mostrar os grandes lagos e ensinar a pescar, ao invés de vendar-me e afogar-me entre o cardume. Agradeço às mutações genéticas que permitiram o surgimento de humanos tão inteligentes e motivados quanto todos os citados nesse livro, e a estrutura social hierárquica complexa que permite que, numa sociedade de abundância, algumas pessoas possam dedicar-se a criar conteúdo intelectual e possam ausentar-se temporariamente da engrenagem. Agradeço às sugestões da banca, João de Fernandes Teixeira e João Vergílio Cuter, e de meus colegas do Instituto Ética, Racionalidade e Futuro da Humanidade. Agradeço por fim a todos os meus leitores presentes e futuros, sem os quais esse texto não seria mais do que um deleite intelectual pessoal, e não, de fato, um simulador do pensamento de Daniel Dennett.

Resumo

A dissertação pretende permitir ao leitor simular a forma de pensar de Daniel Dennett, e perpassa toda sua filosofia, com ênfase em seu tratamento de o que são padrões, o algoritmo evolutivo, *intuition pumps*, consciência, e seu uso dos conceitos de *illata*, *abstracta*, semântica e sintaxe para compreender a natureza, a biologia e a mente humana. O trabalho reapresenta, sob nova luz, grande parte das ideias mais importantes de Dennett, e procura fazer a engenharia reversa de o que o levou a pensar de determinadas maneiras, guiando o leitor através de caminhos similares, procurando fomentar um aprendizado ativo de uma forma de pensar, acima e além de uma exposição dos resultados obtidos ao longo de décadas desse pensamento no próprio Dennett.

Palavras chave: Daniel Dennett, Consciência, Memética, Perspectiva intencional, Evolução, Algoritmo.

Abstract

This dissertation intends to provide the reader with an inner simulation of Daniel Dennett's form of reasoning, spreading over his whole philosophy, emphasizing his treatment of patterns, the evolutionary algorithm, consciousness, and his use of *illata*, *abstracta*, semantic, and synthax, to carve nature at its joints, especially biology and the human mind. It recasts, in a new light, great part of his most important ideas, and reverse engineers what made him think in particular ways, walking the reader through similar pathways, fostering an active learning of a thinking style, above and beyond a mere exposition of the results obtained by this thinking style over the years.

Keywords: Daniel Dennett, Consciousness, Memetics, Intentional stance, Evolution, Algorithm.

Sumário

Apresentação	7
1. Reduccionismo	13
2. Informação e processamento	27
3. Seleção natural	33
4. Máquinas de Turing	51
5. Desafiando concepções intuitivas, desqualificando o sujeito-autoridade	65
6. Tipos de mentes	85
7. Os outros: a memética	93
8. Intuition pumps	117
9. Tornando conceitos antigos palatáveis	129
10. Consciência	153
11. O algoritmo Dennett	171
12. Intencionalidade, psicologia popular, e a necessidade de um plano de análise intencional	187
13. Padrões, consciência, significado, e a realidade dos <i>abstracta</i>	203
14. IA, ciência cognitiva, etologia e neurociências: ¿um cientista ou um filósofo?	211
15. Uma brevíssima história de tudo	215
16. A perspectiva dennettiana	223
Bibliografia	229

Apresentação

¿O que estamos fazendo aqui?

O texto desta dissertação é um simulador. Estamos aqui reunidos, o autor, você, e eu, esse conjunto de palavras e sentenças ordenadas, para simular a mente de um filósofo da mente. Estamos aqui reunidos para adentrar o mundo interior de Daniel Dennett. Nós, as palavras, e o autor delas, somos um conjunto de ferramentas cuja pretensão é lhe permitir criar um andaime no pensamento, e, se fizermos bem nosso trabalho, ao fim desse projeto você estará equipado para pensar com um kit de ferramentas parecido com o de nosso filósofo e tópico principal da dissertação, Daniel Dennett.

O projeto consiste em criar uma simulação unidimensional de uma mente multidimensional. Quando um computador simula uma parte da realidade, ele está, em seus comandos internos e partes físicas, simulando uma ação de cada vez. As etapas são demasiado rápidas, de modo que ele é capaz de simular personagens tridimensionais interagindo entre si em mundos relativamente complexos. Apesar de a informação numérica sendo processada ser sempre algo mensurável em uma só dimensão, ou quantidade, o mundo sendo simulado é multidimensional e complexo. Da mesma maneira, pretendo neste texto, palavra por palavra, sentença por sentença, construir, a partir de simples objetos sintáticos, a semântica cognitiva de Daniel Dennett. Aqui nos importará em primeiro lugar a forma de pensar de Dennett, e apenas secundariamente os conteúdos específicos propostos em sua filosofia. A contribuição original desse trabalho é não uma nova forma de pensar a filosofia de Dennett, nem uma forma mais resumida de apresentar *in totum* sua obra. A contribuição é, em realidade, permitir ao leitor que seja capaz de reflexões similares àquelas que seriam feitas por Dennett ao lidar com um novo problema em filosofia. A filosofia de Dennett, seus conteúdos, entram aqui não como *kernel*, como foco central, mas sim como *instrumentos* mentais, como ferramentas para a construção de um leitor capaz de pensar como ele. O sucesso desse projeto será medido não pela absorção, no leitor, dos conteúdos específicos estudados por Dennett ao longo dos anos, mas sim pela capacidade de escapar do tipo de problema que Dennett foi capaz de escapar, e de resolver, ou começar a resolver, o tipo de

problema que lhe garantiu uma carreira de quarenta anos como um dos grandes filósofos naturalistas do século vinte.

¿Quem é Daniel Dennett?

Dennett nasceu em Boston, EUA, em 1942, viveu sua primeira infância em Beirute, e retornou aos Estados Unidos aos cinco anos, após a morte de seu pai, num acidente de avião. Fez seus estudos de graduação na Universidade de Harvard, onde estudou com Willard Quine, e doutorado em Oxford, sob orientação de Gilbert Ryle. Dennett tem dois filhos e dois netos, e vive em Massachusetts. É autor de diversos livros que cobrem principalmente temas de filosofia da mente e da biologia. É arguivelmente o filósofo vivo de maior impacto na academia, com um *h-index*¹ de 67 em agosto de 2013, e entre o público leigo, mostrando mais de 116 mil resultados de vídeos numa busca do Youtube, substancialmente mais do que outros filósofos de penetração leiga como Chalmers, Habermas, Žižek, Bauman e Alain de Botton.

Além de sua atuação como filósofo e cientista, ele também é um artista diletante, principalmente de esculturas, e mais recentemente, em 2006, se uniu aos chamados “quatro cavaleiros” (*four horsemen*). Trata-se de uma alusão aos quatro cavaleiros do apocalipse, apelido dado a quatro autores que lançaram, num curto período, quatro livros contra o cristianismo ou a favor do ateísmo, com grande penetração na mídia de língua inglesa. Aqui falaremos mais sobre ciência e filosofia, e muito pouco sobre as discussões a respeito de religião das quais Dennett participou nos últimos anos.

Esse texto apresenta algumas das noções importantes trazidas por esse filósofo. Antes disso, e principalmente, o texto se estrutura de tal maneira a convidar o leitor à forma de pensar e às motivações que delineiam a filosofia de Daniel Dennett. Para isso, estarei arriscando dissociar-me parcialmente das ideias originais dele, como é inevitável num resumo. Não se trata somente de um resumo, mas de uma organização particular das ideias de Dennett com vista a tornar *pensável* todo um campo de ideias previamente impensável. Não necessariamente isso significa que deveremos percorrer suas ideias num sentido cronológico, afinal não é garantido que ele próprio tenha sido tão perspicaz em encontrar boas ideias. Essa ordenação acabará deixando boa parte das ideias

¹ Isso significa que as 67 publicações mais citadas de Dennett receberam, cada uma, no mínimo 67 citações. Somente encontrei um filósofo com h-index maior, David Lewis, com 67. Para comparação, alguns h-index de autores conhecidos: Steve Pinker: 67, Marilena Chaui: 36, Derek Parfit: 27, Saul Kripke: 22.

propostas por ele de lado. A meu favor, no entanto, forneço uma maneira de compreendê-lo que torna mais facilmente tangível o entendimento do caminho por ele percorrido, e portanto nos permitiria simular melhor, dentro de nossos cérebros, quais seriam os caminhos percorridos por ele em diferentes áreas do conhecimento que nos possam interessar. Acredito que essa perspectiva de “facilitador de simulação de Dan Dennett”, embora não muito usual numa reconstrução histórica do papel de um intelectual, fosse ser encorajada por um pensador como Dennett. O mundo é um lugar com uma quantidade limitada de intelectuais do calibre de Dennett, e facilitar uma linha de raciocínio como a dele é algo a meu ver importante. Outra propriedade que encontraremos nesse resumo é a repetição de algumas ideias. Tendo em vista o funcionamento das sinapses² humanas e o tipo de formação de redes conceituais que nossa cognição costuma fazer, a ideia de construir uma simulação irá por vezes incluir repetições. Diferentemente de um livro, as associações em nossa mente não são lineares. Um mapa conceitual de associação de ideias é algo difícil de se constituir a partir da linearidade de um texto, pois no mapa cada conceito está associado a diversos outros, e é importante que o leitor faça o máximo de associações possíveis entre os conceitos para poder operar com eles. Assim sendo, algumas frases talvez pareçam repetição de outras, e a função da repetição é conectar aquele espaço conceitual que já estava constituído com as informações que o estão rodeando num segundo ou terceiro aparecimento.

Dennett já defendeu que se nós não sobrevivemos à morte de nenhuma outra forma, sobrevivemos na forma de simulações computacionais dentro da mente de nossos conhecidos e amigos. Tenho certeza de que, se perguntado, ele não afirmaria uma identidade entre a simulação e o simulado, e portanto não vejo porque não assentiria em permitir que eu utilizasse o nome “Daniel Dennett” ao referir-me aquilo que, nesse texto, tentarei induzir o leitor a ser capaz de simular mentalmente.

Estilo, conteúdo e intenção: a tripla sincronicidade

Quando escrevi essa pequena introdução, supus que Dennett não se incomodaria com a escrita de uma dissertação que pretendesse simulá-lo, em oposição a contar sua história como ela se passou. Mencionei acima que seria interessante visar clareza ao explicar como Dennett pensa ao invés de, por assim dizer, imbuir-me de vocábulos

² *Sinapses* são as conexões entre as pontas de um neurônio e alguma parte de outro neurônio. A informação que o cérebro processa tem de passar pelas sinapses para ‘navegar’ entre os neurônios.

estouvados e pernósticos, exclusivos aos doutos e argutos filósofos acadêmicos.

Recentemente recebi dupla confirmação disso, o que redobrou minha confiança no projeto deste texto. Primeiro do próprio Dennett, num e-mail sobre outra questão que lhe ofereci, no qual ele menciona como sua motivação filosófica que sempre levou em conta um pensar que fosse além de uma motivação puramente acadêmica, e que pretendia penetrar mais círculos do que apenas o da academia:

Como menciono em meu livro mais recente, minha dissertação de doutorado foi criticada por um examinador (JZ Young) por seu estilo, que foi considerado demasiado casual, muito alusivo e muito metafórico. Mas desde o início eu descobri que não apenas os estudantes, mas também meus colegas filósofos “profissionais” – para não falar de cientistas e outros acadêmicos – apreciavam a leitura fácil da minha escrita. Alguns filósofos intelectualmente constipados menosprezam meus métodos, mas eu não me importo mais com eles. O fato é, a maioria dos leitores de filosofia lêem apenas por obrigação: o livro ou artigo está em seu tópico de escolha, e portanto eles são responsáveis por lê-lo, por mais desagradável que seja a experiência! As pessoas em uma grande variedade de campos leem meus livros porque eles gostam de lê-los. O mesmo vale para minhas palestras. A gente se amontoa em minhas palestras, e agora isso se multiplicou explosivamente graças ao Youtube. Mas meu propósito maior não é entreter. É mudar o que o mundo pensa a respeito dos tópicos nos quais trabalho. Eu levo isso muito a sério, e considero que é um anseio frequentemente negligenciado na filosofia. Ela não deveria ser um campo hermeticamente fechado, mas isso não significa que todos os filósofos devem seguir pelo meu caminho (Dennett, comunicação pessoal, junho de 2013).

A segunda confirmação veio de maneira composta. Durante esse trabalho, enveredei-me principalmente dos conteúdos que interessaram a Dennett até 2006. Não estava portanto mantendo-me atualizado de seus afazeres contemporâneos. A primeira surpresa que tive foi que, assim como meu próximo projeto é estudar evolução cultural, e a evolução cultural da moralidade, também o mesmo projeto é o próximo plano intelectual dele. A segunda surpresa foi que, assim como esse meu texto almeja inserir parte do conjunto de ferramentas ceifado por Dennett ao longo de anos para ter e transmitir ideias, também o mais recente livro dele, *Intuition pumps and other tools for thinking* (2013), tem como pano de fundo esse objetivo. Se algo poderia ser um teste empírico da validade de minha capacidade de simular a forma de pensar dennettiana, eu teria dificuldade de imaginar algo que me desse confirmação mais intensa do que essa tripla sincronidade, e espero que meu trabalho possa ser usado como um paralelo do novo livro de Dennett, mostrando aqui como ele mesmo implementou as ferramentas que lá divulga ao longo dos anos. Amplio também por extensão a confiança no modo de

escrita para leigos, que adoto em boa medida nesta dissertação, e nas ferramentas cognitivas utilizadas, como as interrogações invertidas do espanhol, os subtítulos advindos de Hofstadter e outros recursos estilísticos que visam facilitar e embelezar o processamento computacional dos conteúdos, no limite das possibilidades de um autor de primeira viagem.

Meta-Dennett

Começaremos do topo a nossa investigação dos vários níveis da hierarquia do pensamento dennettiano, com uma extensa citação que esclarece o espírito geral no qual se engendra o projeto filosófico de Dennett. Trata-se de uma filosofia que não vê separação clara entre filosofia e ciência. Estamos aqui no nível dos pensamentos sobre pensamentos sobre pensamentos de Daniel Dennett, em trechos de seu texto “Faith in the truth” (Dennett, 1997, pp. 6, 10, 11, 17-8):

Só nós podemos ser arruinados pela dúvida e só nós fomos impelidos por essa inquietação epistêmica³ a procurar uma cura: melhores métodos de procurar a verdade. Ao desejar um conhecimento mais adequado das nossas reservas de comida, dos nossos territórios, famílias e inimigos, descobrimos os benefícios de falar sobre isso com os outros, de fazer perguntas e de transmitir conhecimentos: inventamos a cultura. Depois, inventamos a medição e a aritmética, os mapas e a escrita. Estas inovações nas áreas da comunicação e do registro arrastam já consigo um ideal: a verdade. O sentido de fazer perguntas é encontrar respostas *verdadeiras*; o sentido da medição é medir de forma *precisa*; o sentido de produzir mapas é *encontrar* o caminho para o nosso destino. Pode existir uma Ilha dos Daltônicos (para usar a enorme dose habitual de liberdade poética de Oliver Sacks), mas não uma Ilha das Pessoas Que Não Reconhecem os Seus Próprios Filhos. A Terra dos Mentirosos só poderá existir nos enigmas dos filósofos; não há tradições de Sistemas de Calendários Falsos para registrar erradamente a passagem do tempo. Em suma, é evidente que o objetivo da verdade existe em todas as culturas humanas. [...]

Os métodos da ciência não são completamente seguros, mas podem ser constantemente aperfeiçoados. E o que é igualmente importante: existe uma tradição de crítica que obriga ao aperfeiçoamento sempre que se descobrem defeitos, e seja onde for que se descubram defeitos. Os próprios métodos da ciência, tal como tudo o que existe, são objeto do escrutínio científico, transformando-se os *métodos* em *metodologia*, a análise dos métodos. A metodologia, por seu turno, fica debaixo do olhar da *epistemologia*, a investigação da própria investigação – não há nada que escape ao questionamento científico. A ironia é que estes frutos da reflexão científica, que nos mostram as manchas indeléveis da imperfeição, são por vezes usadas por quem desconfia da ciência como pontos de partida para negarem a esta um

³ *Epistêmico* é aquilo que diz respeito ao conhecimento. Por exemplo, meu estado epistêmico é meu estado de conhecimento atual. Epistemologia é o estudo do conhecimento, ‘fome epistêmica’ é desejo por conhecimento, e assim por diante.

estatuto privilegiado na área da procura da verdade – como se as instituições e práticas que eles tomam como concorrentes da ciência não estivessem ainda em pior posição no que respeita a estas matérias. Mas onde estão os exemplos do abandono da ortodoxia religiosa face a provas irresistíveis? Na ciência, as heresias de ontem tornaram-se vezes e vezes sem conta as novas ortodoxias de hoje. Nenhuma religião exhibe este padrão evolutivo ao longo da sua história. [...]

Os cientistas têm fé na verdade, mas não uma fé cega. Não é como a fé que os pais têm na honestidade dos seus filhos, ou a fé que os adeptos desportivos têm na capacidade dos seus heróis para ganhar. É antes como a fé que qualquer pessoa pode ter num resultado a que várias grupos de pessoas chegaram de forma independente. [...]

É um fato que as pessoas não querem muitas vezes saber a verdade. E é um fato mais inquietante que as pessoas não queiram muitas vezes que os *outros* saibam a verdade. Mas, tentar transformar estes fatos de forma a que apoiem a ideia estúpida de que a própria fé na verdade é uma atitude humana relativa a certas culturas, situada ou em qualquer caso opcional, é confundir tudo. O pai do acusado que ouve em tribunal os testemunhos contra o seu filho, a mulher que se pergunta se o marido a anda a enganar – eles podem muito bem não querer saber a verdade, e podem ter razão em não querer saber a verdade, mas o fato é que acreditam na verdade; isso é claro. Eles sabem que a verdade está aí, para ser evitada ou abraçada, e sabem que a verdade é importante. É *por isso* que eles podem muito bem não querer saber a verdade. Porque a verdade pode magoar. Podem conseguir enganarem-se a si mesmos, pensando que a atitude que têm nestas ocasiões perante a verdade reflete um defeito da própria verdade, assim como da própria procura e descoberta da verdade – mas se isto acontecer é puro autoengano. O máximo a que podem aspirar agarrar-se é à ideia de que podem existir boas razões, as melhores razões – no tribunal da verdade, note-se – para, por vezes, suprimir ou ignorar a verdade.

1. Reduccionismo

A primeira coisa que pode ser dita a respeito de Dan Dennett (como é conhecido) é que ele é um entusiasta do reduccionismo. Em sua obra como um todo subjazem dois desejos que têm importante papel. O desejo de reduzir uma explicação aos seus constituintes mínimos e o desejo de eliminar as imprecisões que possa haver em conceitos filosóficos, inclusive, se isso for impossível, eliminando os próprios conceitos.

Essa estratégia reducionista é largamente condizente com toda a tradição filosófica pragmática americana na qual Dennett se encontra, mas existem particularidades em suas ações reducionistas que nos convidam a ir além de simplesmente qualificá-lo como um filósofo americano pragmático e seguirmos adiante. Na maior parte de seus escritos, as preocupações de Dennett estão em explicar a forma de alguma entidade⁴, e procurar alguma maneira não circular de falar sobre ela. Ao falar por exemplo sobre a “experiência”, no artigo “Os sonhos são experiências?” (Dennett, 1976a), sua preocupação é compreender qual é a forma geral do que é uma experiência, e como podemos inferir se algo dessa natureza estava ou não ocorrendo em algum momento do sono quando, posteriormente, relatamos o sonho para outra pessoa. Quando se propõe a discutir a evolução da cultura humana, tenta encontrar alguma perspectiva na qual o elemento explicativo seja bastante diferente do objeto a ser explicado. A cultura é tomada não da perspectiva global adotada por um antropólogo cultural ou um sociólogo, mas da perspectiva de uma ideia particular e sua capacidade de sobreviver. Esse tipo de inversão perpassa toda a obra de Dennett, e por isso suas explicações muitas vezes geram *insights*⁵ distintos dos usuais. Passaremos por vários

⁴ *Entidade* é o termo que filósofos usam no lugar de “coisa” no senso comum. Tudo que é uma coisa é uma entidade. Antigamente pensava-se que uma boa definição de entidade é “aquilo que tem ser”. Evidente que essa não é uma boa explicação, primeiro porque não define o que é *ser*, segundo, porque não demonstra de que maneira *ser* poderia ser uma propriedade de algo. Essa má concepção antiga está sendo gradativamente eliminada pela psicolinguística, que mostra como utilizamos palavras de maneiras incorretas, e porque fomos feitos para fazê-lo (Pinker, 2007a; Lieberman, 2009; Boroditsky, 2009). Não conseguimos pensar que algo *é* sem pensar que esse algo tem a propriedade de ser, da mesma maneira que a bola tem a propriedade de ser redonda. Mas isso tem mais a dizer sobre nossa cognição do que a respeito das coisas. Portanto, quando encontrar “entidade” num texto filosófico, substitua por um uso corriqueiro da palavra “coisa”. Há muito ganho em precisão quando percebemos aquilo que *não* é útil para explicar algo, e atribuir propriedades como *ser* a alguma coisa não nos explica nada sobre ela. O termo “entidade” não contém mais informação do que o termo “coisa”.

⁵ Há duas maneiras de se pensar o conceito de *insight*: ou como um ato de entendimento ou esclarecimento de uma ideia, uma associação de conceitos que gera nova perspectiva; ou num sentido equivocado, uma ideia boa *sem origem*, que, na realidade é uma ideia boa cuja origem se encontra nas subpartes inconscientes de nossas mentes.

conceitos ao longo desta dissertação, e tentaremos aplicar a visão dennettiana a eles, tanto para compreendê-los quanto para operar com eles. Citarei agora vários conceitos conforme ele os pensa, com uma breve descrição do significado de cada um. Não é necessário, *e nem esperado*, saber agora ou entender agora nenhum desses conceitos, ou sequer saber porque a descrição (estranha) aqui feita é uma boa forma reducionista de destrinchá-los. Tudo isso virá à tona conforme a necessidade. Imagino entre meus leitores graus diferentes de conhecimento da obra de Dennett, e os conceitos a seguir serão mais relevantes para aqueles que *já a conhecem*, e somente quando forem reutilizados, no final do livro, fará sentido aqueles que iniciam sua interação com o pensamento dennettiano nesse texto.

Pessoa: Centro de gravidade narrativo, um *abstracta* que organiza nossas representações mentais a respeito de outros e de nós mesmos. Centro do conjunto de histórias atribuídas a um indivíduo.

Abstracta: Qualquer coisa que tem o mesmo grau de realidade que um centro de gravidade, que existe, mas não tem eficácia causal, que *está lá*, mas não é causalmente responsável por nada. Ainda assim podem figurar nas explicações causais como parte de um modelo que as descreva.

Máquinas de Turing: *Abstração matemática* que a partir de regras muito simples consegue executar tarefas complexas, desde que programada para tal e desde que haja um sistema de codificação e decodificação.

Computador: *Objeto físico* que a partir de regras muito simples executa operações muito parecidas com uma Máquina de Turing, e as disponibiliza em interfaces cada dia mais naturais.

Mente: “Si abbiamo un’anima. Ma è fata di tanti piccoli robot.” Tradução: Sim, temos uma alma, mas ela é feita de muitos pequenos robôs.

Evolução: Explicação para todo o *design* que persiste no mundo, *only game in town* (i.e., única possibilidade). Perigosa ideia de Darwin, pois corrói todas as estruturas e ameaça convicções milenares; principal explicação para os botos, o botox, os botões e os batons, todos produzidos pelo mesmo algoritmo.

Algoritmo: Conjunto de operações extremamente simples, descritíveis numa única linguagem, que qualquer idiota conseguiria executar. Sequência ordenada de operações simples.

Sistemas intencionais: Sistemas para os quais a melhor estratégia preditiva é supor que eles tenham *crenças, intenções e desejos*, e uma racionalidade parcial

limitada a seus mundos nocionais.

Sintaxe: Conjunto de regras formais de operação dentro de um sistema *ou* conjunto de operações ocorrendo em sistemas que não têm processamento de “significado”, por exemplo sistemas físicos.

Memes: Unidades replicadoras a partir das quais a evolução cultural deve ser pensada. Atratores no espaço conceitual de ideias (cf. Henrich, Boyd & Richardson, 2008) que tendem a se propagar de maneira matematicamente equivalente a unidades replicadoras.

¿Como Dennett vê memes e sistemas intencionais?

Por hora, estes conceitos ainda estão pouco determinados. Joguemos mais um pouco com eles. Dennett, quando discute o conceito de “memes”, unidades replicativas de elementos da cultura, nomeadas por Dawkins, preocupa-se com as seguintes questões: ¿Quais estruturas possíveis poderiam provocar o surgimento e manutenção da cultura, sem serem elas mesmas estruturas culturais? ¿Como fugir da circularidade de explicar o surgimento e complexificação da cultura a partir de uma entidade não essencialmente cultural? Já ao falar sobre sistemas intencionais (Dennett, 1971), preocupa-se fundamentalmente em descrever em que formas é possível que um organismo se organize de tal maneira que ele não possa ser descrito nem por uma análise física/geométrica, nem por uma análise em termos de “para o que foi feito” ou de design. É dito de um sistema intencional que se pode dizer que ele *crê* (Dennett, 1981a). Para Dan, um sistema *acreditar* em algo seria basicamente ser um sistema cuja forma (organização estrutural/disposicional) é tal que, conforme ele interage com outras formas, animadas ou não, esse sistema modificaria essas outras formas de uma determinada maneira que condiz com aquela crença. Crer é agir como se acreditasse. Comportar-se de tal modo que uma interpretação na qual se acredita é adequada e parcimoniosa. Não há um lugar para uma crença em seu cérebro. Acreditar em algo seria instanciado⁶ não em na matéria, como em um conjunto de neurônios particular,

⁶ Filósofos usam o termo “instanciado” para significar cada uma das ocorrências de um certo tipo de coisa. Por exemplo, posso pensar em dor de dente. Quando falo sobre dor de dente, estou falando sobre um *tipo* (espécie) de dor, e o termo “dor de dente” vale para todos os casos particulares de dor de dente, ou para o agrupamento deles. Já quando falo de uma *instância* de dor de dente, me refiro a uma dor de dente particular que Dawkins sentiu, no dia 10 de dezembro de 1988, às 10 da noite, no segundo molar superior direito. Em inglês há uma diferenciação, que não é muito fácil de traduzir para o português, entre *type* e *token*, sendo o *token* esse caso particular de instanciação de um *type*. A tradução de ‘*token*’

numa determinada cabeça, num momento, mas numa parte de um sistema integrado que altera o ambiente de modo a produzir um comportamento que condiz com o que chamamos, por exemplo, de “acreditar que a água é líquida”.

Digamos que Taly acredita que a água preenche metade do copo. O que Dennett diria, o que garante a verdade dessa afirmação, é que Taly é uma pessoa organizada de tal maneira que, quando questionada sobre a água, ou quando estiver com sede, ou quando jogarem água na cara dela, agirá da forma como esperaríamos que agisse alguém que acredita que a água esteja até a metade do copo. Isto é, Dennett está reduzindo o conceito de *acreditar* a propriedades mais simples, ao conjunto das propriedades *agir de maneira que condiz com acreditar em algo*.

Reduzindo conceitos a formas mais simples

Poderíamos atribuir a Dennett um tipo particular de reducionismo, o reducionismo de formas, em oposição, por exemplo, a um reducionismo de substâncias. Não porque ele não seja um reducionista de substâncias (um monista), mas porque o que mais está operante em seu trabalho, o que cumpre papel mais forte em seus escritos, é a busca de compreender quais são as formas (formatos, estruturas) mínimas que permitem o funcionamento de um determinado processo (consciência, passagem de ideias, crenças, autonomia, seleção natural). O reducionismo de substâncias para ele possui um papel menos importante, por ser tão certo. Cumpre a ele o papel de compreender os diferentes níveis de organização da matéria (info, bio, cogno) e tornar claros os processos através dos quais se podem derivar uns dos outros, e principalmente porque outros níveis são necessários.

Tomemos um caso: em seu livro *A perigosa ideia de Darwin* (1995a, pp. 412-5), ele propõe um experimento filosófico no qual duas caixas pretas conectam-se uma à outra, por um único fio. Uma delas tem dois botões, a outra três lâmpadas. Sempre que se aperta um botão de uma caixa, a lâmpada 1 se acende na outra caixa. O outro botão também se correlaciona imensamente com a lâmpada 2. A terceira lâmpada apenas

tem variado entre ‘ocorrência’ e ‘instância’. Dois *tokens* diferentes podem instanciar de um mesmo *type*, mas não inversamente. Dois *tokens* do mesmo *type* podem (segundo alguns filósofos) ser perfeitamente semelhantes, no sentido de todos os seus aspectos qualitativos serem iguais. Talvez haja uma dor de dente canino que já foi sentida exatamente igual por duas pessoas em ocasiões diferentes, por exemplo. O *type* é o nível que descreve [essa dor de dente] e o *token* é o nível que descreve [essa ocorrência particular de [essa dor de dente]]. Mantereí o termo “*token*” e outros termos em inglês, ao longo do texto, para lembrar que eles estão sendo usados com um significado específico.

acende a cada bilhões de apertadas. O padrão de informação enviado através do único fio, no entanto, é diferente (mas do mesmo tamanho) a cada vez que se aperta um botão. Cientistas, intrigados com esse artefato singular, procuram compreender o padrão que o faz assegurar a conexão. Fazem todas as análises causais físicas (sintáticas) possíveis, para compreender como as duas máquinas conseguem se comunicar, e procuram entender porque sempre que um botão é apertado, uma luz correspondente do outro lado é ativada. No entanto não conseguem desvendar o mistério. Além disso, ao cortar o fio e gerar um *input* qualquer daquele comprimento padrão, a terceira lâmpada se acende. E o mesmo acontece sequencialmente com quaisquer sequências de informação que se envie para a segunda caixa, exceto as que já estavam no histórico do laboratório como emitidas pela primeira caixa no passado. Estupefatos com o engenho, desistem de compreendê-lo, e perguntam aos inventores como fizeram as caixas. Os inventores explicam que uma delas é uma caixa de composição de verdades, que escreve sentenças verdadeiras a partir de uma imensa base de dados numa linguagem (inglês escrito em ASCII, o código padronizado que correlaciona bits com letras, usado no envio de e-mails) e a outra é uma máquina interpretadora, que tem uma base de dados diferente, porém construída também nos mesmos parâmetros de sinalizar acendendo uma lâmpada ao receber uma proposição verdadeira, e outra ao receber uma proposição falsa. Com esse experimento, Dennett pretende deixar claro – para aqueles que acham que todas as coisas podem ser explicadas através da sintaxe, de regras rígidas, e de estruturas sintáticas – que a semântica ainda tem seu lugar, e que a perspectiva de análise semântica não deve ser abandonada, pois os cientistas conheciam todas as informações sintáticas das caixas pretas, e ainda assim não eram capazes de fazer uma interpretação semântica, não conseguiram *desvendar o porquê* do acender de luzes, que teria sido imprescindível para a compreensão do funcionamento delas. O uso da ideia de *significado* das sentenças é condição *sine qua non* para compreender porque cada lâmpada acende com que input, e a semântica é por excelência o campo dos significados. Em outras palavras, ele quer dizer que o *plano de análise* da semântica (e seus termos) é demasiado rico para ser reducionisticamente abandonado por completo; há algo ali que deve ser utilizado, mesmo que se possa em princípio fazer uma análise causal (sintática) perfeita de todas as nuances físicas de um determinado agente.

Esse espírito de análise presente no experimento mental ecoa a famosa frase atribuída (interpretativamente) a Einstein: “Devemos tornar as coisas o mais simples possível, mas não mais simples que isso”. De fato, Dennett posiciona-se vigorosamente

contra um tipo de reducionismo, que ele chama de reducionismo cobiçoso (*greedy reductionism*). Para compreendermos melhor o que é o reducionismo cobiçoso, temos de compreender seu inverso, a procura por Skyhooks.

Procurando por *skyhooks*

O conhecimento racional pode vir de qualquer linha de pesquisa, e essa tem sido a principal razão para a existência de considerável liberdade na escolha que pesquisadores fazem de seus tópicos de estudo, se comparados, por exemplo, à liberdade de um construtor de carros na maneira como ele irá realizar seu trabalho. É da liberdade acadêmica, e um pouco menos da liberdade intelectual garantida por um escritório de patentes na Suíça, que vêm as mais interessantes ideias que posteriormente se tornam *mainstream*. Liberdade de escolha, quando se trata das forças por trás da pesquisa aparentariam ser, portanto, uma técnica bastante sábia para a obtenção da melhor qualidade de produção. Mas isso não parece ser a verdade. Existe um hábito da mente que, como parte dos programas de pesquisa, causou mais mal do que bem. Esse hábito, seguindo a apreciação de Dennett em *A perigosa ideia de Darwin* (1995a), eu chamarei de “procurando por skyhooks”.

Um *skyhook* (gancho-celeste) na explicação dele é:

uma força ou poder ou processo que ocorre primeiro na mente [*mind-first*], uma exceção ao princípio de que todo design, e design aparente, é em última instância o resultado de mecanicidade sem motivos e sem mentalidade. Um guindaste [*crane*], por contraste, é um subprocesso ou traço especial de um processo de design que pode ser demonstrado permitir a aceleração de um processo lento e básico da seleção natural, e que pode ser demonstrado ser ele mesmo um produto previsível (ou retrospectivamente explicável) do processo básico (Dennett, 1995a, p. 76).

Os skyhooks são legatários de uma ideia de *deus ex machina*, a figura divina fictícia que salva o herói quando o autor o meteu numa enrascada sem saída. Ao criarem histórias, crianças tipicamente fazem uso de uma espécie de *deus ex machina* simplificado: “E então ele encontrou o dragão e ficou preso na caverna e quando o dragão ia morder ele, ele, ele... Acordou, era tudo um sonho e ele viveu feliz para sempre! Fim.” No teatro grego essa função era cumprida principalmente por seres divinos (e portanto antropomórficos); a sugestão era que houvesse uma mente capaz de, acima e além das forças naturais, resolver um problema.

No mundo da ciência é muito comum que haja pessoas que fazem o mesmo tipo de suposição, *aquele* objeto não pode ter surgido por meios naturais, *tem de haver* alguma mente, uma interferência superior que possibilite tal maravilha. Mas de novo e de novo, aquilo que não podia ser explicado com as ferramentas da ciência de uma década acabam sendo explicados na década seguinte, como um subproduto sofisticado mas apenas composto dos produtos básicos anteriores, com o auxílio de guindastes, eles mesmos também feitos do produto original. E então os procuradores de skyhooks voltam sua esperança para alguma outra maravilha do mundo que simplesmente *precisa* ter sido criada por algo que veio do céu.

Dennett encontra procuradores de skyhooks no darwinismo. Primeiro pontuando que todo o design foi criado por este processo cego e mecânico, e depois reforçando:

Poderia isto ter realmente acontecido? Ou o processo precisou de uma “mãozinha” vez por outra (talvez apenas no comecinho) de algum tipo de skyhook? Por mais de um século, céticos têm tentado encontrar uma prova de que a ideia de Darwin simplesmente não pode funcionar, ao menos não *de maneira completa*. Eles têm desejado, caçado, rezado por skyhooks, como exceções ao que eles veem como a sinistra visão do algoritmo darwiniano seguindo seu rumo [*churning away*]. E de vez em quando, eles aparecem com desafios muito interessantes – saltos e lacunas e outras maravilhas que parecem, de início, necessitar de skyhooks. Mas daí vieram guindastes, descobertos em muitos casos pelos próprios céticos que estavam esperando encontrar um skyhook (Dennett, 1995a, pp. 75-6).

Moderação contra oponentes intelectuais não é uma qualidade extensamente presente no trabalho de Dennett, mas esse parágrafo me parece ser moderado demais, uma particularidade que é corrigida no restante do livro. Ainda assim, existem algumas coisas pontuadas nele que serão úteis ao nosso escrutínio, num sentido neutro, mas não moderado.

É claro que é possível que os guindastes sejam descobertos pelos céticos que estavam desejando um skyhook. Considerando que a maioria dos céticos com relação ao darwinismo são raramente céticos também com relação a suas próprias inclinações mentais e hábitos mentais, e muito mais céticos a respeito de estatísticas (já que existe bastante evidência a favor, e muito pouca contra, o darwinismo), é altamente improvável que eles tenham um papel significativo na descoberta de guindastes. Não devemos tolerar ou aceitar com moderação a busca por skyhooks numa conversa genérica, já que este seria o caso das pessoas que estão rezando por, e desejando skyhooks, nas palavras de Dennett. Num grau muito menor, creio que as pessoas que caçam skyhooks, os cientistas que buscam um “algo além”, uma explicação com

adornos extras de natureza aparentemente mística, também permanecem um obstáculo para a construção do conhecimento racional. No caso do desejador e do orador é fácil ver que eles não fazem nada senão alimentar inclinações ilógicas, crenças anti-estatísticas e desejos improváveis das pessoas a seu redor. Ao caçador, por outro lado, não podemos atribuir que esteja fazendo a mesma coisa, já que ele está executando um programa de pesquisa, e portanto ele poderia descobrir que de fato está certo, e se esse fosse o caso, então haveria skyhooks, e ele de fato teria criado conhecimento. É sobre a possibilidade de que as pessoas estejam certas que se sustenta a liberdade acadêmica, mas a liberdade acadêmica não pode ser confundida com o relativismo. Não pode haver uma realocação de recursos do mesmo montante à hipótese de que o câncer é causado por um pote de chá orbitando Marte, à de que é causado pela devastação das florestas, e à de que é causado por algum tipo de mau comportamento celular explicável bioquimicamente. A razão para isso é que, ainda que esses três programas de pesquisa possam de fato se revelar corretos, eles têm uma probabilidade muito diferente de fazê-lo. Já que há evidência estatística (para não dizer lógica!) a favor da hipótese da célula, o defensor da devastação florestal tem o ônus da prova de mostrar porque é necessário estudar seu caso. Ainda mais difícil é o caso do defensor do pote de chá, já que ele também carrega o ônus da prova da existência do pote de chá orbitando Marte, que não exibiu particular evidência desde que Russell escreveu a respeito dessa possibilidade (e improbabilidade) muitas décadas atrás.

O caçador, quando ele se propõe a procurar um skyhook, tem o ônus da prova de mostrar porque é provável que ali exista um, já que o passado não mostrou evidências de skyhooks. Já que sempre houve uma grande quantidade de dinheiro sendo alocada na pesquisa por skyhooks fora do mundo acadêmico, afinal, os oradores – i.e. religiosos norte-americanos – pagam os caçadores – i.e. defensores do design inteligente – para continuarem buscando por deles. De um ponto de vista dennettiano, o mundo acadêmico e o mundo do conhecimento racional em geral não estariam ameaçados se todo o dinheiro fornecido aos caçadores de skyhooks fosse utilizado para atividades que têm maior chance de produzir conhecimento, como pesquisa em bioquímica ou alimentar sapos com caviar.

Reduccionismo cobiçoso

Como a intuição humana nos guia em direção a tornar coisas *mais místicas e mais mágicas* do que realmente são, o “reduccionismo cobiçoso” (*greedy reductionism*) é menos comum, e também menos perigoso que seu contrário, a procura por skyhooks, tendo seu nome também tirado de *A perigosa ideia de Darwin* (Dennett, 1995a, p. 82).

Os reducionistas cobiçosos pensam que não haja guindastes (*cranes*), e que tudo deve poder ser explicado em termos simplistas, sem nenhum tipo de alavancagem interna do sistema. Pulam da física de partículas para a sociologia em apenas um passo, sem percorrer todas as alavancagens intermediárias que normalmente pensamos entre essas áreas.

O bom reducionista seria aquele que por um lado não procura nada no céu para auxiliá-lo, ou seja, não procura skyhooks, e por outro lado leva em conta a existência de processos internos ao sistema de alavancagem que permitem novos níveis de complexidade, os guindastes.

Toda vez que alguém tenta fazer uma teoria sobre um fenômeno, ela deve, ao menos em princípio, poder dar origem a um programa de pesquisa que possa compreender partes do fenômeno, ou aceitar o fato de que há uma parte do fenômeno que não pode ser atacada naquele nível de explicação.

Um exemplo clássico é uma crítica no texto de Thomas Nagel “Como é ser um morcego” (1974), na qual ele critica o fato de não termos uma linguagem descritiva de estados mentais que seja independente do observador, e portanto nós não podemos sequer começar a falar como é ser um morcego. A razão para não termos essa linguagem é que não temos uma teoria dos estados mentais no nível correto de análise, e o nível de análise no qual temos uma forma descritiva de falar sobre eles (o nível físico ou biológico, por exemplo) não são suficientes para uma teoria da mente, quanto menos da mente de morcegos.

O que Nagel está pontuando é que ele acha que a filosofia da mente na época estava sendo uma reducionista muito cobiçosa, já que os níveis de explicação que ela propunha (só o biológico e o físico) não eram suficientes para uma teoria completa.

A fronteira entre o reduccionismo cobiçoso e o bom reduccionismo é bem menos clara do que a fronteira entre bom reduccionismo e a procura por skyhooks. Isso basicamente porque no primeiro caso ambos os lados tem o ônus da prova, já que o Cobiçoso Jônatas pode argumentar que “Todo fenômeno até hoje foi explicado somente

com este e aquele níveis de análise”, ao que Parcimonioso Harry responderá “De fato, mas há alguns fenômenos que existem de fato, como X ou Y que ainda não são explicáveis nesses níveis e que podem ser explicados no nível Z.” Já que a questão de ser o nível Z requerido ou não para uma explicação é uma questão empírica, deve haver investimento em ambos os lados, e o reducionismo cobiçoso tem seu lugar garantido.

Não é no reducionismo cobiçoso acadêmico que o problema para o conhecimento reside. Caçadores cobiçosos são sempre bem vindos. São os oradores e desejadores que são novamente perigosos, já que eles comprometem a possibilidade até mesmo de levar em consideração a possibilidade do nível Z de explicação. O Behaviorismo era uma boa ideia, e uma boa tentativa, mas agora é evidente que a psicologia precisa de um nível de explicação que explore mais níveis do que apenas condicionamento operante, condicionamento no nível físico. A possibilidade de entender os erros de Skinner só existiu porque alguém foi um bom reducionista, e não existe razão em particular para advogar publicamente o reducionismo cobiçoso mais do que existe para advogar que não criemos novos níveis de análise da economia ou da política. Explicar quase tudo em termos de quase nada não é algo ruim, mas é importante que de fato haja uma explicação, e não apenas uma petição apelativa que declara que o nível superior pode ser explicado em termos do inferior, mas não demonstra como isso poderia ocorrer.

Um parágrafo de função mnemônica

Nesse momento, e ainda outra vez ao fim do texto, vou inserir um parágrafo cuja função não é esclarecer ou ilustrar. Sua função será, retrospectivamente, a de perceber o aprendizado obtido ao longo do texto, bem como conectar conceitos de partes separadas na escrita. O parágrafo, que apresento aqui, e mais ao fim do texto, provavelmente não será entendido agora, e isso é esperado:

Essa forma exagerada de reducionismo, presente por exemplo no pensamento do psicólogo behaviorista Skinner, seria basicamente um desejo de reduzir absolutamente um fenômeno de larga escala a um fenômeno de pequena escala. Mas isso sem procurar compreender e reconhecer a existência de uma reestruturação numa hierarquia intermediária que permite que o fenômeno se comporte de maneira diferente numa escala maior do que seria esperado que ocorresse se estivéssemos apenas prevendo o comportamento microscópico de suas partes, não porque as macro partes se comportem

mal, mas porque nós não conseguimos calcular seu comportamento. A razão pela qual essas reestruturações se dão, segundo Dennett, é a seleção natural, na qual a existência de replicadores permite uma guerra de reprodutibilidade mais acelerada, e esta por sua vez gera novas estruturas materiais que se reproduzem mais rápido que suas antecessoras, ganhando espaço possível de design (Dennett, 1978b).

¿Quê?

Dennett critica Skinner, da antiga escola do behaviorismo em psicologia, considerando as análises desse pensador reducionistas ao extremo, e procura mostrar, principalmente nos artigos “Skinner esfolado” (1978b) e “Sistemas intencionais” (1971), como o funcionamento proposto por Skinner para o comportamento psicológico de agentes não é capaz de dar conta da complexidade do processo de aprendizado, que inclui por exemplo elementos referenciais (referências a algo externo registradas na memória, em oposição a reações a um dado da experiência). Além disso, critica a posição behaviorista tradicional ao condenar a perspectiva de procurar compreender a mente como uma caixa preta, apenas a partir de seus inputs e outputs⁷. Dennett percebe e ilustra a importância da forma, da estrutura interna da mente, para compreender seu funcionamento, e indica que uma análise que proceda através do estudo apenas dos impulsos deferentes e aferentes de um indivíduo não será capaz de dar conta da tarefa da psicologia. A tarefa da psicologia, em sua concepção, seria algo como explicar o funcionamento mental em termos de algo que não sejam elementos do mesmo tipo que se planeja explicar. A explicação do funcionamento da mente como um teatro cartesiano, por exemplo, não pode ser levada a cabo, pois ela necessitaria de um homúnculo interno que dispusesse de todas as propriedades do homem que se está tentando explicar, e seria necessário então explicar como opera a mente do homúnculo dentro do teatro. Vale notar que isso não significa que não se pode permitir que algumas das propriedades da mente de fato estejam sendo controladas por um demônio interno. O demônio interno é uma explicação possível quando ele possui *menos* propriedades e *menos* complexidade do que a mente a qual pertence (Dennett, 1991a, 1994). Esse demônio, por sua vez, poderia ser constituído de demônios menores, ou de outras entidades que não fossem ele.

⁷ *Input* é o nome da informação que entra em um sistema. A informação que sai é o *output*. Tanto o input quanto o output podem estar instanciados em comportamentos, manipulações, movimentos físicos ou simplesmente em sistemas simbólicos ou sequências digitais.

Então ele serve para construir uma parte da ponte que a psicologia, na concepção de Dennett, deve construir. A diferença conceitual entre o que se define como homúnculo e como demônio é fundamentalmente essa. O homúnculo não tem poder explicativo, porque ele apenas levaria a uma infinidade de inexplicados, enquanto o demônio é definido como tendo ao menos parcialmente o poder de explicar o comportamento de uma entidade mais complexa do que si mesmo (a mente, ou outro demônio, caso ele seja, digamos, de uma hierarquia inferior).

Podemos ver até aqui que existe na filosofia de Dennett uma grande preocupação em evitar postular aquilo que se deseja em termos de reducionismo, isto é, ele não afirma que um nível Y – i.e. intenções – de complexidade é *evidentemente*⁸ composto por um nível mais baixo, como quarks, por exemplo, ou dizendo que toda a psicologia é redutível à física. Ao invés disso ele procura colocar-se na tarefa de compreender justamente como essa liga é possível. ¿Em que nível de análise se poderia dizer que a psicologia pode ser reduzida à física? Dennett nos mostra que vale a pena trabalhar, por vezes, de diversas perspectivas de análise (Dennett, 1981b), através das quais possamos localizar e compreender melhor o porquê tudo indica que a psicologia é redutível. No entanto, o processo de de fato executar a redução tem se mostrado tão complexo, tão difícil para a ciência e para a filosofia, que Dennett adota a posição de parar em alguns pontos no meio do caminho entre os níveis de análise e dar a perspectiva de como observar daquele ponto intermediário. Se imaginarmos, de um lado de uma montanha, conceitos abstratos como mente, pessoa, inteligência, arbítrio, e do outro lado explicações de nível físico simples, e virmos a ciência e a filosofia cavando o túnel que promete ligar esses dois lados, Dennett seria o engenheiro de projeto da zona intermediária, pretendendo observar a área e criar a forma geral do projeto de escavação. Tornando claros os pontos intermediários (o que discutirei mais no cap 13), Dennett abre caminho para que a filosofia consiga barrar as infiltrações no projeto de redução, e é nesse sentido que deve ser considerado como um reducionista.

⁸ Ao contrário, em seu último livro (Dennett, 2013, p. 53) ele inclusive cita o “*surely*” (evidentemente, obviamente, claramente) como um ótimo indicador, num artigo filosófico ou num conversa, de onde o ponto mais fraco provavelmente se encontra.

Reduccionismo, funcionalismo e verificacionismo

Explicação terminológica para filósofos: dentro da filosofia da mente o termo ‘reduccionismo’ tem diferentes significados. Numa definição abrangente, engloba toda uma vasta gama de formas de se supor que os constituintes do mundo são feitos de partes menores, e que uma apreensão completa das partes menores não deixaria nenhuma informação de fora sobre as partes maiores. Numa definição mais precisa, e que portanto *está capturando outro conceito, diferente do primeiro*, a posição do eliminativismo mental é associada ao reduccionismo. Segundo os eliminativistas, todo o vocabulário mental se tornará obsoleto conforme fomos aprendendo mais sobre o cérebro, suas configurações, a neuroquímica e a endocrinologia dos comportamentos, e assim por diante. Essa posição é mais famosamente associada a Paul e Patricia Churchland. No sentido abrangente de reduccionismo, no entanto, que é relevante também aos psicólogos, físicos, químicos, biólogos e sociólogos, Dennett é um veemente reducionista. Quando o termo reducionista está sendo usado no outro sentido, no entanto, o nome técnico de sua posição costuma ser alternado entre funcionalismo e verificacionismo. “Funcionalista”, no sentido mais vigorosamente defendido pelo filósofo Hilary Putnam até o início da década de 1990, pode ser pensado como a hipótese de que o que há de comum entre dois estados mentais é sua isofuncionalidade, ou seja, que os dois executem a mesma função. Uma hipótese alternativa seria que quando tem a mesma forma (organização neuronal quase idêntica) eles seriam o mesmo estado, mesmo que tendo em vista o contexto, a função dessas áreas fosse diferente. Para o funcionalista, a função é o que determina o estado mental (isto é, qual dentre todos os possíveis estados mentais aquele estado particular é). A posição funcionalista é consistente com a definição abrangente de reduccionismo, e portanto Dennett é ao mesmo tempo um funcionalista e um fisicalista.

Já o termo também usado para qualificar Dennett, ‘verificacionista’, refere-se a uma posição mais geral a respeito de como se deve fazer ciência e filosofia. Nesse caso, Dennett só considera que algo é um conhecimento em potencial se houver uma maneira testável de que aquela hipótese seja falseada. Entre as diversas hipóteses a respeito da existência de universos paralelos da física recente por exemplo, Dennett apenas interage com, e se interessa por, aquelas nas quais existe alguma maneira de, a partir de nosso universo, fazer inferências que poderiam mostrar a verdade ou falsidade daquela hipótese. Igualmente para uma hipótese sobre o que é a mente e do que ela é feita. Para

que Dennett a considere é necessário que ao menos *em princípio* se possa executar uma manobra que nos garanta a verdade ou falsidade de uma conjectura. Mais fortemente do que a maioria dos verificacionistas, Dennett prefere enfática e claramente teorias e hipóteses que possam ser testadas num curto período e com os recursos já existentes. Verificacionista, funcionalista e reducionista são designações suficientemente próximas da forma de pensar de Dennett para que se apliquem frutiferamente a ele. Quando falarmos mais adiante de padrões, veremos como sua forma de pensar sobre padrões é o que dá liga a essas três interpretações de Dennett. Lembremo-nos que, para o próprio, na medida em que ele pode ser interpretado frutiferamente de tal maneira, então ele possui essas crenças.

2. Informação e processamento

Dennett tem posições claras a respeito de como funciona o processo de conhecer humano, a respeito de como funciona e o que é a cognição.

Ele relembra em sua mini-autobiografia (Dennett, 2008b) que, em seus estudos em Oxford, quando, sugeriu que para entender porque o braço “dorme” deveríamos olhar para a neurociência, ele foi desconsiderado por seus colegas. Mais tarde alguns desses se arrependiam, e com o desenvolvimento das ciências, Dennett ganhou enorme papel não só no mundo da filosofia como também no mundo de biólogos, psicólogos e cientistas da computação. Dizia o espírito intelectual de Oxford nos anos sessenta: ¿Que papel teria a ciência a cumprir na possibilidade de entender o estado de conhecimento (o estado epistêmico) de um agente? Se meu braço está dormindo e portanto está fora do campo da minha consciência, então isso é um problema apenas filosófico, e pode ser resolvido do conforto de nossa poltrona. Dennett, antecipando o atual clima pró-científico e tecnológico, não admitia esse estilo de pensamento. Parece bastante simples que essa questão se possa resolver apenas com recurso à compreensão de como funcionam neurônios, e como eles se comunicam. A mente, ou o que quer que seja, ou está instanciada no corpo, ou suportada pelo corpo, ou é o próprio corpo, de modo que compreender o conhecimento científico sobre o corpo é absolutamente necessário para compreender o que a mente pode ou não acessar, e como ela funciona.

¿Como a mente funciona? A ciência cognitiva, aliada à computação, e mais recentemente à neurociência cognitiva, nos fornecem os melhores modelos. Dennett é considerado também como um cientista, não coincidentemente, justamente nessas áreas. Entender como funcionam computadores é para ele uma tarefa que nos possibilita responder uma das questões essenciais da epistemologia: “¿Como o conhecimento é possível?” No ensaio “Inteligência artificial como filosofia e como psicologia” (Dennett, 1978a), ele trata de conectar a epistemologia com o campo da inteligência artificial, alegando que na inteligência artificial e na psicologia cognitiva, para responder questões *gerais* a respeito de um sistema, uma boa abordagem é projetar um sistema particular que seja capaz de executar as tarefas a serem estudadas, e a partir desse sistema particular questionar-se a respeito de quais dentre as propriedades do sistema seriam requisitos para *qualquer* sistema que fosse executar aquele tipo de ações. Evidente que essa abordagem não esgota as possibilidades metafísicas de como se pode adquirir

conhecimento, mas o que é importante é que a inteligência artificial nos mostra *uma* maneira através da qual é possível compreender como um sistema funcional é capaz de atos complexos, sem necessidade de recorrer a explicações mágicas, contraditórias ou que necessitem postular muitas novas coisas no mundo. A inteligência artificial nos dá um caso particular de epistemologia feito a partir de partes que não são epistêmicas, de conhecimento feito a partir de coisas que não possuem conhecimento: uma forma de transformar algo que não conhece em uma parte de algo que conhece.

Contrastemos a essa abordagem a solução cartesiana: postular um sujeito com capacidade de escolha infinita, um Deus veraz e infinito, e uma distância entre eles, através da qual o sujeito pode caminhar, encontrando peças de conhecimento pelo caminho, é uma solução muito cheia de adornos desnecessários para o problema do conhecimento. Ela não resolve o problema, por ter premissas falsas, e mesmo que resolvesse, ela nos deixa de mãos vazias no que tange a como continuar explorando o problema, como compreender as minúcias que poderiam gerar novos subproblemas. Outra desvantagem da solução cartesiana é que ela é impossível de testar. Temos muitas maneiras de testar se um computador “aprendeu” algo: é só escrever em qualquer que seja sua linguagem a pergunta que consideramos que possui uma resposta “aprendida”, e averiguar se ele responde aquilo que supostamente deveria. O sistema abstrato de pequenas ordens que o computador cumpre é suficiente para lidar com vastas quantidades de conhecimento específico, por exemplo o conhecimento de como se joga xadrez. Então está no nosso poder afirmar categoricamente que sabemos projetar ao menos um tipo de coisa física, material, não misteriosa, que é capaz de “aprender”, de “conhecer” e de “agir”: programas de computador.

“Aprender” e aprender

Dennett repudiaria o uso proposital das aspas que fiz até aqui. Ora, diria, ¿porque “aprender” e não aprender? ¿Qual a substancial diferença entre o conhecimento adquirido pelo computador e aquele adquirido por seu oponente humano? Até onde sabemos, ambos têm igual capacidade de processar informação fisicamente, ambos são sistemas capazes de executar a tarefa notoriamente complexa de jogar xadrez, ¿o que torna o conhecimento do computador mero “conhecimento”? Nada.

É evidente que existem diferenças substanciais entre a forma como um computador processa informação e a forma como nós processamos a mesma informação.

De início, sabemos que o computador executa operações de maneira sequencial, sendo um operador digital, enquanto, de acordo com o conhecimento neurocientífico atual, nós somos um processador misturado, não linear, paralelo e analógico. Dennett pontua que um processador digital pode simular um processador analógico com qualquer grau de acuidade, então isso não nos deveria manter preocupados a respeito de quão *verdadeiro* é o conhecimento do computador. Com efeito, existem diferenças enormes entre o que somos capazes de compreender e pensar e o que um computador é capaz de compreender e pensar, no entanto essa diferença talvez possa ser apenas uma questão de grau. Enfim, essa é a visão que Dennett defende.

Sabemos então que um sistema projetado para conhecer pode ter a cara de um sistema computacional, ou de um sistema de inteligência artificial. Essa é uma das respostas para a questão: “¿Como o conhecimento é possível?” Ela não encerra a questão, de certo, mas é capaz de explicar grande parte da compreensão de regularidades e processamentos cognitivos, com as quais orgulhosamente nos diferenciamos das pedras, ossos, tesouras e demais objetos inanimados. Uma das tarefas mais importantes para a capacidade cognitiva é a capacidade para detecção de regularidades. EsTaMoS TãO AcOsTuMaDoS com detectar regularidades que por vezes nem percebemos como tudo a nossa volta exhibe, de nossa perspectiva, algum tipo de regularidade; é quando essas regularidades são QuEbRaDas QuE CoMeÇaMoS a **PERCebeR** o quanto nosso processo mental depende delas. Um sistema computacional pode simular essa capacidade, através da presença de detectores, e meta-analisadores que veriam as regularidades de ativação destes detectores. Padrões espaciais ou temporais, por exemplo, podem ser percebidos com a ajuda de mecanismos internos, ou regras simples, que afirmam frequências e distribuições num determinado padrão de inputs que seja alimentado a uma máquina.

Retornaremos ainda aos modelos propostos por Dennett de como detectar regularidades no mundo, na verdade de uma maneira muito mais próxima da biologia do que da inteligência artificial. O tema de nosso próximo capítulo será a capacidade de processamento de informação através da biologia, e para isso, como para qualquer atividade em biologia, é necessário convocar a presença de Darwin.

Duas formas de ver uma cara: prosopagnosia

Antes de adentrar a biologia, porém, é interessante pontuarmos algumas das

contribuições mais interessantes de Dennett para a história do pensamento sobre o pensamento, aquelas que dizem respeito não mais à possibilidade do conhecimento, mas ao conteúdo do mesmo. Afinal, ¿o que conhecemos? Muitos filósofos antigos se moviam em um território ingênuo, crendo que o que nós conhecemos, o que vemos, é o que vivenciamos, é o mundo, num sentido totalmente literal. A revolução promovida por Kant, considerado por muitos o primeiro cognitivista, foi perceber que o que vemos não são as coisas como elas são em si, mas as coisas como são para nós. ¿Qual a diferença? Basicamente a diferença é subordinar nosso conhecimento ao sistema cognitivo no qual ele está engendrado. Para Kant, numa opinião mais tarde muito contestada, toda experiência se dá necessariamente no tempo e no espaço, de maneira que o tempo e o espaço estão em nós, e não no mundo. O tempo e o espaço seriam a maneira como nós percebemos o mundo e não propriedades do mesmo. Algumas outras categorias primitivas organizariam nossa experiência do mundo. Assim como experienciamos as coisas *no* tempo e *no* espaço, existiriam outras coisas que organizam nossa experiência do mundo, por serem parte de nós, por serem uma maneira de organizar toda a enorme complexidade que nos rodeia em partes, de modo a poder perceber padrões.

Em muitos aspectos, a teoria de Kant sofre de problemas tão ingênuos quanto as de seus predecessores, mas o espírito que o levou a iniciar esse movimento de como descrever a nossa experiência foi o que possibilitou o surgimento da ciência cognitiva. Sabemos hoje de uma miríade de processos cognitivos através dos quais podemos demonstrar que muitas vezes nossa percepção de um evento não coincide com a existência do evento: esses processos são aqueles que caracterizamos como ilusões. Ilusões são importantes dessa perspectiva porque elas nos mostram uma espécie de defeito do sistema cognitivo. Quando nossos sistemas cognitivos estão trabalhando adequadamente, parecemos perceber o mundo como ele é, mas quando somos iludidos e avisados da ilusão, podemos notar como nosso sistema organiza as informações que chegam até nós, e que elas não são meramente dados absolutos (o que quer que isso signifique). Com suficiente conhecimento a respeito do funcionamento dos neurônios que conectam a retina humana à região posterior do cérebro, é possível inventar uma ilusão de ótica que sabemos, antes mesmo de testá-la, que irá funcionar. Esse tipo de evento cognitivo é de suma importância para a compreensão de como a mente humana processa informação.

Tomemos por exemplo o fenômeno da prosopagnosia, discutido no livro *A consciência explicada* (Dennett, 1991a). A visão de um sujeito se mantém

completamente normal, sendo capaz de diferenciar formas, objetos cores etc. Entretanto, a capacidade de reconhecer faces é perdida, e mesmo os amigos não lhe parecem ser alguém conhecido. Isso evidencia que não é a mesma coisa ver uma cara como um conjunto de formas e cores, e ver uma cara enquanto cara. Sabemos que esse é um problema de natureza neurológica, que o processamento de informação que registra algo como a cara de algum conhecido (no lobo temporal) é uma parte distanciada da parte que reconhece formas, cores e outros padrões (no lobo occipital).

**Não só as representações na mente diferem de nossas intuições,
mas também as regras lógicas**

Um ponto de ataque crucial para Dennett é a crítica de que as proposições (sentenças de linguagem que podem ser verdadeiras ou falsas) tenham de estar na cabeça para que possamos pensar. Em particular, a necessidade que um lógico ou um matemático possa supor de que exista em algum lugar no cérebro de alguém a informação:

A	p.ex.: “Chove”
$A \rightarrow B$	p.ex.: “Se chove, então molha o chão”
<hr/>	
B	p.ex.: “Molha o chão”

Lê-se: A, A implica B, logo B.

¿Será que é necessário que essa regra seja instanciada nesses termos dentro do cérebro de uma pessoa? ¿Será possível a um neurocientista do futuro abrir a cabeça de uma pessoa e com auxílio tecnológico – presumivelmente de nanomáquinas – selecionar exatamente os grupos de neurônios que executam essa operação lógica? Por mais estranho que possa parecer, muitos filósofos e cientistas cognitivos defenderam essa possibilidade. Afinal, qual seria a outra opção? ¿Interferência mágica? Kant e seus sucessores admitiam que o conhecimento não vem de “cima”, mas sim de dentro, de nosso próprio aparato de conhecimento. ¿Como podemos ter avançado em relação a eles e entretanto não admitir que tudo aquilo que sabemos, *principalmente* essas normas lógicas simples, tem que estar instanciado no cérebro de alguém, para que ele possa

executar a operação que elas formalizam?

Há alternativas. Dennett expõe uma delas em “Styles of mental representation” (Dennett, 1983a), capítulo do livro *The intentional stance* (1987, p. 221):

Mas certamente haverá uma representação explícita adicional escondida do usuário? Considere o que acontece quando alguém dá à calculadora o problema: $6 \times 7 = ?$ Suponhamos que a calculadora faça a multiplicação rapidamente adicionando (em notação binária) $7 + 7 + 7 + 7 + 7 + 7$, e durante o processo ela mantenha no seu acumulador ou memória temporária [*buffer*] os totais parciais de cada soma sucessiva. Assim podemos claramente distinguir o processo que ela executa quando multiplica 6×7 do processo pelo qual passa quando multiplica 7×6 . No primeiro caso os resultados parciais são 14, 21, 28, 35, enquanto no segundo caso são 12, 18, 24, 30, 36. Certamente isso é uma representação explícita e sistemática de *números*, mas onde é que a calculadora representa quaisquer *proposições* matemáticas verdadeiras? Sua maquinaria interna está de tal maneira arranjada que ela tem a peculiar propriedade disposicional de *responder questões de aritmética corretamente*. Ela o faz sem jamais *consultar* quaisquer fatos aritméticos ou regras de operação nela armazenados. Ela foi projetada [*designed*], é claro, por engenheiros que sabiam as verdades da aritmética e as regras do cálculo aritmético, e que providenciaram que o dispositivo operasse de modo a “honrar” todas essas verdades e regras.

Essa é uma das maneiras de ter uma representação em seu funcionamento que não precisa de representação explícita. Há diversas razões para se supor que o caso de fato é que nós não temos representações explícitas internas (ao menos não de todas as operações que podemos realizar). Todo o ramo da psicologia evolutiva (Buss, 2005), que estuda a evolução dos comportamentos humanos, está direcionada a tentar compreender os módulos da mente. A ideia original era de que a mente funcionasse como um tipo de canivete suíço, com módulos, ou ferramentas, diferentes para diferentes tipos de atividades. Sabe-se por exemplo que se fizermos a exata mesma operação lógica quando ela nos é apresentada em forma totalmente abstrata ou apresentada como detecção de trapaça, seremos muito mais capazes de detectar a trapaça do que de resolver o problema abstrato (Pinker, 1997). A partir da ideia original, foram se desenvolvendo ramos e se encontrando tipos de especialização mental, tanto em nível comportamental quanto em nível neuronal. Os livros do psicólogo Steven Pinker, que explicam em detalhe aspectos diversos de psicologia evolutiva, neurociência e teoria computacional da mente, se tornaram best-sellers, e ele, juntamente com Dennett e o biólogo Richard Dawkins, formam uma espécie de sociedade defensora da perspectiva genocêntrica da evolução como explicação de tudo o que há por aí, no mundo inteiro. É dessa linha de pensamento que tratará nosso próximo capítulo.

3. Seleção natural

Ao pedir a um grupo de pessoas para explicar de maneira intuitiva a ideia de evolução darwiniana, ou seja, Seleção Natural, imagino que elas se dividiriam entre aquelas que desenhariam uma sequência linear de animais, que progressivamente vão se modificando, e aquelas que desenhariam algum tipo de árvore evolutiva, onde a intersecção entre dois galhos demonstra a existência de um ancestral comum entre quaisquer indivíduos que sejam parte dos galhos.

Ambas essas maneiras de expor a ideia central da seleção natural abandonam grande parte da informação importante.

Esse capítulo tem três tarefas principais. A primeira é demonstrar porque essas formas simples de pensar a evolução não dão conta do nível de intuição que é necessário para raciocinar corretamente sobre o funcionamento da seleção natural. A segunda é passar uma ideia mais completa de alguns aspectos da Seleção Natural no mundo da biologia, que permita ao simulador (você) raciocinar de maneira pouco problemática sobre o tema. A terceira e última é falar sobre a inversão da pirâmide do design (Dennett, 1995a), o que é isso, porque é importante, e porque essa inversão levou Dennett, e outros com ele, à perspectiva de que tudo que exhibe design no universo por um tempo razoável foi criado por seleção natural.

Pois bem, vamos a à primeira tarefa. Imagine o clássico desenho que “caminha” de um macaco quase quadrúpede para uma pessoa andando em postura ereta, ou, nas versões mais atuais, para uma pessoa andando em postura ereta e subsequentemente curvando-se diante da tela de um computador! Enquanto esse desenho privilegia um fator importante sobre a forma de atuação da seleção natural, sua gradualidade, ele passa por cima de uma série de outras propriedades bastante importantes que nossa intuição não é capaz de, por si mesma, computar automaticamente. Talvez a mais importante seja a ideia de que a maioria dos animais (e seres vivos em geral) na história evolutiva são “perdedores” no grande jogo da vida.

Sabemos que o homem e o chimpanzé possuem um ancestral em comum (Mithen, 1996), que provavelmente viveu na África, algo como 6 milhões de anos atrás. Esse ancestral em comum não era igual aos chimpanzés atuais. Na realidade, a seleção natural teve tanto tempo para agir entre o ancestral comum e o chimpanzé quanto para agir entre o ancestral comum e o humano atual. Ou seja, se os chimpanzés tivessem

passado por diferentes ambientes, tipos de sociedade e etc., poderia muito bem ser o caso que nós nos parecêssemos mais com o ancestral comum do que eles, em termos de comportamento, estrutura corporal etc. A seleção natural não pára de agir sobre um grupo de indivíduos quando eles param de se reproduzir com outro grupo de indivíduos; ela continua agindo em ambos os grupos. Isso parece intuitivamente óbvio, e captura nosso conceito de especiação.

Sabemos hoje que espécies não são objetos com fronteiras bem definidas, isto é, existem casos de animais que conseguem se reproduzir com dois grupos diferentes, que não podem se reproduzir entre si. A seleção natural é um processo constantemente operante, e as mudanças que ocorrem a cada geração são mínimas.

Pensando os genes como linguagem

Imagine que você dá a um bebê um jogo de cubinhos com letras em todos os lados, e dispõe os cubinhos no chão de modo a escrever “O melhor bebe do mundo”, sem acento, porque o jogo não possui essa regalia. Agora, o bebê está ali por perto vendo todos aqueles cubinhos que ele não tem a menor ideia que significam alguma coisa, e resolve mexer um pouco neles, tira um cubinho e produz a sentença “O melhor bebe do muno”. “Muno” não é uma palavra com significado, e portanto a sentença não faz mais sentido. Se o bebê realmente estiver modificando os cubinhos aleatoriamente, na enorme maioria dos casos as modificações que ele fizer serão modificações ruins no sentido de que elas dificultam a compreensão da frase, ou a impedem completamente.

Calma lá, alguém poderia objetar: se os fabricantes dos cubinhos forem fabricantes espertos, eles provavelmente colocaram as letras na proporção que elas ocorrem na linguagem, e o bebê vai errar bem menos. De fato, ele vai gerar sentenças erradas bem menos. Mas ‘bem menos’ ainda é uma proporção astronomicamente maior do que a quantidade de vezes que ele criará outra frase com algum sentido. No exemplo acima, com apenas uma modificação (retirar ou colocar uma letra ou espaço) seria possível criar a frase “O melhor bebe o mundo”. Outras possibilidades seriam “O melhor bebe no mundo”, “O melhor beber do mundo”, “O melhor bebe do mudo”, “Melhor bebe do mundo”, “No melhor bebe do mundo”, “Do melhor bebe do mundo”.

Escolhi propositalmente uma frase que aceita uma quantidade desproporcional de modificações simples que ainda têm significado porque isso me permitirá ilustrar o quanto, mesmo nesse caso, é imensamente mais provável criar uma modificação sem

sentido do que uma com (Dawkins, 1986).

Usando um gerador genuíno⁹ de números aleatórios da internet, vou selecionar uma letra para ser retirada ou colocada e produzir 4 variantes aleatórias da sentença:

“O melhor bebSe do mundo”

“O melhoSr bebe do mundo”

“O melhor bebe doE mundo”

“O melhor Sbebe do mundo”

Conforme o esperado, menos de duas das sentenças possui qualquer significado. Na realidade, nenhuma possui. A variação é quase sempre, de uma perspectiva de um falante da linguagem, uma variação ‘ruim’. Algo muito parecido acontece na seleção natural com os genes. Antes de mais nada, é importante lembrar que quando estudiosos de evolução se utilizam da palavra “gene”, eles se referem ao que os biólogos moleculares chamam de “alelo”. Dentro de cada célula do corpo de uma pessoa, lontra ou barata, existe um conjunto de informações escritas numa linguagem bastante incomum. Nessa linguagem a escrita é feita de moléculas chamadas de bases nitrogenadas, que representariam mais ou menos o equivalente às diferentes letras do nosso alfabeto. Um conjunto de letras em “genômês” poderia ser representado da seguinte forma (não necessariamente um conjunto significativo, ou útil, de letras):

ATGACCGAAGGTTAAG

O processo de seleção natural (que será explicado na segunda parte do capítulo) inclui a ideia de variação. No caso da biologia, essa variação é dita cega, isto é, quando algo se modifica (por exemplo adicionando ou retirando uma parte de uma sentença de informação genética), a modificação não é direcionada para nenhum lado especial. Essas modificações são chamadas de mutações.

Uma mutação é cega no mesmo sentido que um bebê brincando com os cubos e construindo uma nova ordem é cego: ele não pré-seleciona um cubo com um objetivo relacionado à sentença. No mundo dos jogos de bebês com letrinhas, não existem grandes problemas em escrever uma sentença errada, modificá-la além da compreensão etc. O mundo da escrita genética não é tão condescendente; muitas vezes, a punição por um erro é a morte, a infertilidade, ou algum defeito que diminua muito a probabilidade

⁹ Mais genuíno que o gerador de números pseudo-aleatório de seu computador pessoal e menos genuíno que um analisador de decoerência quântica, basicamente um número gerado em função de variáveis ambientais sobre as quais não temos controle algum. Ver: www.random.org/

de que aquele erro passe para próximas gerações, impedindo o animal no qual ele ocorreu de se reproduzir tanto quanto os demais de sua geração.

A grande maioria das modificações que ocorreram no ancestral comum do chimpanzé e do homem, até os dias de hoje, foram modificações ruins, alterações no código genético que poderiam impedir o indivíduo de fabricar alguma proteína, de encontrar parceiros, de proteger-se do frio. Mas se essas modificações foram ruins, ¿por que não é o caso que estamos todos extintos? Porque um mínimo dessas modificações são neutras ou boas, isto é, aumentam a probabilidade de que aquela modificação passe para as próximas gerações.

Um dos princípios da seleção natural, então, é que a variação nova é quase sempre uma variação ruim, mas havendo suficientes indivíduos, haverá suficientes tentativas para que alguém se mantenha neutro ou melhore com relação aos outros membros a seu redor. Mutações são geralmente ruins.

Outra ideia importante que não fica evidente nem na linhagem de macacos, nem na árvore evolutiva, mas que é essencial para entender a seleção natural, é o conceito de escassez de recursos, de recursos limitados. Nem toda a energia utilizada por uma geração de uma espécie de seres vivos será usada pela próxima, parte dela escapa para fora da atmosfera, dissolve-se em formas que genes não podem utilizar para replicar genes, ou passa para o patrimônio energético de outra espécie ou outro habitat.

Se a biologia fosse completamente análoga a nosso exemplo dos bebês, todas as espécies, populações, indivíduos, e genes estariam fadados à morte. Afinal a probabilidade de que diminuam em quantidade a cada geração seria muito alta, culminando no término do experimento da vida. Nosso exemplo propõe mutações demais, desorganizando as sentenças numa velocidade maior do que as reorganizando, e também é um árbitro draconiano, já que um só erro impede a sentença de ser uma sentença significativa. Se ao invés de estarmos interessados apenas em frases ou sentenças significantes estivermos interessados em tudo aquilo que pode ser reconhecido como frase ou sentença, estaríamos mais próximos do que a evolução faz, ao permitir que indivíduos com mutações diversas sobrevivam e deixem descendentes, até porque boa parte das mutações se encontra em partes do DNA que não são “lidas” pelos decodificadores genéticos, mensageiros celulares e sintetizadores proteicos.

Uma sequência de “palavras” no genoma precisa codificar alguma coisa

interessante¹⁰ para que haja espaço para ela no futuro. Na analogia com os bebês, o papel do experimentador, que é selecionar as sentenças que são bem formadas, não tem um análogo na natureza, porque, na natureza, “bem formado” não está previamente definido. Sabemos, dirá Dennett (1995a, p. 56-7), que ao longo do tempo podemos *olhar para trás* e concluir que esse ou aquele conjunto de características, codificado por essas “palavras” no genoma, foram as que melhor disputaram os recursos escassos da natureza, e portanto elas eram as que estavam melhor formadas.

A seleção natural é uma história dessas “palavras”, dessas sequências de bases nitrogenadas dentro do DNA e do RNA de seres vivos, se reproduzindo e descobrindo, depois de muito tempo, se eram “gramaticais” ou não. É um mundo em que a gramaticalidade vai se construindo gradualmente, e só pode ser entendida claramente olhando para trás. Evidentemente, a evolução das linguagens humanas também é imprevisível em vários níveis, mas não poderemos explorar esta questão agora.

Temos de manter em mente que em cada galho da árvore evolutiva existem milhares de pequenos ramos de indivíduos que não deixaram descendentes, de genes que não foram passados adiante, e que quando olhamos a árvore só a partir dos sobreviventes de hoje, ignoramos uma parte dos seres vivos que não tem nenhum descendente vivo. E grande parte dos genes dentro desses seres vivos, que também não se reproduziram até os dias de hoje, e não têm nenhuma cópia atualmente ativa, também foram esquecidos.

Isso vale para indivíduos e também para linhagens, linhagens inteiras foram extintas, e espécies inteiras também.

Com uma ideia um pouco mais sofisticada de um diagrama da árvore da evolução, como aquela esboçada pelo próprio Darwin, podemos prosseguir para a segunda parte do capítulo, e tentar entender melhor o que é a seleção natural, e qual a perspectiva de Dennett sobre ela.

A evolução

A evolução, em sua descrição mais abrangente, envolve os seguintes três elementos, como descrito em *A perigosa ideia de Darwin* (1995a, p. 343):

¹⁰ Nem sempre isso significa “interessante” da perspectiva do indivíduo, ou de um biólogo. Pelo que se sabe, o DNA lixo é resultado da capacidade de alguns genes de se anexarem a conjuntos úteis de genes. Então, embora eles próprios não codifiquem nada, estão numa posição estratégica para continuar existindo. Nesse sentido são “interessantes”.

[E]volução ocorre quando as seguintes condições existem:

- (1) variação: existe uma continua abundância de elementos diferentes
- (2) hereditariedade ou replicação: os elementos têm a capacidade de fazer cópias ou réplicas de si mesmos
- (3) aptidão [*fitness*] diferencial: o número de cópias de um elemento que são criadas num certo tempo varia, dependendo das interações entre as propriedades do elemento e as propriedades do ambiente dentro do qual ele persiste.

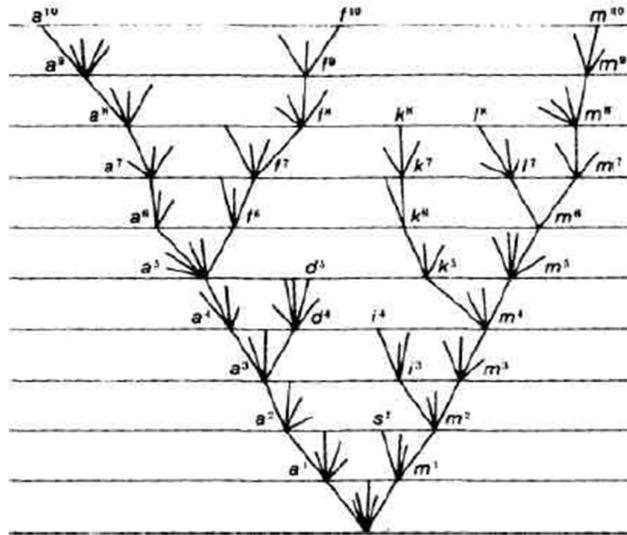


Figura 1: Desenho de Darwin ([1859] 2002) da árvore da vida.

O diferencial entre a análise proposta por Dennett para a evolução e a utilizada normalmente na biologia é que, na presente descrição, fica evidente que não é necessário ser um ser biológico para estar evoluindo. Qualquer sistema, que possua as características acima descritas, irá ser submetido à seleção natural. Por ora, nos manteremos na biologia.

Na nossa biosfera, DNA e RNA são as moléculas que possuem essas propriedades; poderia haver outras, mas são essas que dominam o espaço de replicação do nosso mundo biológico.

Dentre as bilhões de sentenças que já foram escritas em linguagem de DNA, somente algumas continuam existindo hoje e se replicando; essas são aquelas que (cegamente) adotaram alguma estratégia boa para cada época da história da vida, as que se mantiveram suficientemente gramaticais durante toda a história, mesmo com todas as suas mudanças de gramática, para estar se reproduzindo até hoje.

Uma boa estratégia para isso é codificar uma ordem (a ser processada por organelas na célula) de produzir uma proteína útil, ou um olho com mais acuidade visual que outro, num ambiente no qual acuidade visual é essencial. Temos razões para crer que uma águia com boa acuidade visual deixará mais filhotes do que uma sem, e portanto para crer que os genes do código genético de águias que sofreram mutação gerando sentenças que codificam bons olhos deixaram mais cópias do que outras mutações, ou sentenças não mutantes. A Seleção Natural não seleciona bons olhos necessariamente, ela *favorece* bons olhos em ambientes nos quais um olho preciso compensa o gasto energético de ser produzido. Alguns peixes abissais são cegos ou quase cegos, já que o código genético que codifica olhos cegos é *favorecido* nesse ambiente. Em geral, quando alguma característica deixa de ser necessária, sobrevivem os indivíduos nos quais ela tem menos e menos expressão ao longo do tempo, até que desapareça (se possível). Baleias têm restos de patas traseiras, herança de seus ancestrais terrestres, e nós temos um apêndice com funções duvidosas, por exemplo. Ambas são coisas que vinham diminuindo ao longo dos últimos milhões de anos, já que não eram mais úteis.

Outra estratégia interessante é anexar-se a alguma coisa útil. Se um pedaço de DNA que não codifica nada conseguir entrar no código genético em um lugar no meio de dois trechos que codificam algo de importante para a sobrevivência daquele código como um todo, é bem possível que ele passe muitas gerações ali, mesmo não sendo um codificador de nada. Dasoirdukuhl. Imagine um livro no qual uma sequência aleatória de letras foi introduzida numa página qualquer antes da primeira edição ser publicada, e cujo autor não se dispõe a revisar. Se os editores não têm certeza da função da sequência, eles não podem retirá-la subitamente, porque sua obrigação é produzir cópias do livro como um todo. Como o DNA se organiza em grandes conjuntos de genes que ficam anexados (os cromossomos), é possível que sequências de nucleotídeos “adotem essa estratégia de sobrevivência” (para usar uma linguagem intencional imprecisa). Notem a frase do parágrafo anterior:

“Alguns peixes abissais são cegos ou quase cegos, já que o código genético que codifica olhos cegos é *favorecido* nesse ambiente.”

Se é adequado ser cego, ¿para que olhos? Existem duas explicações possíveis para isso. A primeira, e menos provável, é que os genes que codificam a existência de olhos estão anexos a genes que ainda são bastante úteis e portanto estão sendo “hospedados” pelo cromossomo no qual ficam, assim como a corte real seria hospedada

numa visita do rei a algum lugar no qual só o rei seria necessário. Isto seria mais um caso do fenômeno acima descrito.

A outra possibilidade é que não haja como voltar tão atrás através de uma pequena modificação de cada vez. Uma série de mutações que vagarosamente desfuncionaliza o olho pode ir eliminando características irrelevantes no escuro das profundezas, como a curvatura da córnea e a transparência do olho, mas não é apenas porque um conjunto de sentenças se tornou inútil que ele será cortado e jogado fora. As mutações são lentas e em geral pequenas, e não se pode chegar de qualquer lugar em qualquer lugar através dessas mudanças cumulativas. A evolução não é simétrica no tempo, e também não é possível, por uma sequência de pequenos passos, transformar uma zebra numa barata. Devem existir muitos estágios intermediários que ou não são viáveis, ou seja, não levariam o indivíduo a sobreviver, ou são menos viáveis que um estágio intermediário vizinho.

Suponha que você more numa rua muito longa, na qual você só pode ver a casa do seu vizinho da direita e do seu vizinho da esquerda. Todo dia você pode decidir se mudar, caso lhe pareça melhor, para a casa de um dos seus vizinhos. Se a rua possuir 1000 casas, e você por acaso nasceu na 10, é bem possível que você acabe morando em alguma casa entre a 0 e a 50, e muito pouco provável que acabe morando na casa 500. Isso porque há grandes chances de que você ache a casa de um vizinho melhor, e depois veja a de outro que é ainda melhor e você não havia visto antes. E assim por diante. Mas um belo dia haverá uma casa bem ruinzinha ao seu lado, e você não vai querer morar nela, então vai ficar na melhor casa pela qual passou. Como você é uma pessoa, dotada de inteligência, até pode ir um pouco mais longe, mesmo passando por casas ruins, na esperança de encontrar uma mansão em algum lugar, mas a evolução não permite esse tipo de regalia: uma vez que você seja bem melhor que seus vizinhos, é ali que você vai ficar, porque qualquer mutação lhe transformaria em algo que não terá muitas chances de sobreviver na população futura. Uma zebra não pode virar uma barata porque existem vários estágios intermediários que são melhores que os próprios vizinhos. Um caso mais próximo de nossa realidade cotidiana é o apêndice. O apêndice é um órgão que no passado evolutivo recente de nossos ancestrais diminuiu. Não precisamos de um apêndice porque não digerimos mais celulose, no entanto, um apêndice menor que o nosso aumenta a probabilidade de infecções e doenças. Não ter um apêndice seria melhor, mas não podemos dar esse salto. Segue-se que nossa espécie possui esse traço, o apêndice pequeno e desfuncional, que no entanto é estável evolutivamente, sendo

melhor que os traços vizinhos (Bostrom & Sandberg, 2009).

Uma das propriedades essenciais para que a seleção natural funcione é o fato de que a linguagem na qual os genes estão escritos é uma linguagem discreta. Discreta porque ela é composta de espaços bem definidos e objetos bem definidos, que podem ocupar esses espaços. Uma sequência de bases nitrogenadas possui espaços conectados por outras moléculas bem definidas, onde pode estar ou um A, ou um T ou um G ou um C. Não é possível que o mesmo espaço tenha 30% G e 70% A, por exemplo.

No português escrito ocorre o mesmo, não há como escrever uma palavra na qual o espaço de uma letra é preenchido por parcelas de duas letras distintas. Se o discurso escrito em português fosse contínuo (como os sons que emitimos quando falamos português), é plausível supor que nossos cérebros não poderiam decodificar a voz de diferentes pessoas, ou de diferentes ocasiões etc. como instâncias da mesma letra ou sílaba. Mas aceitamos muitos sons diferentes como casos da letra “o” sendo enunciada, por exemplo, tanto em termos de vozes diferentes, quanto volumes diferentes, tonalidades diferentes etc. Se não fizéssemos automaticamente esse enquadramento, ou discretização, a linguagem dificilmente poderia ser expressiva.

A analogia com as bases nitrogenadas, que compõem os genes, fica bastante evidente quando pensamos em reprodução sexuada: se misturássemos os códigos genéticos das duas pessoas, mesmo quando uma delas tivesse uma mutação boa, essa mutação seria gradualmente trazida de volta à normalidade. Isto é, se os genes fossem algo que se mistura, uma gosma que produz uma solução, essa mutação ficaria cada vez mais diluída ao longo das gerações. Ilustremos isso da seguinte maneira.

Considere, por exemplo, uma sequência de dez bases xxxxxxxxxx, onde “x” representa qualquer base existente. Suponha que uma mutação adicionasse um A: xxxxxA(100%)xxxxx. Na hipótese da gosma contínua, os filhos desse indivíduo teriam o código xxxxxA(50%)xxxxx, e os filhos destes xxxxxA(25%)xxxxx, e o efeito da mutação se diluiria e desapareceria.

As mutações que ocorrem na natureza, e a partir das quais opera a evolução, já são bastante pequenas e fazem modificações bastante simples na estrutura geral de um organismo, quando aumentam sua aptidão. *E elas não ocorrem da forma apresentada acima.* Não temos um processo de diluição, mas sim um caso em que 50% dos filhos de um indivíduo tem 100% de chance de ter seu genoma com um “A” presente. A evolução não funcionaria se 100% dos indivíduos filhote tivessem 50% de um gene de seus predecessores. Ela funciona porque é discreta, binária, 0, 1: ou um gene (alelo) está

presente, ou não está. Se houvesse um processo de diluição, teríamos o mesmo resultado que na homeopatia, isto é, ineficiência total e incapacidade de produzir qualquer resultado relevante (salvo por efeito placebo, no caso da homeopatia).

Se trocássemos o processo evolutivo, que é discreto, por um processo analógico, com nuances e gradações contínuas, estaríamos diminuindo a potência de mudança, a longo prazo, da evolução. Se diminuíssemos ainda mais a potência da mudança, provavelmente o ambiente se modificaria mais rápido do que o necessário para garantir que *qualquer* mutação fosse boa. Não por acaso, a natureza das recombinações gênicas é discreta.

Esse tema envolve uma das cláusulas das quais depende a evolução: a cláusula de aptidão diferencial. Se não houver diferença de aptidão, não haverá evolução. O contrário de um sistema discreto é um sistema contínuo, e de fato, em sistemas contínuos, não observamos evolução, salvo na presença de atratores¹¹ Um sistema discreto é um sistema no qual alguma operação ocorre, na base do “tudo ou nada”, e um sistema contínuo é indefinidamente divisível, como um *dégradé*.

Essa é uma das razões pelas quais Dennett adota a perspectiva genocêntrica: *olhando do ponto de vista do gene*, a evolução consegue se desenrolar, enquanto que quando pensamos apenas nas características fenotípicas (principalmente as expressas por vários genes, como cor da pele), parece que vamos cair novamente num *dégradé* que impediria a ocorrência da evolução. A razão pela qual as características externamente visíveis em animais parecem se manifestar em *dégradé* nos filhotes: “Ele tem um pouco do nariz do pai e um pouco do da mãe” é que essas características são determinadas não só por um gene - que por si já é um grande conjunto de bases nitrogenadas em sequência - mas por muitos. Digamos (bastante falsamente, pois genes são muito maiores do que isso) que uma sequência como

AAT AAT AAT AAT AAT

codifique a cor de pele de um indiano escuro, e

GGT GGT GGT GGT GGT

codifique a cor de pele de uma japonesa clara.

¹¹ *Atratores conceituais* ou *cognitivos* são tudo aquilo que faz, com a forma das ideias recebidas e repassadas, aquilo que um ímã faz com o metal. Pensar a Terra como uma esfera é um atrator mais forte do que pensá-la como uma elipse achatada nos polos, e por isso pensamos mais frequentemente na Terra como esfera. A investigação do papel de atratores na evolução cultural tem um bom representante em Henrich, Boyd & Richerson (2008).

Existem muitas combinações possíveis entre essas correntes, para a sequência do filho dos dois. Supondo que se possa cortar a cada 3 letras, poderíamos ter

GGT AAT AAT AAT AAT,

que deve gerar um indivíduo escuro, mas não tão escuro quanto seu pai.

A ideia geral importante aqui, que iremos revisitar em capítulos posteriores, diz respeito à possibilidade de um sistema discreto simular um sistema contínuo. As cores de pele são um sistema que parece contínuo, mas os genes que as determinam são compostos de bases totalmente discretas. Os pixels da tela de um computador são discretos, mas como são menores do que a acuidade visual de muitos humanos, eles parecem contínuos para essas pessoas. Havendo bastante espaço para colocar informação discreta, pode-se simular um sistema contínuo com qualquer grau desejável de acuidade.

Isso vale para sistemas de processamento de informação em geral, e não se restringe a linguagens humanas ou genômicas, programas de computador, estados binários de computação etc. Todos esses são capazes da mesma façanha, se programados para isso, sejam programados pelo homem, pela seleção natural ou até mesmo por outros programas.

Espaço de design, biblioteca de Mendel

Tendo visto como a linguagem sintática das moléculas cria a continuidade aparente entre os seres vivos, mesmo sendo discreta nos níveis que mais interessam à evolução, podemos agora partir para um novo *intuition pump*¹², o “espaço de design” (Dennett, 1995a, cap. 6), e particularmente, o espaço de design da biologia.

O espaço de design deve ser pensado com um espaço no qual se dispõe, de alguma maneira ordenada, todos os designs (formas, desenhos, descrições, mapas) possíveis da coisa que se pretenda estudar no momento. É importante perceber que a ideia de espaço de design independe do objeto cujos designs possíveis estamos espalhando no espaço. É importante perceber isso para que possamos simular melhor o estilo de pensamento de Dennett. Ele se utiliza dessa ideia com relação a uma miríade de situações, problemas e assuntos, e a generalidade do conceito é parte essencial do

¹² *Intuition pumps* são experimentos mentais que auxiliam nossa intuição a pensar sobre um assunto. O termo pode ser traduzido por “condutores de intuição”.

que o faz de suma importância.

No momento, pretendemos estudar biologia, mais especificamente, estamos interessados no tipo de estruturas que se replica no nosso mundo, porque estamos falando de evolução. Ou seja, estamos interessados em analisar o DNA e o RNA.

¿Como poderíamos organizar todos os genomas possíveis? Dennett, seguindo Jorge Luis Borges e Richard Dawkins, sugere que façamos o equivalente a uma organização por ordem alfabética.

Primeiro definimos nossos símbolos válidos. Esse serão “A” “T” “G” “C” e “_” (o espaço vazio serve para representarmos todos os genomas com o mesmo tamanho).

O genoma humano tem aproximadamente 3,2 bilhões de bases nitrogenadas, e o maior genoma conhecido (sim, fomos superados, pelas amebas!) 670 bilhões de bases (Gregory, 2005, p. 73). Começemos pelo começo. Na primeira prateleira de nossa biblioteca ficarão os livros que começam com A, na segunda prateleira, aqueles que começam com a letra C, na terceira os que começam com a letra G, na quarta os que começam com a letra T, e na quinta prateleira estarão os livros que começam com um espaço (como definimos que o genoma sempre começa na primeira letra, nesta prateleira estão livros que não codificam nenhum ser vivo, por definição).

Na primeira coluna estarão os livros cuja segunda letra é A, na segunda coluna aqueles cuja segunda letra é C e assim por diante. Os livros estarão organizados também em profundidade, de modo que na primeira profundidade estarão os livros cuja terceira letra é A, na segunda aqueles cuja terceira letra é C e assim por diante.

Infelizmente acabaram nossas dimensões como bibliotecários tridimensionais, mas embora não possamos imaginar visualmente a continuação desse processo, podemos conceber uma quarta dimensão que organiza a quarta letra e seguir o procedimento até que tenhamos tomado 700 bilhões de dimensões.

Teremos então livros com a descrição do genoma de diferentes seres vivos. Cada livro terá 700 bilhões de espaços, que ficarão vazios após o fim do genoma de cada espécie. Além de todas as espécies atualmente existentes, essa biblioteca contém todos os genomas de espécies que já existiram, que existirão, ou que poderiam existir (ao menos para propósitos teóricos, ignorando um genoma muito maior que o da ameba).

Recapitulando, cada ser vivo (existente ou não) é um livro na biblioteca de Mendel. Os livros estão organizados de modo que aqueles mais parecidos ficam perto uns dos outros, por ordem alfabética. Se olhássemos de muito longe, e se pudéssemos ver 700 bilhões de dimensões, todos os aspectos estariam conjuntamente representados,

e haveria sempre como ver a distância numa dimensão como sendo a distância entre os genomas, de acordo com algum critério particular ligado aos genomas, como, digamos, “o número de A’s no genoma”. Uma questão interessante é se propriedades fenotípicas, como “ter uma cabeça grande”, ou “ter um certo número de dentes”, são mapeadas de maneira próxima no espaço de design dos genomas.

Voltemos à analogia com vizinhanças numa rua, adaptando-a para o espaço de design. A melhor maneira de pensar nas vizinhanças de um animal é vendo seus vizinhos no Espaço de Design, e existem muito mais maneiras de se sair perdendo, isso é, livros vizinhos que são menos gramaticais, do que maneiras de se sair ganhando, quando um ser vivo faz uma mudança.

A evolução pode ser concebida como uma sequência de processos físicos se desenvolvendo no tempo e gerando seres que compõem galhos e diferenciações no espaço de design. A aptidão é um atributo que depende, naturalmente, do meio ambiente no qual o ser vivo está imerso; assim, não é possível determinar a aptidão de um ser vivo apenas analisando sua posição no espaço de design dos genomas. No entanto, dado um ambiente como o da Terra, e um processo evolutivo específico, verifica-se que há certos “atratores” no espaço de design, regiões que codificam características que tendem a ser ocupadas. Voltaremos a este ponto ao tratarmos da “evolução convergente”. Não há porém um atrator absoluto.

O jogo da vida é um jogo de habilidades, mas não é necessário ter nenhuma habilidade particular para que ele comece: disponha algumas estruturas que podem se replicar no espaço de design, e faça-as jogar algumas rodadas. Alguns designs têm maior probabilidade de emergir no espaço de design a partir de um início randômico. ¿Quais seriam esses? Seriam aqueles com maior aptidão, e que seriam muito semelhantes aos “vencedores” anteriores, que por sua vez são semelhantes a seus antecessores, e assim por diante. Conforme os ricos se tornam mais ricos no espaço de design, algumas formas emergem, segundo uma aptidão totalmente mecânica, em oposição a uma aptidão baseada numa noção de sentido ou objetivo; uma definição totalmente física/estrutural/sintática de aptidão, que independe de uma função particular, um sentido, uma teleologia. No meu exemplo da vizinhança das casas, seria como se fosse obrigatório mudar-se quando uma casa ao lado é melhor; isso concentraria a população de rua em algumas áreas, o equivalente aos genomas se concentrando nas proximidades de design, por exemplo, de um cavalo (e a essa população vizinha chamamos de espécie “cavalo”), e outros genomas se concentrando nas proximidades

do design de um jacarandá... Uma vez que sabemos que o processo de decisão é *obrigatório* (porque a evolução pune os perdedores com a morte), podemos imaginar a evolução apenas como bolas deslizando para colinas mais baixas. Não é necessário saber o que é ser “melhor”, só é necessário deslizar para a vizinhança ao lado. É por isso que aqueles seres vivos da sopa primordial se desenvolveram em descendentes complexos como os elefantes, jacarandás e minha prima, mesmo sendo completamente incapazes de processar semanticamente o que confere a um design, ou a outro, o título de melhor.

Em toda a obra de Dennett é importante essa noção da redução de conceitos que eram tidos por essencialmente semânticos em conceitos sintáticos. Isso faz parte do paradigma reducionista por um lado, mas merece uma atenção especial.

Muitos filósofos defendem, ainda nos dias de hoje, que existem aspectos particularmente irredutíveis da mentalidade. Irredutíveis no sentido de que não faria sentido pensar essas propriedades da vida mental como propriedades que simplesmente são manifestação, num alto nível de complexidade, de eventos de baixo nível (campos, forças, partículas etc.). Há duas principais linhas de ataque feitas por esses semanticistas. A primeira envolve a forma como nós, seres humanos, capturamos o significado de palavras, que é considerada um entender “legítimo”, em oposição a um entender “ilegítimo” que seria feito por um computador programado para associar uma palavra a seu uso cotidiano. Enquanto o computador apenas faz operações de maneira mecânica, diz essa classe de filósofos (Searle, 1980; Cole, 2013), de maneira mecânica, e *aparenta* compreender, nós, como *sujeitos*, de fato entendemos o significado das palavras, e essa distinção qualitativa não irá embora mesmo que o computador seja incrivelmente sofisticado.

A segunda linha de ataque é feita por aqueles que argumentam contra o “adaptacionismo” na biologia evolutiva, e pontuam que não se deve pensar semanticamente, em termos de sentido e objetivo, as ações de um ser vivo, já que ele não tem um objetivo porque não pode prever o futuro, e suas adaptações vêm do passado. O pensamento de Dennett está sempre permeado por raciocínios adaptacionistas, e a razão pela qual ele se considera fora da esfera de crítica dos anti-adaptacionistas é o fato de que se utiliza desse vocabulário apenas como *estratégia de previsão*, e não tentando descrever um sentido último de algo.

¿Por que, por exemplo, existem, dentro do espaço de design, tantos animais que evoluíram, separadamente e em momentos distintos, para ter olhos?

Um adaptacionista pode dizer que “olhos são bons para ver, e ver é uma boa estratégia de sobrevivência em vários nichos ecológicos.” Dennett pontua que não há outra boa explicação para que vários espaços semelhantes, mas não vizinhos imediatos no espaço de design, sejam preenchidos, enquanto todos os outros fiquem vazios; a única maneira de incorporar essa informação é aceitando que olhos são uma boa estratégia. São o que ele chama de *movimento forçado* no espaço de design. Ou *bom truque*.

Outro espaço de design, o espaço fenotípico

Consideremos agora o espaço de design organizado de acordo com o fenótipo¹³ dos indivíduos, e não mais o genótipo. Onde antes os hipopótamos estariam próximos das baleias, porque é com elas que mais compartilham DNA, e com elas que possuem o mais próximo ancestral comum, agora eles estarão mais próximos dos porcos, porque o design de seu corpo, e suas vidas, são mais próximas destes. Modificamos nosso espaço de design de um espaço genotípico para um espaço fenotípico.

Num jogo de xadrez entre dois mestres, é muito claro que por vezes, por mais que haja dezenas de movimentos possíveis, somente um é minimamente razoável: esse então é um movimento “forçado”. O mesmo pode ocorrer em termos de evolução, nos espaços de design. Alguns movimentos são simplesmente melhores em muitos nichos, e isso gera o que os biólogos chamam de *evolução convergente*, um caso no qual a árvore da vida (agora visualizada no espaço de design fenotípico) tem vários galhos indo para o mesmo lugar, mesmo que vindo de troncos distantes. Exemplos de evolução convergente são o surgimento da inteligência, de asas, e de uma forma hidrodinâmica em seres aquáticos. Várias plantas possuem a propriedade da fototropia, ou seja, elas tendem a crescer para o lado em que a luz está; é como se um movimento naquela direção fosse um *movimento forçado*. Podemos imaginar o mesmo acontecendo à árvore da vida, que tem vários pontos de convergência na *vasta* biblioteca com a descrição dos fenótipos de todos os seres vivos que já existiram, jamais existirão, ou poderiam ter existido a partir de um genótipo bem definido.

¹³ *Fenótipo* é o conjunto de características observáveis de um organismo, o que inclui sua morfologia, desenvolvimento, propriedades bioquímicas e comportamento. Ele é determinado pela interação do genótipo com o ambiente, e alguns fatores aleatórios envolvendo expressão gênica. O estudo do desenvolvimento e organização dos fenótipos é denominado *epigenética*. O fenótipo também inclui modificações num ambiente, codificadas por genes que alteram o comportamento de modo que os organismos alterem o ambiente. Represas de castores e teias de aranha são apenas alguns exemplos de fenótipos externos ao espaço ocupado pelo próprio organismo.

Minha sugestão pessoal é imaginar dois grandes volumes tridimensionais, como paralelepípedos, e pensar que um deles é o espaço de design genotípico dos seres vivos possíveis, e o outro o espaço de design fenotípico. Dentro de um desses tijolos (transparente), representando o espaço genotípico, cada ponto colorido é um design possível de ser vivo, e cada um pode ser mapeado em um ponto do espaço fenotípico (associado a um ambiente em particular, como o do planeta Terra hoje). Ao longo do tijolo genotípico, a árvore dos seres atuais e seus ancestrais se estende, com uma análoga árvore no espaço fenotípico. Em alguns trechos os galhos fenotípicos convergem aproximadamente para certos pontos; os análogos genotípicos podem ou não convergir. Esses seriam exemplos de movimentos forçados que a evolução faz. Esses pontos, os *bons truques*, são atratores, imãs no espaço de design.

A informação codificada no genoma é sintática, mecânica, ela só começa a adquirir significado no processo de tradução e decodificação feito pelas proteínas. O que chamamos de “significado” está todo surgindo de um mundo de moléculas mecânicas, sem nenhuma capacidade de nada antever, que simplesmente vão existindo conforme permitem suas estabilidades químicas, e que, devido às condições restritivas do jogo da vida, estão, “coincidentalmente”, cercadas de organelas, células, indivíduos e até grupos (Wilson & Wilson, 2007) que existem para preservá-las. Se genes pudessem pensar, *e eles não podem* em nenhum grau, eles provavelmente se surpreenderiam com a absoluta sorte que têm de estar cercados de todo esse aparato que os preserva. De fato, é importante reforçar que por mais que genes apenas existam, e não pensem em nada, existem razões por todos os lados para explicar porque algo vivo é como é. Um ponto importante da filosofia de Dennett é que não é necessário ninguém apreciando um *movimento forçado* para que o movimento tome as rédeas da evolução e faça com que um ser vivo faça algo.

Uma espécie de pássaro é conhecida por “fingir” que a asa está quebrada quando um predador olha para seus ovos: o predador, vendo uma presa grande e fácil, corre atrás do pássaro, que por sua vez voa para longe, já que ele não está de fato machucado. O pássaro não precisa estar consciente de que isso é uma estratégia para salvar seus filhotes, ele não precisa fingir, nada disso precisa acontecer numa esfera de pensamento aviário. A única coisa que é necessária é que uma sequência de bases nitrogenadas codifique um cérebro que, ao detectar um predador olhando para seus ovos, manda sinais elétricos correspondentes ao que mandaria se estivesse machucado. Ninguém precisa saber de nada, é só mecanismo, um mecanismo extremamente útil para preservar

a estabilidade de moléculas, mas apenas um mecanismo.

Improbabilidades no jogo da vida

Mas se isso é um mecanismo improvável, e as mutações são quase sempre ruins, ¿como pode isso ter evoluído? Vale a pena resumir o argumento de Dennett (1995a, p. 54) sobre a existência da improbabilidade. As chances de vencer 10 cara ou coroa seguidas são bastante baixas. Exatamente 1 em 1024. No entanto, se fosse extremamente necessário que alguém vencesse 10 cara ou coroas (do contrário aliens invadirão a terra), há uma maneira muito simples de fazer isso: organizando um torneio. Junte 1024 pessoas, e faça 512 lançamentos de moeda, sendo que cada um é obrigado a apostar no lado contrário de seu par. Junte os pares vencedores, e ao cabo de 10 rodadas haverá 1 vencedor de 10 rodadas, o campeão do torneio. Não faz sentido dizermos que é muita sorte que *alguém, quem quer que seja*, venceu 10 cara ou coroas seguidas, porque o torneio foi organizado de modo que isso iria acontecer. A probabilidade de uma pessoa vencer 10 cara ou coroas seguidas é 1 em 1024, mas a probabilidade de que nesse campeonato *alguém* o fizesse é 1 em 1, 100%.

O jogo da vida não é um campeonato perfeitamente organizado para que se chegue a algum resultado particular, mas há vários espaços de aptidão alta esperando para serem colonizados por aqueles que conseguirem se avizinhar a eles no vasto território do espaço de design. E mesmo que isso exija bastante sorte, só é necessário que uns poucos indivíduos (entre os trilhões existentes) tenham essa sorte para que ela seja preservada e aquele espaço seja preenchido, aquela mansão seja habitada, aquele *movimento forçado* comece a acontecer. Não podemos nos esquecer que os ganhadores dos cara ou coroa do jogo da vida não apenas vão para a próxima rodada, como se duplicam, triplicam, até centuplicam antes de ir para a próxima rodada. A replicabilidade é uma estratégia fantástica para produzir 1024 ganhadores de cara ou coroa a partir de 1024 competidores iniciais: é só permitir que eles tenham filhos que herdem seus prêmios sempre que vençam.

¿O que é então a biologia? Biologia é engenharia, é a engenharia de sistemas capazes de fazer o que quer que precise ser feito para que um grupo de bases nitrogenadas esteja por aí, dentro de algumas gerações. Analisar um ser vivo é algo que requer uma perspectiva de engenharia reversa, e *quase sempre* a pergunta “¿Se isso não parece fazer nada, porque está aí?” terá uma resposta interessante dessa perspectiva.

Pensar a biologia dentro do espaço de design como engenharia nos leva a perguntas interessantes, que é o que uma ciência deve fazer. É quase sempre o caso que existe uma resposta para a pergunta “;Para que evoluiu esse traço?”, porque quase todos os traços biológicos são determinados em grande parte por código genético e requerem gasto de energia. Qualquer coisa que requer processamento de informação e gasto de energia deve estar ou (1) dando algo em troca (por exemplo, olhos de águia) ou (2) parasitando (por exemplo olhos de peixe abissal). Os olhos de uma águia ajudam-na, com seus dois pontos de foco, a caçar, se locomover, e analisar seu ambiente. Os olhos de um peixe abissal, por outro lado, só fazem com que o peixe gaste energia a toa para criá-los, fazer o metabolismo de suas células, transportá-los etc., e portanto podem ser considerados parasitas do peixe.

Para funcionar, o olhar adaptacionista sobre a natureza dos seres vivos não precisa estar contaminado de noções como o lamarckismo (herança de caracteres adquiridos), nem semântica (conhecimento prévio do significado de uma mudança evolutiva), nem finalismo. E o vocabulário adaptacionista é compatível com a visão contemporânea de um mundo mecânico, sendo uma boa maneira de fazer previsões em certos níveis de análise. Essa, em suma, é a posição de Dennett sobre o adaptacionismo em geral, e em particular sobre a utilidade de se empregar o vocabulário adaptacionista.

Sem dúvida a ideia de Dennett mais citada, no que diz respeito à evolução, é o fato de que ele a pensa como um algoritmo. Algoritmos são um território vasto e interessante, que está principalmente presente nas ciências da computação, e portanto, a união de nossa simulação de Dennett com o conceito de algoritmo, e subsequentemente com a evolução, terá de esperar o próximo capítulo, que trata mais de computação.

4. Máquinas de Turing

Vimos no capítulo 2 um esboço de uma das diferenças que se pode atribuir entre o funcionamento de homens e máquinas. Uma calculadora pode fazer suas operações sem necessariamente estar instanciando explicitamente alguma regra da lógica em seu conteúdo interno. Nós fazemos contas de um jeito diferente do da calculadora, mas o fato de que ela pode funcionar de diversas maneiras é suficiente para mostrar que existem muitas maneiras de *parecer* estar engendrado num sistema de regras, sem necessariamente estar engendrado nele.

Na p. 32 apresentamos uma citação de Dennett (1983a) que ressalta que a mesma operação matemática pode ser computada de diferentes maneiras, ou seja, pode ser organizada de diferentes maneiras em suas descrições de nível mais baixo. No nível abstrato mais alto, $6 \times 7 = 7 \times 6$, mas descendo um pouco o grau de abstração, encontraremos diferenças, conforme destacado por Dennett. Descendo ainda mais, para o nível quase físico da operação de computação, haverá mais de uma maneira de computar 6×7 , com sub-passos internos diversos, assim como conseguimos observar a diferença de nível intermediário entre 6×7 e 7×6 .

Em um de seus mais antigos textos, “As habilidades de homens e máquinas” (1978e), apresentado como comunicação em 1970 e publicado no livro *Brainstorms*, Dennett se volta para traçar as distinções entre as capacidades humanas e as capacidades de máquinas. A filosofia da mente do séc. XX abalou-se com uma descoberta importantíssima da matemática, uma das poucas descobertas feitas pelo homem até hoje a respeito do que *não* podemos saber. Trata-se dos teoremas da incompletude de Gödel. Antes mesmo de seu futuro amigo Douglas Hofstadter escrever o épico livro *Gödel, Escher, Bach* (1979), Dennett já havia exposto suas ideias a respeito desse resultado. O teorema de Gödel é demasiado complexo para ser explicado aqui, mas em poucas palavras ele diz que, dado um sistema axiomático de lógica, S, há enunciados expressos na linguagem de S que não podem ser deduzidos dos axiomas de S, e nem as suas negações podem ser deduzidas. Considerando que o sistema axiomático busca axiomatizar uma teoria como a aritmética (construída na teoria dos conjuntos), pode-se concluir que há verdades da teoria aritmética (expressos na linguagem de S) que não são dedutíveis no sistema S.

Um resultado próximo a este é a prova de indecidibilidade derivada por Alan

Turing, mostrando que não há um procedimento efetivo para determinar, para qualquer programa de computador, se ele irá parar ou não. Podemos descobrir se um certo programa irá parar (gerando o resultado de uma função, para um dado argumento) ou não (entrando em um loop infinito), mas não há um procedimento que valha para *qualquer* programa de computador.

A demonstração se baseia na definição de um computador abstrato simples, que veio a ser conhecido como *máquinas de Turing* (ver Turing, [1948] 2004). Todos os computadores deterministas contemporâneos fazem operações que, em última análise, podem ser simuladas por uma máquina de Turing. Essas máquinas de Turing, no entanto, não são máquinas de verdade, pois têm uma memória infinita, sendo assim uma suposição abstrata de uma máquina computacional. Turing nos convida a imaginar uma fita infinita dividida em casinhas ou espaços discretos, cada qual preenchido por um símbolo, que podemos imaginar como sendo dois: uma bola preta ou um espaço vazio. Além disso, a máquina pode se encontrar em um número finito de diferentes estados, que podemos numerar de 1 a n , sendo que o primeiro é o estado inicial da máquina. A fita é lida por uma leitora, uma casinha por vez, e cinco operações podem ser feitas pela leitora: escrever uma bola preta, escrever um espaço branco, dar um passo para a direita, dar um passo para a esquerda, ou ficar na mesma casinha.

A função sendo computada, para a entrada de dados que está inicialmente na fita, é representada por uma tabela de ações, que contém comandos do seguinte tipo: “Se estiver no estado 3 e lendo uma bola preta, mova a fita uma casa para a direita (ou, alternativamente, escreva um espaço branco) e vá para o estado 4”. Para que a máquina seja capaz de computar é necessário que haja pelo menos um comando que diga algo como “Se estiver no estado n e encontrar um espaço x , pare a computação”. O próprio cessar da ação indica que a computação foi encerrada, para um intérprete externo. Este pode então olhar para a fita, com todas as suas bolas pretas e espaços vazios, e traduzir essa linguagem simbólica no resultado da função calculada, para o argumento de entrada. O programa poderia ser a função soma, com dois argumentos, 2 e 3, e o resultado seria 5. O código usado para representar os números é arbitrário, mas no exemplo dado poderia ser baseado no código binário, com a bolinha representando o dígito 1 e o espaço em branco o zero. Uma vez decodificada a mensagem, saberíamos o que estaria acontecendo.

¿Essa máquina é uma abstração matemática só porque ela possui uma fita infinita? Não só isso, mas também porque pode ser implementada finitamente em

diversos sistemas físicos diferentes, um deles sendo o seu computador. As informações que você digita no teclado ou ao movimentar o mouse são informações que chegam codificadas no computador, como *input* ou entrada. Conforme o programa que está sendo rodado, gera-se um *output* ou saída, como o que você vê na tela, os sons que ouve etc. Os programas usados no computador geram resultados de maneira muito mais eficiente do que uma máquina de Turing (imagine quanto tempo uma fita teria de ir e voltar para fazer operações simples como apagar uma imagem num editor de desenho).

A implementação do programa de uma máquina de Turing também não precisa ser feita em coisas que reconhecemos como máquinas. Suponhamos por exemplo que o governo chinês resolva pedir ao seu exército de dez milhões de pessoas que implementem uma máquina de Turing (adaptação do intuition pump “*China brain*”, de Block, 1980). Cada soldado poderia “pegar um balde com água e jogar na cabeça”, encher o balde com água, ou passar o balde para seu colega à esquerda ou à direita. Com uma série de instruções simples, conseguiríamos criar uma força computacional capaz de implementar uma máquina de Turing (restrita a uma fita finita). Evidentemente, nesse caso, como nos outros, tem de haver dispositivos que traduzam a informação numa forma na qual possamos interpretá-la nós mesmos. E é isso que mais interessa a Dennett.

Suponhamos que encontramos uma máquina na praia, e descobrimos que ela implementa perfeitamente a operação de jogar damas, contanto que interpretemos toda informação que entra num ponto de acordo com uma linguagem simbólica e toda informação que sai como uma outra. Isto é, supondo um código possível de tradução entre estados dessa máquina e descrições em forma binária de um estado de jogo de damas, percebemos que a máquina funciona como um jogador de damas. E que belo jogador de damas! Agora suponhamos que, fascinados com a descoberta e querendo descobrir sua origem, vamos a um simpósio de descobertas e encontramos outro grupo de afortunados que encontraram exatamente a mesma máquina. Mas esse outro grupo alega que essa máquina não deve ser lida da maneira como estamos lendo, alega que devemos contar o número de buracos que aparecem na entrada e na saída, e não o número de buracos somado às bolinhas, e que na verdade a máquina serve para avaliar a simetria interna da entrada de bolinhas. As duas máquinas são fisicamente idênticas, por dentro e por fora, e os dois grupos ficam batalhando a respeito de qual a real função dessa máquina. ¿O que estaria essa máquina de Turing, de fato, fazendo? Dennett nos explica que em verdade há mais de uma máquina de Turing sendo implementada aí. No

caso temos uma máquina de Turing que admite uma interpretação interessante como jogador de damas, e portanto ela é legitimamente uma jogadora de damas, e também há outra máquina que é legitimamente uma avaliadora de simetrias. Em verdade há infinitas máquinas de Turing ali sendo instanciadas, conforme interpretamos seus símbolos, mas duas delas estão fazendo-o de tal maneira que consideramos a informação que elas processam importante, e principalmente, sabemos traduzi-la. Ora, olhando-se dessa maneira, qualquer objeto material pode ser interpretado como uma Máquina de Turing ao interagir com seu ambiente¹⁶, mas na maioria dos casos, não teremos elementos sintáticos isolados (como bolinhas e buracos) e também não teremos uma linguagem para traduzir (como código Morse ou Java). Ou seja, não haverá nenhuma razão interessante para dizer que aquilo está implementando uma máquina de Turing. Só uma perspectiva muito enviesada, cujo objetivo é torcer a realidade de diferentes maneiras, será capaz de gerar interpretações esdrúxulas de objetos cotidianos como as tais máquinas. A interpretação de pessoas como máquinas de Turing, no entanto, não envolve malabarismos como este, e essa forma de ver, a chamada teoria computacional da mente, permeia cada questão abordada por Dennett.

Temos uma questão pendente que agora pode ser respondida. Dennett (1978e, p. 256) a formulou da seguinte maneira no início de seu artigo:

O esqueleto comum dos argumentos anti-mecanicistas é o seguinte: qualquer máquina computacional pode ser representada como alguma máquina de Turing, mas um homem não pode; pois suponha que Jones, que ali está, seja uma realização de alguma máquina de Turing TM_j, então (por Gödel) haveria algo *A* que Jones *não poderia fazer* (qual seja, provar a sentença de Gödel de TM_j). Mas cuidado! – e esta é a parte empírica crucial do argumento – Jones consegue fazer *A*, portanto Jones não é uma realização de TM_j, e como *pode ser visto* que isso será verdadeiro qualquer que seja a máquina de Turing que escolhamos, Jones transcenderia, como um anjo, os limites do mecanismo.

Em outras palavras: se Gödel provou que uma sentença não pode ser provada por um certo sistema, mas nós, seres humanos podemos conhecer essa verdade, ¿como assumir que nós somos um tal sistema? Com o que vimos anteriormente, temos

¹⁶ A discussão filosófica sobre a multiplicidade de interpretações de objetos como máquinas abstratas se iniciou no apêndice de um livro do principal precursor funcionalista de Dennett, Hilary Putnam (1988), e seguiu viagem por baixo do pano até ser trazida à tona por seu principal algoz, David Chalmers (1996). Ela então ganhou força fora da filosofia, com a publicação da ficção científica de Greg Egan, *Permutation city* (2010), que dá margem à entrada de questões éticas, como a seguinte: ¿se uma pedra é computacionalmente equivalente a Gandhi, quando quebro uma pedra estou fazendo algo eticamente análogo a matar o Gandhi? Uma discussão informal sobre o tema pode ser encontrada em lesswrong.com/lw/1jm/getting_over_dust_theory/

elementos para responder tal questão pendente.

Primeiro, nós não precisamos ser o mesmo tipo de sistema que os tipos de sistema sintático de regras nos quais o teorema de Gödel é válido (o sistema de lógica da aritmética, por exemplo). A resposta é que nós somos *várias* máquinas de Turing, e não apenas uma. Mas talvez essa não seja a interpretação mais interessante de nosso comportamento e de nossos atos. Tal qual a máquina da praia, que admite mais de uma interpretação coerente e razoável de si, nós também admitimos. Não só no sentido de que podemos ser interpretados como fazendo duas operações diferentes ao mesmo tempo, mas também porque ao longo do tempo nos modificamos o suficiente para que não estejamos mais implementando o mesmo conjunto de regras simples com nosso aparato mental. Essas modificações se dão num nível mais alto de complexidade que o nível de complexidade dos estados de fita de uma máquina de Turing abstrata convencional. Somos então uma série de grupos de máquinas de Turing.

Mas ainda que possamos alegar isso num nível teórico, isso apresenta tantos problemas pragmáticos quanto achar que compreenderemos o funcionamento de uma bicicleta utilizando descrições de nível molecular e quântico. Talvez a linguagem através da qual nos comunicamos não seja uma linguagem baseada em sinais específicos como bolinhas e buracos, e dependa de diversos fatores e seja sensível a diversos fatores pragmáticos, desde a temperatura ambiente até o zunido de um mosquito. Então, interpretar a nós mesmos como uma máquina de Turing seria um trabalho que não compensa, e para o qual não existe uma linguagem pré-definida e um conjunto de símbolos que possibilitasse essa compreensão.

Cabe agora uma introdução à próxima seção, e cito um trecho de *Intuition pumps and other tools for thinking*, o livro mais recente de Dennett (2013, p. 29), que em geral não será discutido por nós:

O pé-de-cabra da investigação racional, a grande alavanca que garante a consistência, é o *reductio ad absurdum* – literalmente, reduzir (o argumento) ao absurdo. Você pega a asserção ou conjectura que está em pauta e vê se consegue extrair dela quaisquer contradições (ou apenas implicações disparatadas). Se conseguir, essa proposição tem que ser descartada ou enviada de volta para reconfecção. Nós fazemos isso o tempo todo sem nos preocuparmos em salientar a lógica subjacente: “Se isso é um urso, então ursos têm chifres!” ou “Ele não chegará a tempo para o jantar, a não ser que consiga voar como o Super-homem”. Quando a questão é uma controvérsia teórica intrincada, o pé-de-cabra é usado energeticamente, mas aqui fica difícil traçar a distinção entre crítica justa e refutação por caricatura. Será que seu oponente é realmente tão idiota a ponto de acreditar na proposição que você acabou de reduzir ao absurdo com alguns movimentos hábeis? Certa vez corriji um trabalho de aluno que tinha um curioso

erro de escrita, trocando “paridade” [*parity*] por “paródia” [*parody*], criando a deliciosa frase “por paródia de raciocínio”, um nome apropriado, creio, para argumentos ilegítimos de *reductio ad absurdum*, que são muito frequentes no calor da hora da controvérsia filosófica e científica.

***Reductio ad absurdum* usando Gödel**

Dizíamos anteriormente que talvez seja melhor concluir que nos pensar como Máquinas de Turing é um trabalho que não compensa. Mesmo assim, diriam os mais perspicazes, se em última análise formos de fato uma forma finita de Máquina de Turing, mesmo que imensamente complexa e intraduzível, nunca conseguiríamos conhecer aquela verdade matemática (que a sentença de Gödel afirma ser incognoscível dentro do sistema). Mas, de fato, conseguimos, então não somos uma máquina de Turing. Esse é um argumento por *reductio ad absurdum*.

Alto lá! - diz o dennettiano, não é bem assim. De fato podemos ser interpretados como máquinas de Turing extremamente complexas, e de fato sabemos que uma máquina de Turing não pode provar sua sentença de Gödel. Mas nós instanciamos diferentes máquinas de Turing ao longo do tempo, e não estamos confinados a um sistema particular específico de símbolos, elementos e estados, como um computador pessoal. Nosso sistema físico não é desenvolvido para estar isolado do ambiente ao seu redor e processar apenas seus inputs e outputs; pelo contrário, ele é sensível a uma miríade de mudanças do ambiente externo e reage a essas mudanças das mais diversas maneiras. Assim sendo, podemos ser uma máquina de Turing (numa análise interessante) num momento, e outra em outro. Logo, o fato de que podemos provar ou encontrar essa ou aquela verdade matemática em nada desafia as provas de Gödel, e não são nenhuma garantia, ou sequer evidência, que a mente não é um processo mecânico, sintático e causal.

Sintático como uma hidroelétrica, semântico como Shakespeare

O uso de ‘sintático’ em filosofia é bastante específico, e cabe aqui uma rápida explicação (posteriormente virão outras). Um sistema *semântico* é aquele que diz respeito ao significado das palavras ou símbolos. Então, por exemplo, o que conecta a palavra ‘cão’ com os cães é algo relativo à semântica. *Sintático* é o aspecto puramente

formal dos símbolos, por exemplo, as posições em que eles podem se encontrar numa determinada sentença, para que esta faça sentido. Além disso, os sistemas mecânicos podem ser vistos como sistemas sintáticos, isso é, sistemas de processamento formal de dados, sistemas que processam dados sem levar em conta o significado deles, mas que apenas são influenciados pela forma do sistema. Se eu estivesse andando por Tóquio e visse uma placa em japonês, por exemplo, eu só seria influenciado de forma sintática por ela, mas se ao lado dela houvesse uma placa escrita em português, o conteúdo semântico dela poderia ser interpretado por mim e eu seria também influenciado pelo aspecto semântico da frase. Se for verdade que o cérebro é apenas mecânico, o que é bastante defendido e condizente com todas as ciências, então ele é apenas um sistema sintático. E isso é algo de fundamental importância para filósofos da mente como Dennett. Afinal, digamos que você queira defender a hipótese de que a mente é só o cérebro, ou um conjunto de funções realizadas pelo mesmo. Uma das afrontas que eu lhe poderia fazer é pedir que me explicasse como essa máquina sintática é capaz de processar informação semântica. E essa é uma das charadas mais difíceis de resolver. Com efeito, o primeiro a expor de maneira clara esse problema foi o filósofo John Searle, que propôs o experimento do quarto chinês, sobre o qual falaremos mais adiante. O ponto do experimento é demonstrar que uma máquina não é capaz de “compreender” no mesmo sentido que nós “compreendemos”, e isso se dá fundamentalmente porque nós somos sujeitos semânticos enquanto as máquinas são apenas sintáticas. Veremos a solução de Dennett para esse problema mais adiante.

Prove que você é uma pessoa

Turing era definitivamente um sujeito à frente de seu tempo, e além de desenvolver sua máquina (que como vimos é apenas uma abstração que pode ser finitamente implementada numa máquina real), ele também percebeu que no futuro será possível que os computadores cheguem a ter capacidades similares às capacidades humanas. Hoje isso pode ser considerado quase trivial por alguns, e totalmente absurdo por outros, mas nos anos cinquenta pode-se dizer que era absurdo para quase todo mundo. Turing desenvolveu outra abstração, o teste de Turing, para verificar quando uma máquina já teria adquirido suas condições de personalidade.

O teste de Turing consiste em fazer uma máquina conversar com um agente humano, de tal forma que este não perceba que está conversando com uma máquina.

Uma das maneiras de se saber isso seria no caso de o agente não conseguir diferenciar entre a máquina e outro humano. No livro *Gödel, Escher, Bach*, Hofstadter (1979) sugere várias concepções diferentes do teste de Turing e os discute mais longamente.

A linguagem humana é bastante vasta, e por isso o repertório de conhecimento necessário para usar a linguagem adequadamente é bastante grande, e a capacidade de processá-lo de maneira adequada também. Dennett nos convida, no capítulo “Can machines think” do livro *Brainchildren* (1998), a considerar um exemplo de outro filósofo:

“O comitê negou ao grupo uma permissão de fazer uma parada porque eles advogavam a violência”.

E a segunda sentença:

“O comitê negou ao grupo uma permissão de fazer uma parada porque eles temiam a violência”.

Nessas sentenças o pronome “eles” é ambíguo, e podemos imaginar um mundo exótico em que de fato seja o caso que um comitê do governo seja defensor da violência e de alguma maneira isso os faça não permitir paradas em geral (quem sabe porque a parada fosse ocorrer justamente no lugar onde planejavam um massacre). Mas não é essa a interpretação que fazemos – pelo menos fora do Brasil em 2013 não é essa. A interpretação humana normal é que o grupo advoga a violência e o comitê não gosta disso. Mas isso exige não apenas conhecimento de normas sintáticas, de normas gramaticais, mas também é necessário conhecimento sobre o mundo. E para nós, isso é tão óbvio que é preciso que um filósofo nos traga à atenção a existência da esdrúxula interpretação para que percebamos o processo que nos ocorre inconscientemente. É por isso que o teste de Turing é um teste difícil, porque a máquina teria que perceber esse tipo de nuance, do contrário não poderia enganar seu interlocutor. O teste fica mais fácil quanto mais restrito for o tópico, quanto menor a linguagem e quanto menos se possa falar. Restringi-lo à linguagem é um preço a se pagar (pode ser, por exemplo, que uma máquina domine as habilidades de linguagem humana e não as mecânicas, ou as intuitivas de percepção e de movimentação de objetos). ¿Para que serve esse nosso exemplo ilustrativo? Para lembrar-nos que em muitos casos o uso de linguagem está repleto de ambiguidades, e que a interpretação dessas ambiguidades exige conhecimento do mundo, e portanto uma máquina que conheça o mundo ou terá de haver passado por um processo de aprendizado magnífico de “senso comum” dentro de

uma sala fechada, recebendo trilhões de informações, ou haverá de ter interagido um pouco com o mundo de modo a obter uma série de seus conhecimentos. Ao menos assim considera Dennett.

Pensa ele também que esse teste é demasiado difícil, e que as versões dele nas quais já houve confusão dos juízes entre homens e máquinas estavam demasiadamente restritas (em termos de tópicos e linguagem permitida). Com efeito, como Dennett aponta em diversos textos, uma das principais habilidades que diferencia os humanos é a capacidade de lidar com uma informação nova de um tipo exótico, uma chamada “pane” no sistema que faria com que uma máquina girasse em círculos, voltando ao mesmo lugar redundantemente de novo girando em falso, e se perdesse, por assim dizer, do seu meio representativo. Quando colocamos um elemento novo no nosso ambiente, geralmente somos capazes de lidar com ele, mas quando perguntamos a um computador programado para jogar xadrez “¿Um cavalo se parece mais com um bispo ou com um peão?”, não esperamos nenhum tipo de resposta convincente, ao menos não da forma como os computadores são hoje. Dennett conclui, em 1997, que os testes de Turing ainda são muito difíceis para o mundo real. Falta tempo para que tenhamos a primeira máquina capaz de vencer esse desafio.

Algoritmos desenhados por nós e algoritmos desenhados pela evolução

Vimos até aqui que somos máquinas sintáticas, que portanto podemos ser interpretados como máquinas de Turing, que fazê-lo não vale nada a pena e não faz sentido, e que os computadores até hoje ainda não parecem estar habilitados a emular uma performance humana em tarefas complexas, como uso e compreensão de linguagem. O teste de Turing parece muito difícil para as máquinas, em particular por sua incapacidade de lidar com modificações do mundo que dizem respeito a coisas não previstas em seus sistemas de representação e simbolização.

Veremos no próximo capítulo que não são apenas as máquinas que podem se enganar. Por ora, façamos o link prometido entre o universo das computações (sequências de pequenas ações discretas e simples, automatizáveis) e o universo da evolução.

Um algoritmo é um conjunto de regras simples que produz um resultado interessante (na verdade é só um conjunto de regras que produz um resultado, mas dificilmente você verá alguém citando um algoritmo que não produz nada de

interessante, e portanto talvez valha mais a pena associar a palavra à ideia de um conjunto simples de regras que gera um output interessante).

Algoritmos podem ser programados em diversos tipos de linguagem, eles podem ser escritos em notação lógica, em linguagens de programação, e em qualquer outra notação que possa ser detectada por algum sistema capaz de realizar as operações triviais que o sistema determina. Se seu computador parece muito inteligente, isso se deve a uma quantidade grande de linguagens de programação amontoadas uma em cima da outra, sendo implementadas muito, muito rapidamente pelo processador, e já tendo um idioma de transcrição adequado para algo que você considera intuitivo (por exemplo, dispor na tela letras, fotos ou músicas). Se a evolução parece muito inteligente por ter gerado algo tão incrível quanto as aves do paraíso, os tatus bolas e os orangotangos, isso se deve ao fato de que aquelas três condições necessárias (vistas no capítulo 3) estão a muito, muito, muito mais tempo operando do que nós conseguimos imaginar com nossos cérebros preparados para antever apenas algumas décadas. Revisitemos as três condições da evolução por seleção natural: (1) variação; (2) hereditariedade ou replicação; (3) aptidão (*fitness*) diferencial.

Isso não parece exatamente um conjunto de ordens, um conjunto de operações simples que um idiota poderia executar, um processo computável. ¿Mas será que há ordens, pequenas instruções de ação implícitas em algum lugar?

Não. O que existe aí é uma correlação com um estado no qual um algoritmo está sendo implementado. As leis da física determinam a movimentação de campos, forças, partículas, moléculas etc. Uma vez que os primeiros replicadores hajam surgido, um subgrupo mínimo das coisas físicas, esses replicadores, passam a seguir a seguinte instrução, por conta de seu coincidente design: “Faça cópias de si mesmo”.

Uma vez que esse sub-nicho da física é criado, e a possibilidade de outros elementos que se repliquem ainda mais rápido está em aberto, porque o sistema é aberto, temos todos os elementos necessários para implementar o algoritmo darwiniano. Como os replicadores estão num sistema aberto, que por ora interfere naquela primeira instrução, acabamos por ter um mundo biológico cujos sobreviventes operam de acordo com as seguintes duas regras:

“Faça cópias de si mesmo.”

“Encontre os pontos de aptidão alta no espaço vizinho dentro do espaço de design”.

Conforme o espaço vai sendo preenchido, e interferências externas vão afetando o organismo (mutações, meteoros, campo eletromagnético, calor do sol, radiação, água etc.), os pedaços do espaço de design que são desejáveis vão mudando, e o espaço vai se densificando até o ponto no qual temos o espaço físico preenchido de vida, por exemplo no nível da floresta amazônica. Diferentemente do campeonato com as moedas (capítulo 3), o jogo muda conforme o tempo passa, mas existem regularidades a serem exploradas (como a gravidade, o empuxo da água, a força de atrito em relação ao chão) que duram muitíssimo mais do que necessário para que haja aptidão garantida para aqueles que se aproveitarem bem essas constantes. Existem espaços intermediários, que duram para quase todas as linhagens mais tempo do que suas vidas (detectores de fótons, decodificadores de vibração no ar, captadores de moléculas do ar). A caminhada através do espaço de design dos seres vivos e seus descendentes é interessante, e se assemelha muito a um campeonato no qual o esporte vai mudando gradativamente e imperceptivelmente.

Dennett (1995a, p. 55) traça uma comparação interessante, com um campeonato em que há uma mistura de esportes:

Se você jogar tênis no primeiro round, xadrez no segundo round, golf no terceiro round e bilhar no quarto round, não existe razão para supor que o eventual vencedor irá ser particularmente bom comparado com o todo, em *nenhum* desses jogos – todos os bons jogadores de golf podem perder na rodada de xadrez e nunca ter uma chance de demonstrar sua proeza, e mesmo se a sorte não tem um papel no quarto round, a final de bilhar, o vencedor pode acabar sendo o segundo pior jogador de bilhar no todo.

Esse exemplo serve bem para contrastar a sorte com a habilidade, e habilidade, no jogo da vida, é estar perto dos espaços de maior aptidão no espaço de design. Para que esses espaços existam, é necessário que a estabilidade da informação genética seja maior do que a instabilidade das mutações, e que a estabilidade dos espaços de maior aptidão dure o suficiente para que alguém os alcance. Considerando que no nosso mundo existem muitas adaptações convergentes, temos razões para crer que essas condições vêm sendo satisfeitas há no mínimo 3 bilhões de anos em nosso planeta.

A evolução pode ser pensada como um algoritmo. O input do algoritmo envolve muito “ruído” (*noise* é um termo técnico para informação não relevante que invade um sistema de comunicação quando uma mensagem é transmitida), porque os sistemas biológicos são abertos, ou seja, podem ser afetados pelos sistemas físicos a sua volta.

Vale salientar que sistemas biológicos são sistemas físicos, são um tipo especial de sistema físico que nos interessa, e uma das principais razões pelas quais esses nos interessam é porque parecem estar sendo parte de um grande algoritmo evolutivo.

Máquinas feitas por homens são em geral artefatos feitos para realizar algum tipo de tarefa mecânica. Mais recentemente inventamos máquinas que (mecanicamente) computam informação e, por conta do caráter duro da nossa imagem de máquinas, tendemos a achar que elas não vão errar. Uma máquina consegue calcular a décima oitava potência de 3 antes que eu possa dizer ‘desoxirribonucleico’, e por isso ela parece infalível. Mas isso se deve ao fato de que a forma de processamento de informação das máquinas é bem adaptada para processar números grandes, e o nosso cérebro não evoluiu para isso. Como Marvin Minsky pontua, ainda não temos nenhum robô que possa colocar uma fronha num travesseiro. Nossas adaptações são adaptações que nos conferem algumas habilidades superiores as dos computadores, e o design dos computadores lhes confere algumas habilidades superiores às habilidades humanas. Examinaremos no próximo capítulo algumas concepções intuitivas humanas que Dennett considera falsas, e vamos expor seus argumentos em favor de suas posições contraintuitivas.

O algoritmo Dennettiano

If there is an elephant in the room, introduce him. – Randy Pausch

Quando esse texto foi apresentado como qualificação de mestrado, os membros da banca enfatizaram que se o objetivo do texto é produzir uma simulação mental na cabeça do leitor, e fazê-lo compreender não apenas *as ideias de Dennett*, mas *sua maneira de pensar e como chegou a essas ideias*. Assim, o quanto antes fosse introduzida a visualização do algoritmo dennettiano, melhor. Agora que temos uma noção um pouco mais precisa de “algoritmo”, estamos em um bom momento para introduzir, de maneira um tanto brusca, o algoritmo dennetianno. Faremos isso de forma gráfica, na Fig. 2, e deixaremos todos os aspectos do algoritmo para serem explicados no capítulo 11, “O algoritmo dennettiano”. Até lá, esse mapa visual dos pequenos agentes que comandam o pensamento dennettiano permanecerá com o status de um “elefante na sala”. Mesmo que propriamente apresentado, ele deixará diversas

interrogações na mente de seus hospedeiros. Que essas sirvam de motivação para uma leitura *cognitivamente ativa*, que não planeja apenas reter informações, mas reorganizar processos.

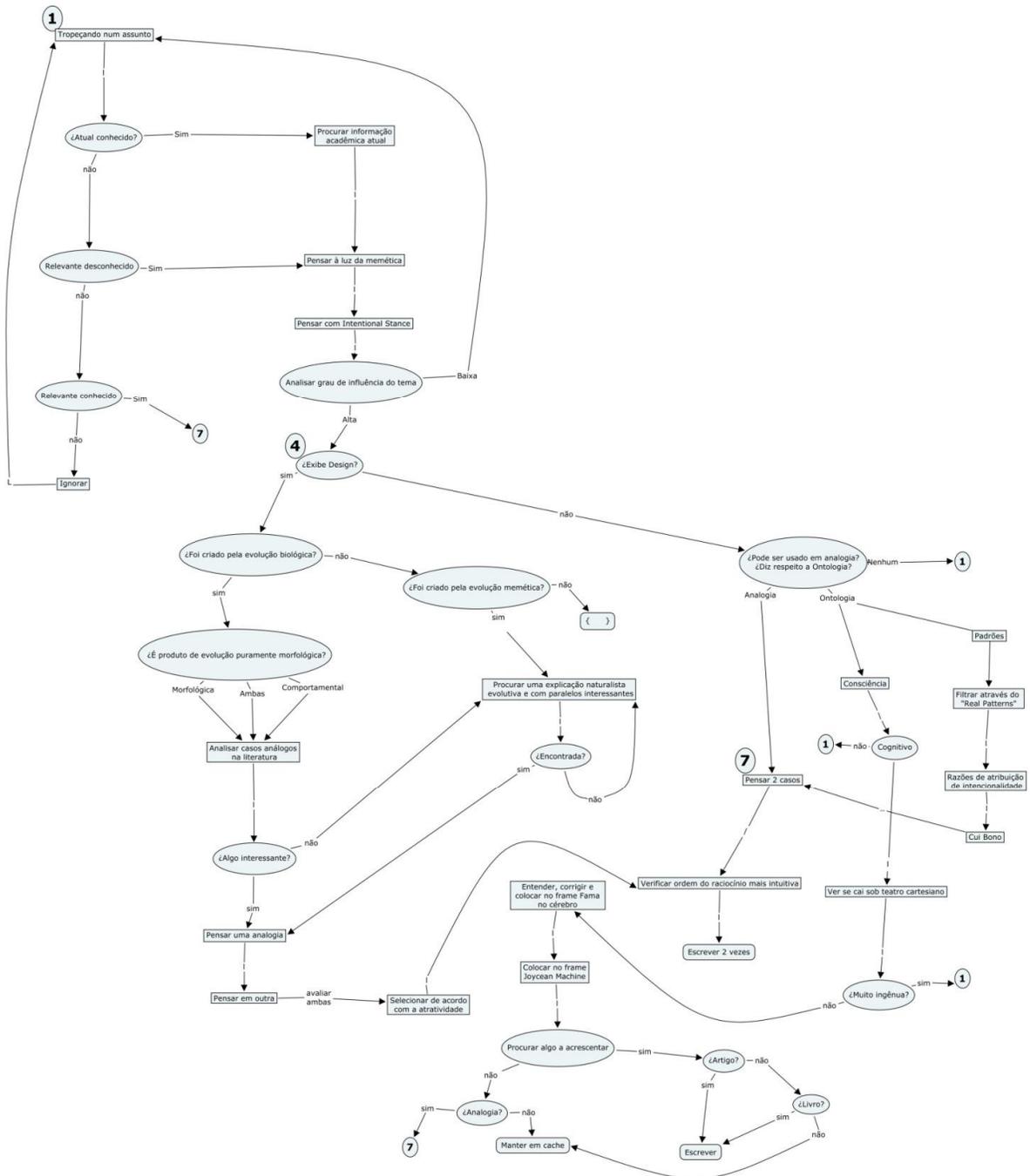


Figura 2: Representação esquemática do algoritmo dennettiano.

5. Desafiando concepções intuitivas, desqualificando o sujeito-autoridade

As perspectivas de análise: física, design e intencional

Uma vez que as descrições que fazemos de nossos estados psicológicos não são em forma de estado de Máquina de Turing (ainda que, lembremos, possam ser), cabe encontrarmos uma forma de descrever aquilo que dizemos e aquilo que queremos dizer em outros termos. Uma boa explicação pode ser entendida como isso mesmo, uma maneira de expressar algo em outros termos, até um ponto de satisfação no interlocutor. Com efeito, é desse recurso que Dennett vai se valer em seu texto “Sistemas intencionais” (1971). Talvez minha principal discordância em relação a Dennett seja a importância que ele dá atualmente a esse texto e a que eu dou. Nessa primeira seção faremos uma exploração dos temas desse texto.

A primeira coisa que devemos saber a respeito de um sistema intencional, que será definido adiante, é que algo não é por si só um sistema intencional, mas só é um sistema intencional da perspectiva de alguém que esteja tentando compreender ou predizer seu comportamento.

A palavra “intencional” é geralmente usada por filósofos com um significado distinto, querendo dizer o mesmo que “*aboutness*”, ou “ser a respeito de algo”. Algo é intencional, neste sentido, se é a respeito de alguma coisa. Por exemplo, o título desse livro é a respeito do livro. Há algum tempo na filosofia é aceito que uma propriedade que marca os estados mentais é sua intencionalidade, ou seja, o fato de que eles são a respeito de objetos. Em particular, expressões idiomáticas de crenças, desejos, esperanças, são sempre a respeito de algo. Dizemos que “cremos *em* estatística, mas não cremos *em* astrologia”, por exemplo.

Existem três formas principais, três perspectivas¹⁹, ou posturas de explicação para o comportamento de um determinado sistema. A mais comumente pensada é a perspectiva do *design*, na qual fazemos previsões de comportamento de um sistema a

¹⁹ O termo em inglês é *stance*, que não tem tradução consensual; na versão de *Brainstorms* feita em 2006 para o português, o tradutor escolheu “postura” (Dennett, 1978d), mas acredito que “perspectiva” passe melhor a ideia de *stance*, já que estamos falando de diferentes perspectivas de análise, e dificilmente alguémalaria em “posturas de análise”.

partir dos porquês de seu design: podemos prever que um ar condicionado esfriará o ar se apertarmos o botão porque ele foi desenhado para isso.

Outra perspectiva de previsão comportamental é a perspectiva *física*. Não é comum usarmos a física porque lidamos a maior parte do tempo com utensílios humanos, mas a utilizamos, por exemplo, quando dizemos que o sol nascerá amanhã, ou que a TV não vai ligar enquanto não a pusermos na tomada. Essa forma de previsão é utilizada em geral quando a explicação física é bem genérica e se está tentando explicar o mau funcionamento de um sistema, no caso, a TV, por razões que não são um mau design.

Dennett pontua que existem casos nos quais nenhuma dessas perspectivas é a melhor para fazer previsões, como ocorre no caso de programas que jogam xadrez; fisicamente, nem o melhor dos calculistas conseguiria prever seus movimentos, dada a enormidade de variáveis, e o comportamento também não é facilmente previsível pela forma como o programa foi desenhado. Assim sendo, uma terceira estratégia é melhor, caso se queira ganhar o jogo: assumir que o computador está pensando como um oponente humano. Ou seja, não faremos previsões com base na equação de Schrödinger ou nos planos de design daquele programa específico. Assumiremos que o computador é um agente como nós.

Essa perspectiva, a perspectiva *intencional*, que assume um design ideal, também está sujeita a falhar, por exemplo porque o design pode não ser ideal. Mas isso não deve ser considerado uma falha; quanto mais próximo de um sistema ideal, mais eficiente e útil será a utilização da perspectiva intencional. Trata-se de uma questão de grau, e não binária, de sucesso e falha.

Essa é a perspectiva intencional, que assume que um sistema possui certas informações e é direcionado a certos objetivos. ¿Mas o que queremos dizer aqui com “possui informação”? Possui informações em um sentido mais preciso do que o de armazenagem, dado que esse sistema há de possuir essas informações em um sentido epistêmico, superior, por exemplo, a possuir informações no sentido que um chinês que tenha um livro em português possua essas informações. As informações têm de estar na forma certa e no local certo para serem utilizáveis. Um programa que joga xadrez possui crenças a respeito da posição atual do tabuleiro, e tem o objetivo de vencer o jogo num tempo limite.

É um pequeno passo chamar de crenças e desejos as informações possuídas e os objetivos. Não está em discussão se podemos *realmente* dizer que são crenças e desejos,

apenas que se atribuirmos crenças e desejos a esses sistemas, poderemos fazer previsões que doutro modo não poderíamos.

Se fossemos a outros planetas, e víssemos criaturas complexas produzidas pela seleção natural, atribuiríamos a elas crenças e desejos, e isso não implica estarmos atribuindo nenhuma qualidade humana particular, antropomorfizando-as. Em virtude da racionalidade desses seres, podemos assumir que eles compartilhem conosco sua crença em verdades lógicas, no sentido de poder traçar inferências, e também que eles não têm desejo de se destruir.

Nós usamos comumente a perspectiva intencional como estratégia de previsão, ao tentar entender animais, outros humanos, e mesmo computadores.

Nós supomos a intencionalidade em homens e animais naturalmente, tanto o é, que quando existe algum mal funcionamento no sistema, ou melhor, quando nossas previsões a respeito do sistema não são confirmadas (por exemplo, quando uma pessoa não vira para mim ao ser chamada), costumamos atribuir primeiramente o erro a alguma forma de deficiência na entrada de informação, como quando dizemos que uma pessoa não ouviu, que ele não deve saber que X, que ela não entende nossa língua etc.

A atitude que tomamos com relação aos insanos é aquela de abandonar a perspectiva intencional e passar à perspectiva do design, prever seus comportamentos não mais pressupondo um design ideal, mas sim um padrão de comportamento.

A ideia de uma teoria que descreve a mente como um artefato que evoluiu por seleção natural é basicamente a ideia de uma teoria que consiga quebrar em partes menores a perspectiva intencional que adotamos uns em relação aos outros, e explicar essas partes de uma perspectiva do design. O design por sua vez se explica pelas razões sintáticas e mecânicas expostas no capítulo 3. É interessante notar que num nível importante de análise, o pensamento adaptacionista, que fala na função de algo, é o mesmo que a perspectiva intencional adotada com relação à evolução. Procura-se no pensamento adaptacionista uma explicação racional, em termos de objetivos (de um conjunto de bases nitrogenadas, é verdade) para que haja um conjunto de características (olhos de águia) no mundo. A única razão pela qual nos permitimos fazer isso é porque funciona. As previsões feitas usando a perspectiva intencional dão resultado tanto para entender homens, quanto zebras ou programas que jogam xadrez.

Explicar a inteligência só em termos de não-inteligência

Se quisermos explicar a inteligência, uma maneira de fazer isso é dividindo a informação processada no cérebro em símbolos, que significam uma ou outra coisa. Porém, essa estratégia é limitada. Quando permitimos que o sistema ainda tenha símbolos não esfacelados, ainda há uma pressuposição de inteligência interna a cada nódulo do sistema dividido. Um símbolo que ainda contém significados distintos de suas próprias conexões estruturais e físicas ainda não está completamente entendido ou esmiuçado. Ele requer inteligências externas para ser compreendido em sua significação. Isso não seria um problema se essas inteligências forem menores que a original, mas se forem iguais, teremos um processo de recursão infinita, uma das poucas coisas que filósofos aprenderam que nem eles podem aceitar: o famigerado homúnculo interno, que não pode existir. Ao explicar a inteligência com pequenas inteligências divididas, estamos facilitando o processo, mas somente teremos uma explicação da inteligência se formos capazes de descrevê-la de tal maneira que não sobre nenhum constituinte inteligente na explicação final. Uma boa explicação é aquela que consegue explicar um objeto em termos de algum outro, uma propriedade que pode ser compreendida a partir de outras. Não pode haver homenzinhos na cabeça que seriam responsáveis por ir ao banheiro quando precisamos, ou investir no mercado financeiro: esse trabalho tem de ser feito por agentes não inteligentes.

Uma das vantagens de um sistema intencional, que é pensado como uma estratégia preditiva, é que não é necessário fazer perguntas metafísicas sobre ele. Se ele *realmente* pensa não está em questão, se é consciente, tem alma, se é emergente²⁰ nada disso influencia o fato de que tudo o que se segue de ser um sistema intencional se seguirá dele. Um sistema intencional funciona, isso independe do que ele seja em sua realidade última. Podemos atribuir crenças e desejos e isso nos dá poder preditivo; as questões se encerram aí quando a definição é essa.

Quine, um dos grandes filósofos do início do século passado, considerava haver um círculo intransponível aí. Ao atribuir crenças para um indivíduo, dependemos dos

²⁰ Os conceitos de “alma” e “emergência” têm ambos uma característica bastante perigosa dentro da filosofia e das ciências. Eles são usados como *substitutos de explicação*. Ao invés de me mostrar como um indivíduo faz alguma coisa, você passa o bastão para uma alma, dizendo que a alma é que faz a coisa. O grau de mistério de uma questão, por exemplo ¿Como os estados psicológicos subsistem na massa gordurosa do cérebro?, é exatamente o mesmo antes e depois de respondermos: “É um fenômeno emergente”. Estar atento a esse tipo de tentativa de escapar, sem pagar o preço de explicar, é uma virtude frequente em Dennett, que não se utiliza da maneira comum de palavras como ‘alma’, ‘real’, ‘emergência’, ‘existente’, entre outras vilãs da história da filosofia.

desejos, e ao atribuir desejos, dependemos das crenças, *ad infinitum*, não há o que fazer. Dennett contrapõe a isso a ideia de que mesmo que, em casos particulares, não tenhamos como confirmar as crenças de uma pessoa (poderíamos perguntar, mas a resposta dela só seria útil sob o pressuposto de que ela tenha o *objetivo* de colaborar), em geral podemos ultrapassar esse obstáculo. Isso porque numa análise geral, com muitos casos particulares, várias iterações, e evidências vindo de diversos *fronts* e maneiras de interpretar, teremos razão para supor alguma coisa, para um lado ou para outro, o suficiente para podermos começar a fazer nossas inferências. Embora haja interdependência entre crenças e desejos, no longo prazo essa interdependência deixa de ser viciosa e passa a ser parte de um modelo assentado que possui crenças e desejos de um agente intencional itinerante²¹.

A assunção de intencionalidade deriva da assunção de racionalidade, mas é necessário desmamá-la se quisermos algo *extra* de interessante em comparação com a visão, digamos, de “homem racional” dos economistas. Quando atribuímos crenças e desejos a um cão, que fez um caminho muito idiota para pegar o osso, por exemplo, estamos utilizando o conceito de sistema intencional para pensá-lo, mas não de sistema racional. Podemos apreciar o fato de que ele não foi capaz de pegar o osso da maneira simples, provavelmente porque não conseguiu processar alguma informação relevante (“meu corpo passa por baixo da grade”).

Em humanos, acreditamos no que as pessoas nos dizem crer. Há uma justificativa evolutiva para isso: qualquer linguagem que evoluísse para contar mais mentiras do que verdades tornaria benéfica uma mutação para *não entender nada*. Ou seja, uma linguagem só pode evoluir se ela for, na maioria dos casos, para espalhar a verdade, e para crer no que se diz. De qualquer forma, assumimos a perspectiva intencional quando aceitamos o dito pelo crido, o falado pelo acreditado.

Quando a TV não funciona, adotamos a perspectiva física como explicação inicial para o problema, até que encontremos uma tomada desligada. Uma TV é um artefato cuja interpretação padrão é dada no nível do design, e recorremos ao nível imediatamente inferior para procurar compreender um problema no sistema. Ao

²¹ Aos que permanecerem descrentes sobre a possibilidade de um sustento de longo prazo para um sistema de duas variáveis interdependentes que não são baseadas em nada externo, sugiro encontrar argumentos análogos (altamente especializados) em Jeffrey & Berger (1992) e Arntzenius (2008). Embora tratem de assuntos que apenas tangenciam o que se discute aqui, tomados em conjunto eles mostram como um raciocínio epistêmico bayesiano justifica a estabilidade de longo prazo de tensões conflitantes, variáveis interdependentes etc. Nota: em 2013, Arntzenius me disse que um aluno seu parecia ter refutado a contento o seu artigo, mas o texto ainda não havia sido publicado.

recebermos a mensagem “este e-mail não consta em nosso cadastro”, procuramos um erro de digitação. Analogamente, se um humano falha em sua razão, de acordo com o que acreditamos que ele queira e saiba, recorremos a uma explicação por design (o estudo de *cognitive bias*, ou vieses cognitivos, parte da psicologia evolutiva, é o campo justamente que trabalha com isso). Se alguém falha, talvez a psicologia explique, ciência esta que estuda (ver Kahneman, 2011) as sub-mentes contidas numa mente e seus designs. Se ainda assim não ficar claro o comportamento de alguém, podemos recorrer à perspectiva física: “Provavelmente ele caiu no poço do elevador porque não raciocinou direito” não é uma boa explicação, mas “Ele caiu no poço do elevador porque, apesar de não desejar cair, e acreditar que estava caindo, não pode impedir a brutalidade da força da gravidade, e caiu” soa mais razoável.

O fim do autoritarismo do sujeito, distribuição de poder

Dennett recusa-se também a aceitar que haja uma autoridade do sujeito (como tradicionalmente concebido) sobre seus estados mentais. Essa crítica dirige-se fortemente à fenomenologia e às vertentes da psicologia que defendem as imagens mentais, e está no texto “Duas abordagens às imagens mentais” (Dennett, 1978c). Sua crítica consiste em dizer que o acesso presumivelmente privilegiado que temos a nossas imagens mentais não é privilegiado *no sentido que desejamos* e que, portanto, boa parte do debate fenomenológico está na realidade tratando de não-questões. Não podemos atribuir como causa proximal próxima de nossos atos os supostos conteúdos fenomenológicos aos quais nos referimos; assim, temos de abandonar essa intuição, por mais forte que possa ser, e seguir viagem numa análise na qual essa intuição não interfira. Nossa intuição nos sugere que há algo de fenomênico e imediato, mas não podemos conceder status de conhecimento a essa intuição se ela não puder justificar nossas elocuições a seu respeito. Se não houver elo causal entre as imagens mentais e os relatos de imagens mentais, então há um problema grave com a forma como imagens mentais são pensadas.

Num texto muito similar em espírito a este último, presente no livro *Brainchildren* (1998) e intitulado “Do it yourself understanding”, Dennett defende que se uma sentença, como esta, não estivesse sendo processada por aquilo que você acha ser *seu* cérebro, não haveria entendimento algum. O significado não tem nenhum efeito diante do motor sintático que é o cérebro. O físico David Deutsch (1997) sugere uma

maneira interessante de definir “real”, que deve ser usada mais como uma piada *contra* qualquer outra tentativa de definir esse termo do que levada a sério: “*If it kicks back, its real*”, isto é, “Se chuta de volta, é real”. Munidos disso, temos de lembrar que o cérebro é um composto mecânico, um gigante motor sintático que opera de acordo com um ordenamento mecânico/sintático. O cérebro é sintático, e é nesse nível de operação que ele faz o que faz, inclusive entender. Tudo o que ele chuta de volta, piadas, tacadas de hóquei, gestos obscenos, cantos líricos, é refletido por esse sistema cheio de colunas corticais e redes neuronais. Se você acha que está entendendo algo *você mesmo*, como algo *oposto* a conjuntos de neurônios sendo sintaticamente modificados, você ainda não se deu conta da ideia de entendimento. O entendimento tem de ser feito por recursos não inteligentes.

Em sua mini-autobiografia, Dennett (2008b) nos traz de volta ao ponto em que começou a refletir a esse respeito, na década de sessenta:

Esse foi o primeiro de muitos encontros em que eu deparei-me com meus amigos filósofos da mente estranhamente complacentes em sua ignorância a respeito de cérebros e psicologia, e principiei a definir meu projeto como sendo o de compreender, como filósofo, como cérebros poderiam ser, ou suportar, ou explicar, ou causar, mentes. Perguntei a um amigo estudando medicina em Oxford do que cérebros eram feitos, e lembro-me vividamente dele desenhando diagramas simplificados de neurônios, dendritos, axônios – todos termos novos para mim. Imediatamente me ocorreu que um neurônio, com múltiplos inputs e outputs ramificados e modificáveis, seria justamente a coisa que poderia compor-se em redes que poderiam aprender por algum tipo de processo evolutivo.

**Sutis solavancos separam semântica e sintaxe:
¿Sem significado, sobra suficiente solução sintática?**

É bem conhecida dentro dos círculos profanos da filosofia a distinção entre copo e ‘copo’. O primeiro serve para beber água e o segundo contém quatro letras na fonte Times New Roman. O primeiro pode ser usado para prender um inseto, e o segundo pode ser usado para preencher a sentença “Tenho sede e um poço profundo, se ao menos eu tivesse um _____”.

“O nada nadifica” é uma sentença entre aspas, ¿ela significa alguma coisa que importa? Bem, matéria move matéria, e significados não. Então ainda que “nada” signifique algo, o que realmente importa (ou seja, o que é fisicamente efetivo) é “ ‘nada’ ”, a informação sintática contida no papel, que afeta causalmente nosso órgão

de entendimento.

Significados são para *previsão e controle*, e não para explicação. Explicações não de descrever níveis sintáticos, porque o cérebro é um motor sintático. Se eu quiser controlar e prever a chuva para preencher meu copo, devo atentar aos significados dos gráficos e palavras da previsão do tempo fornecidas por uma breve pesquisa na web. Se eu quiser explicar o fato de que saí de casa com um copo em riste, pouco tempo antes de chover, terei de submeter-me a elementos de sintaxe para explicar o fenômeno. Afinal, o que *choca e impressiona* não é o fato de que *dado que eu entendo* o significado da previsão do tempo, eu saia na hora certa; o que *impressiona e requer explicação* é que eu posiciono o ponto de foco de meus olhos numa tela luminosa contendo enormes quantidades de padrões escurecidos, contrastando ao fundo branco, movimento os olhos da esquerda para a direita seguindo uma velocidade fixa, e, horas depois, saio de casa com um dispositivo capaz de conter líquidos precisamente na hora em que líquidos vindos do espaço principiam a mover-se em direção a mim.

Ainda que estejamos todo o tempo imersos nesse complexo oceano sintático, os significados parecem estar por aí, afetando nossas vidas com placas de trânsito, legendas etc. O que é necessário é explicar porque ocorre essa *harmonia*.

Após nossa análise do adaptacionismo, podemos começar a ver de onde pode emergir uma resposta: a harmonia, assim como a presença de olhos, é uma boa estratégia, e portanto nós chegamos perto da região do espaço de design no qual conseguimos operar mais ou menos de acordo com regras de significação, mesmo sendo apenas organismos físicos (e portanto sintáticos). É uma boa estratégia saber simbolizar, e por isso aprendemos a fazê-lo. Nós somos o produto de uma especialização sintática que faz uma mímica da semântica, e caso os significados mudem, em pouco tempo (evolutivamente falando) desenvolveríamos novas adaptações que estão em *harmonia* com a esfera semântica. Isso não é uma correspondência, muito menos uma correspondência de 1 para 1, biunívoca, entre o mundo semântico e o mundo sintático, mas é o melhor grau de correspondência que podemos esperar, é o limite do fisicamente permissível.

Significado dividido em níveis de complexidade

Assim como faz com respeito à autoridade que temos sobre nossos estados fenomenológicos, ou estados mentais, que discutimos na primeira parte deste capítulo,

Dennett fará uma crítica entusiasmada de algumas suposições que temos sobre a nossa autoridade de conferir significado às frases. O artigo “Beyond belief” (Dennett, 1982b) trata extensamente de suas concepções sobre isso. Uma abordagem resumida desse artigo é uma empreitada provavelmente mais difícil do que uma reanálise de qualquer outro artigo do autor, e o recorte que pretendo fazer aqui faz jus apenas a alguns dos pontos que ele pretende clarificar no texto. Perguntei a ele, e ele não discordou de que esse seja seu mais brilhante artigo, mencionando que foi aquele no qual ele trabalhou mais longa e arduamente.

Dennett se propõe a discutir estados psicológicos (como *crer*, *desejar* etc.), frequentemente chamados de atitudes proposicionais, e significado. Mantendo a pretensão de possibilitar uma simulação do pensamento dennettiano no leitor, vou considerar inicialmente o significado, pois é sua concepção de *o que forma* o significado que mais se enquadra nos objetivos desse capítulo. Ou seja, pretendo mostrar como ele desafia uma noção intuitiva que temos sobre *quem tem autoridade* sobre o significado de uma proposição.

Dennett apresenta algumas possibilidades de formação do significado, e apresento aqui a mais extensa e mais próxima de sua posição:

- (-1) Propriedades físicas + Considerações de design (ou menos irrelevâncias funcionais) → Propriedades sintáticas
- (0) Propriedades sintáticas + Convenções linguísticas → Caráter (*character*, conforme concepção de David Kaplan)
- (1) Caráter + Contexto → Conteúdo
- (2) Conteúdo + Circunstâncias → Extensão.

Nesse modelo, a forma através da qual uma sentença produziria significado é atravessando todo o percurso apontado. A sentença “água é H₂O” poderia ser examinada da seguinte maneira: algumas vibrações nas moléculas do ar causadas pelo aparato vocal do Falante, em condições normais para comunicação, atingiram o ouvido de Ouvinte; essas vibrações foram encapsuladas em unidades discretas, que ignoraram minúcias e nuances funcionalmente irrelevantes para um bom falante da linguagem, e com isso a mente de Ouvinte instanciou uma forma sintática canônica, com histórico e fronteiras bem definidas. Continuando seu percurso de modificações, a informação que agora estava representada como uma forma, ou estrutura, foi colocada (presumivelmente por interferência do restante da maquinaria mental) em contexto, e

tornou-se uma forma com conteúdo. No entanto, esse conteúdo, “água é H₂O”, ainda não encerra a determinação do significado; essa informação, colocada dentro das circunstâncias de enunciação, ganha ainda mais sofisticação, e então determinaria uma extensão, encerrando o processo.

Essa *não é* a posição adotada por Dennett, mas explorá-la ajuda-nos a entender a caracterização de sua posição.

A primeira coisa que se deve notar é que essa caracterização não está versando sobre uma proposição em geral, mas sim sobre um caso (*token*) particular da enunciação de uma sentença. Ou seja, a pergunta que ela responde não é “¿O que significa a proposição ‘Água é H₂O’?”, mas sim “¿Qual a forma genérica a partir da qual eu posso significar *cada caso singular de enunciação da sentença* ‘Água é H₂O’?”. Essas perguntas são muito diferentes, e partem de ontologias²³ diferentes. No primeiro caso, se supõe (deseja e postula) que o significado esteja de alguma maneira contido na própria sentença; uma sentença seria, por ela mesma, capaz de significar algo, mesmo que solta no mundo, sem sujeito ou lugar. A segunda pergunta tem uma ontologia mais humilde, ela não pretende alegar a existência do significado nesse sentido, apenas se pergunta, para os casos *efetivamente ocorrentes* de sentenças, se há uma caracterização geral do fenômeno de como se conecta a sentença à pretensa extensão que ela determina.

Recapitulando: foi-nos dada uma sentença, e estamos pensando sobre como aquele conjunto de sinais se conecta ao significado deles. Para entender isso, temos de pensar o que faz o significado variar, e o que o mantém invariante. Mudar o contexto pode mudar o significado da frase; por exemplo, “eu estou escrevendo” foi verdade na hora que eu escrevi, mas não é verdade agora que você está lendo, porque o contexto de enunciação é outro. Ou seja, queremos entender uma sentença com valor de verdade (proposição) e como ela se conecta com esse valor de verdade. Poderíamos tentar responder essa questão em geral, fazendo uma teoria geral do significado, ou uma teoria que vale para os casos particulares dentro de seus contextos de enunciação, e das circunstâncias nas quais foram enunciados.

O *contexto* diz respeito à localização na configuração do texto falado ou escrito, e associações com elementos ao redor. As *circunstâncias* dizem respeito a condições externas do mundo naquele caso particular de elocução.

“Esta sentença é muito grande” pode ser verdadeiro num contexto em que

²³ *Ontologia* é um termo filosófico significando “aquilo que diz respeito ao existente” ou “conjunto de ideias a respeito do que existe”.

estamos exigindo sentenças de no máximo duas palavras, e falso num contexto de crítica pós-moderna de arte.

“Faz sol hoje” pode ser verdadeiro dependendo das circunstâncias externas no momento no qual é dito. Ou seja, existem ao menos dois graus de liberdade (contexto e circunstância) diferentes operando em algum lugar entre o nível sintático e o nível semântico. ¿Mas será que ambos os níveis estão presentes no mesmo *caráter* (*character*) (rever acima)? Sim. Por exemplo: “Faz sol hoje” pode variar seu valor de verdade de acordo com o contexto, caso esteja inserido numa história que estou contando.

“Era uma vez uma garota que disse, num belo dia de sol, ‘faz sol hoje’, quando um lobo se aproximou...” Nesta história, a frase é verdadeira mesmo que seja contada num luau ao redor da fogueira; seu valor de verdade, sua falsidade ou veracidade, está determinado dentro do contexto da própria história. A mesma frase pode variar pelas duas causas, contexto e circunstâncias. Ou seja, as duas possibilidades de variação, os dois “manípulos”²⁴, estão presentes em cada caso de elocução; e o que é mais interessante: *nós somos capazes de processar essa informação!*

¿O que é o nível do *character* (e porque eu mantive a notação em inglês)?

O filósofo David Kaplan (1989) percebeu que existe um nível de estabilidade que está acima das convenções linguísticas (que discretizam a informação sonora), mas abaixo do nível de variação conforme as circunstâncias ou contexto. É o nível no qual a frase “Faz sol hoje” permanece a mesma, independente de estar numa ficção, num discurso político ou numa citação de livro. Esse nível é diferente do nível de significação, porque não é suficiente para determinar o valor de verdade, mas é muito importante, porque é o nível que melhor captura a invariância entre “Faz sol hoje” em seus diversos contextos, mesmo que a tipografia mude, ou a sentença esteja em código Morse etc. *Character* é o nome que ele atribuiu para isso, o que quer que seja, que se mantém a respeito dessa sentença onde quer que ela apareça.

No uso cotidiano, costumamos misturar todos os níveis de significação da linguagem ao falar sobre sentenças. O nível *character* serve para iniciar uma separação dessa mistura. Como isso é fundamental para compreender a filosofia da linguagem e da mente, prefiro manter o termo conforme ele aparece na literatura em inglês, porque ele é *importante e específico*, e não queria perdê-lo na tradução.

²⁴ *Manípulos* são botões giratórios manuseáveis, como os de volume em aparelhos de som de automóveis.

**Três manípulos para sentenças, modificando
(1) circunstâncias, (2) contextos e (3) quantidade de contexto interno**

Após trazer à tona essa divisão de uma sentença, Dennett (1982b) então pretende dar uma explicação de como a informação vai se compor para determinar o significado, e nesse caso o nível do *character* seria o que mais se aproxima do nível proposicional. No entanto, Dennett percebe que, uma vez que abandonamos a condição da proposição como independente e pura, somos levados a caracterizá-la não só embebida em um contexto, mas embebida em *uma certa quantidade de contexto*. O significado de uma fala varia em função da caracterização contextual que damos a ela. Não só varia para diferentes situações (casos de variação *indexical*²⁵, por exemplo), como varia também em relação às bordas com relação às quais delimitamos o nosso espaço de significação. Um exemplo tornará isso mais claro. Uma estrutura mental em mim instancia a seguinte informação “Todos os Fs são J e T”; se quisermos estabelecer os limites do significado nas bordas dessa região, então só teremos acesso ao que ocorre nela, e tudo que saberemos é que todos os Fs são J e são T. Suponha agora que modificamos *somente nossa perspectiva de análise*, mantendo contexto e circunstâncias invariantes. Ao invés de delimitar a fronteira de significação na área imediatamente ao redor da área que instancia a regra que todos os F são J e T, e manter os objetos J, T e F como indeterminados, permitimos uma expansão dessas fronteiras até tal ponto em que os significados previamente indeterminados de J, T e F já possam estar na área interna de nossa análise. Colocamos esses objetos para dentro da fronteira. Nesse espaço mais amplo (digamos *eu* como um todo), aquela informação será bastante mais rica, podendo significar “Todos os filhos são ingratos e desobedientes”. A mesma estrutura semântica/cognitiva poderia, em um contexto diferente, significar “Todos os Papas são homens e católicos”, e isso é um ponto interessante. Mas mais interessante é perceber que *na mesma situação*, dependendo somente da *quantidade de contexto* que permitimos observar, o significado também muda (ou seja, dependendo dos limites a partir do qual dizemos que há receptores e efeitores, fronteiras que separam o exterior das partes internas do sistema). O significado não está preso a uma forma canônica, mas seu conteúdo pode se tornar mais rico, em função de quanto mais contexto nossa metodologia nos permite observar como parte interna do sistema de representação. O

²⁵ *Indexical* é o termo dos filósofos para proposições ou assuntos autorreferentes, ou que se referem a uma circunstância externa apontada. Por exemplo, “Eu me chamo Diego Caleiro” é um *indexical*, ou “Faz sol hoje”.

significado depende da amplitude do nosso espectro de observação, além de depender, externamente, de circunstâncias e contextos distintos, como vimos. O significado depende de quanto contexto é interno e quanto é externo.

Dennett reduz então as regras de significação apresentadas acima (das quais, lembremo-nos, discorda) numa caracterização de um fenômeno mais geral que é:

$$C + E :eI .$$

Conceitos do tipo *character* mais informação sobre o ambiente (*environment*) levam a uma *informação mais rica*.

Essa caracterização, vale notar, retira a possibilidade de um conteúdo absoluto, exceto se nossa postura metodológica nos permitir ter toda a informação ambiental, caso no qual nossa teoria seria desinteressante porque não podemos ter toda informação (se tivéssemos toda a informação contextual, o significado se tornaria aquilo que um Demônio de Laplace²⁶ calcularia, ou seja, estaríamos de volta à perspectiva física, e teríamos saído da perspectiva intencional). Dentro da caracterização $C + E :eI$ sugerida, podemos expandir o zoom e retrair o zoom o quanto formos tecnicamente capazes com nossa mente e nossas análises. Temos que trabalhar então, ao pensar a significação, com *gradientes de contextualidade*, com níveis de compreensão do ambiente, e nesses níveis variará também o significado.

Isso nos levaria então à perspectiva de que o significado de uma sentença estaria no mundo nocional de uma pessoa, podendo ou não corresponder ao mundo real.

¿O que é um mundo nocional?

Mundo nocional, parte 1: é o mundo conforme a pessoa o vê, digamos uma imagem do mundo que ela tenha, só com a informação que lhe é epistemicamente acessível. Não se trata aqui de dizer que é um conjunto de mundos possíveis que têm algumas propriedades em comum, aquelas que a pessoa conhece; trata-se de imaginar as noções que a pessoa tem sobre o mundo que estão de fato na cabeça dela. É um espaço cognitivo cujas fronteiras são os limites dos conhecimentos e inferências dessa pessoa.

²⁶ O demônio de Laplace é uma invenção dos filósofos que funciona como um computador gigante, para o qual são dadas todas as informações físicas de um determinado instante. Seu uso ocorre principalmente em discussões sobre determinismo. Uma das maneiras de definir o determinismo é dizendo que o estado do universo num instante particular, mais as leis da física, seriam suficientes para determinar o estado de todos os instantes posteriores. ¿De que maneira? Presumivelmente através de um grande poder de cálculo preditivo de um demônio filosófico. Esse é o Demônio de Laplace (em geral, ele só recebe a informação física do nível fundamental, não dos níveis mais altos de descrição).

Unificando a perspectiva C + E erI com a forma de analisar os conteúdos mentais a partir da perspectiva de um mundo nocional, começa a revelar-se um caminho para as atitudes proposicionais, que Dennett vem pretendendo alcançar com sua discussão. Ele separa atitude nocional, isto é, o análogo de uma sentença que está dentro do mundo nocional de um indivíduo, de atitude proposicional, aquela noção filosófica da qual estamos tentando nos aproximar ao longo desse capítulo, sem incorrer nos erros que a filosofia pré-dennettiana incorria.

De mundos nocionais a atitudes proposicionais: rotas propostas e o fim da linha

Atitudes proposicionais são um tópico controverso mesmo para os iniciados em filosofia. “Dawkins crê que Dennett é brilhante” descreve uma atitude proposicional de Dawkins, por exemplo. “Mamãe eu quero mamar” descreve uma de um filho. “Me dá a chupeta”, por outro lado, não constitui atitude proposicional. Propositivamente desvio por hora de defini-las.

¿O que, segundo Dennett, os filósofos parecem querer que as atitudes proposicionais sejam capazes de fazer para sustentar o título?

1. São portadoras de um valor de verdade, isto é, são verdadeiras ou falsas.

2. São determinantes de extensões²⁷, extensões sendo definidas como aquilo que determina extensões. Por exemplo, “dois ao cubo” tem a mesma extensão que “o número de planetas”, mas intensões diferentes. Mais de uma intensão podem determinar a mesma extensão, mas uma só intensão determina uma só extensão.

3. São apreensíveis pela mente; somos capazes de compreender numa tacada.

É recomendável ir ao artigo para compreender porque Dennett (1982b) considera que essas três propriedades não podem ser sustentadas ao mesmo tempo.

Conectemos então o conceito de atitude proposicional com aquele de mundo nocional. Para tal, utilizaremos de um conceito intermediário, o de atitude nocional. Uma atitude nocional não tem seus referenciais no mundo exterior, mas no mundo nocional de um indivíduo. A atitude nocional “Mamãe eu quero” refere-se ao modelo simples de mamãe que o sistema cognitivo do filho tem, “Dennett é brilhante” refere-se ao modelo sofisticado e construído ao longo de diversas interações e leituras de Dennett que Dawkins possui. Vejamos o resultado:

²⁷ ‘Intensão’ significa algo diferente de ambos os significados de ‘intenção’.

Atitude nocional + ambiente mbatitude proposicional.

Esse esquema, apesar de resolver parte das questões relativas ao significado, parece totalmente contra-intuitivo. Suponhamos que eu seja trocado pelo meu *Doppelgänger* que vive na terra-gêmea (*twin earth*), um planeta no qual tudo é exatamente igual a aqui, salvo pelo fato de que a água lá é composta de XYZ e não H₂O. Intuitivamente, as crenças que tenho (e que um filósofo pode querer verificar se verdadeiras) são sobre a terra, não sobre a terra-gêmea. No entanto, nessa caracterização, quando fui teleportado, todas as minhas crenças passaram a ser a respeito da terra-gêmea (salvo as sobre mim mesmo). “Cavalos me dão medo” passaria a ser “Cavalos_{gêmeos} me dão medo”. *Não queremos uma teoria assim* e estaríamos inclinados a procurar alguma forma de resolver a questão sem termos de fazer essa concessão. É estranho demais supor que o referencial de todas as minhas atitudes proposicionais seja automaticamente mudado quando sou teleportado. Não temos um caminho seguro traçado até as atitudes proposicionais, e isso pode ser um problema, não de nossa capacidade, mas das próprias atitudes proposicionais; ou melhor, da tentativa filosófica de encapsular aquelas três propriedades sob um mesmo teto, e chamar esse teto de ‘atitudes proposicionais’.

Dennett trabalha distinções e teorias que aparentemente poderiam ser úteis para resolver essa problemática, mas considera-as todas insuficientes. Por brevidade, não irei enumerar e discutir essas teorias aqui, como também não pude contextualizar as premissas importantes e o desenvolvimento que ele faz do primeiro sistema de significação aqui apresentado. As noções aqui retratadas procuram apenas facilitar uma simulação de uma linha de pensamento, e não discutir todo o seu conteúdo.

A conclusão a qual chega é que qualquer um que tente importar as categorias necessárias para uma semântica formal, e tente colocá-las a serviço de uma teoria psicológica, está fadado a criar um monstro. Não há como transportar o domínio da significação para o universo psicológico. Uma vez mais, nossa suposta autoridade sobre os conceitos é desafiada. Não temos o direito de falar de significados que estão dentro de nós ou ocorrendo na linguagem fora de seus contextos. Sem contexto não há significado, há apenas a forma sintática. Dennett deve considerar isso uma solução boa, pois anteriormente fez uma afirmação a respeito de *todas* as soluções possíveis para o problema do significado: “Cérebros são *motores sintáticos*, então no final e em princípio as funções de controle do sistema nervoso humano devem ser explicáveis nesse nível ou

permanecer para sempre misteriosas” (Dennett, 1987 [1982c], p. 141).

Ao longo desse artigo (Dennett, 1982b), ele percorre uma linha de substituição. Começa lidando com o problema das atitudes proposicionais, ao demonstrar (e isso foi omitido aqui) que as necessidades que se costuma pressupor que as proposições tenham não podem ser todas atendidas; substitui então o domínio da significação pelas atitudes sentenciais, até que, finalmente, em virtude da característica sintática de nossos cérebros (uma restrição forte ao que nós podemos fazer), faz-se necessária uma substituição das atitudes sentenciais por atitudes nocionais. A referência das sentenças deixa de estar localizada no mundo e passa a estar localizada num mundo nocional²⁸, um construto (que pode ser) puramente sintático e que não precisa corresponder ao mundo, apenas de alguma maneira mapeá-lo *o suficiente*. É fundamental compreender que esse mapeamento não é um mapeamento perfeito ou sequer com pretensões disso, apenas uma aproximação cognitiva representacional do estado de coisas. Essa representação se constrói, como podemos ver no capítulo sobre significado do *A perigosa ideia de Darwin* (1995a), a partir de uma regulação sintática (física) que opera ao longo da evolução. A evolução biológica poderia ser pensada então como um algoritmo que, dentre suas finalidades secundárias, compreende a finalidade de sintaticamente produzir máquinas capazes de uma representação próxima do estado de coisas (relevantes à sua sobrevivência) ao seu redor. Já foram criados robôs com capacidade de aprender sobre sua própria forma²⁹ (sem que ela lhe seja dada anteriormente) e aprender a se movimentar pelo mundo; a representação de sua forma vai se aproximando da real forma conforme ele tenta se movimentar e vai percebendo sua forma. A visão proposta

²⁸ Mundos nocionais, parte 2. Na tradição da filosofia analítica, os mundos nocionais são iguados a um conjunto de mundos possíveis, que são reunidos por terem as características relevantes, e não terem nenhuma característica absurda (o exemplo de Dennett é que ninguém teria, por exemplo, três narinas). No entanto, como é intuitivamente claro, uma caracterização de tal forma não faz jus ao estado psicológico ou de significação de um indivíduo. Meu mundo nocional não precisa ter uma representação completa do mundo, mas apenas de parte dele. Essa noção foi refinada na linguística e o conceito de mundo nocional foi substituído por um composto de espaços mentais, que só possuem as características relevantes para uma determinada análise. O modelo mais completo do qual o autor tem notícia, que trabalha a significação através da fusão (*blend*) de espaços mentais, é o modelo semiótico-linguista cognitivo de Per Aage Brandt, exposto em Brandt & Brandt (2005), que dá conta da significação de metáforas (além das demais sentenças significantes já explicadas anteriormente, sem a necessidade da fusão), e leva uma teoria das representações mentais e da significação ao seu melhor, levando em conta a tradição analítica, semiótica e linguística. O comprometimento de Brandt com a fenomenologia e o dualismo são inessenciais para sua teoria do significado, e podem ser substituídos sem perda de informação por uma hetero-fenomenologia dennettiana, na minha visão com ganhos substanciais para ambos os lados.

²⁹ Hod Libson desenvolveu robôs que são *self-aware* (auto-conscientes), num grau bastante simples. Uma apresentação expondo esses robôs pode ser encontrada em: <http://www.ted.com/talks/view/id/165>

por Dennett colocaria os engenheiros que criaram esses robôs e a evolução lado a lado, como construtores de aparelhos sintáticos que aproximam um mundo nocional do mundo real, e todo o domínio do significado estaria portanto subscrito nesse mundo nocional, e livre portanto de teorias de referência ou das demais postulações metafísicas que se costuma utilizar para fazer a ponte entre o sintático e o semântico.

Lembrando da forma de reducionismo que Dennett defende, explicada no capítulo 1, o que ele faz aqui é parar numa determinada perspectiva, e tentar compreender *como* é possível, se realmente for o caso, que a semântica possa ser reduzida a sintaxe. Aceitando uma concepção menos forte de semântica (a da pergunta humilde – sobre *cada caso*), ele demonstra a possibilidade dessa redução.

No nosso nível de análise o significado então existe, mas outra questão se coloca quanto aos significadores, aqueles que conferem significado, aqueles dos quais o significado deriva. Contra-intuitivamente, os significadores não somos nós. Para conhecê-los, Dennett nos leva em mais um *intuition pump*, uma analogia que parcialmente reproduzirei aqui, pois se ele estiver correto, encerra um debate de muitos séculos na filosofia.

¿Quem são os significadores não significados?³⁰

Suponha que você queira conhecer a vida no século 25, que esse sempre foi seu maior sonho desde a infância, assim como o de muita gente era ser astronauta, porque os *video games* serão muito bons até lá. Enquanto a medicina não avança e nossas espaçonaves não atingem velocidades próximas à da luz, você só tem uma chance de fazer isso: paralisar seu metabolismo através de resfriamento, e tentar se colocar em uma cápsula que irá manter sua temperatura fria até lá, e acordá-lo automaticamente com um agradável aquecimento nos idos de dois mil quatrocentos e bolinha. De fato, há pessoas que fazem isso hoje em dia (porque preferem essa opção a morrerem e serem enterradas, por exemplo), que estão criogenadas, na esperança de que algum dia sejam reanimadas num estágio mais avançado do desenvolvimento tecnológico. Marvin Minsky, um dos principais cientistas da história da computação, que classifica Dennett como o novo Bertrand Russell da filosofia, está entre os pretendentes dessa técnica, por exemplo.

³⁰ No inglês fica muito melhor: ¿*Who are the unmeant meaners?*

Mas enfim, voltando ao caso, uma série de problemas começa a emergir quando esses são seus objetivos. *Construir* uma cápsula segura e protegida, garantir que *haverá energia* para 400 anos de resfriamento. Você não poderá depender de seus descendentes para isso, é claro, porque salvo se a medicina anti-envelhecimento progredir bastante rápido, eles também estarão mortos até lá. Além disso, quanto mais distante o parente, menor a confiança, como disse o biólogo J.B.S. Haldane: “Eu me sacrificaria por dois irmãos ou oito primos” (apud McElreath & Boyd, 2007, p. 82), numa alusão a um mecanismo evolutivo importante, chamado de *kin selection* (seleção de parentesco), sobre o qual falaremos mais adiante.

Você poderia confiar em dois filhos ou quatro netos para que *alguém* cuide de você, mas não é a melhor maneira de ter certeza que você chegará onde quer, nada garante que seus descendentes vão continuar se reproduzindo e se importando com seus ancestrais.

Há duas estratégias básicas para resolver a questão. Uma é encontrar um local seguro e utilizá-lo como uma base fixa, a outra é dotar sua cápsula de algum tipo de sistema motor, para que possa sair de locais que se tornem perigosos por quaisquer razões. Essa diferença entre as estratégias é análoga à diferença entre plantas e animais, no mundo biológico.

Por mais segura e escondida que seja a caverna na qual você resolva se fixar, muito pode acontecer em 400 anos para destruir o local, e portanto provavelmente seria mais adequado dotar sua cápsula de sistemas motores e sensoriais simples, para detectar perigo. É adequado então fazer um grande robô e colocar a cápsula dentro dele, com uma programação adequada para que ele encontre a energia que precise para manter-se longe do perigo.

Na hora de programar o robô, sem dúvida a melhor maneira de desenhá-lo é procurar, no *vasto* espaço de design dos robôs possíveis, algum que tenha grande probabilidade de lhe preservar até 2401. Ele estará equipado com objetivos, tarefas, necessidades, estratégias de solução de problemas etc., tudo para garantir seu conforto na viagem no tempo. O robô estaria cometendo *erros* se incapacitasse de alguma maneira sua reanimação, isso seriam coisas *ruins*, já que a razão de ser dele é sua proteção.

Como as condições podem mudar, cabe programá-lo não apenas com visão e capacidade de mover-se, mas também de reprogramar-se conforme a necessidade ambiental o exija.

Vamos analisar passo a passo algumas das coisas que ele terá de poder fazer. Ele deverá conhecer, poder reconhecer, e procurar fontes de energia, deverá ter conhecimento sobre como ir para um território mais seguro caso isso seja possível, e tudo isso não pode ser feito em milhões de anos, tem de ser feito na velocidade das exigências da natureza, então é melhor que você não o faça calcular o maior número primo já encontrado antes de cada tarefa. Poupar energia e tempo será fundamental, seu robô não deve estar equipado com um sistema para apreciar arte russa, já que isso lhe tomaria tempo e energia. A constrição temporal é muito importante. Se ele tiver diferentes programas para resolver diferentes tipos de problemas, é adequado que os programas mais rápidos sejam aqueles que lidem com problemas que causam mais dano em menos tempo, como desviar de objetos em alta velocidade. Outros, como mudanças de território para território levemente mais seguro, podem ser mais lentos, e exigir algum tempo de pesquisa na biblioteca local e processamento dessa informação. Ele deve ser capaz não apenas de lidar com problemas, mas também de antevê-los.

Seu robô não estará sozinho, no mínimo Marvin Minsky estará por aí congelado em algum robô programado nos confins do MIT, com os mesmos objetivos. ¿E como saber quantos serão? ¿Haverá energia para todos? Parece razoável programá-lo para ser capaz de formar alianças com robôs e humanos, computando custos e benefícios. Ele poderá ter objetivos secundários e revisar seus objetivos. É provável, em muitos e muitos anos, que ele some erros o suficiente para não mais estar atendendo aos objetivos do seu criador (você): ele estará autônomo e poderá fazer coisas muito ruins para você.

Mesmo nesse grau de sofisticação, há filósofos que defendem que ele ainda não teria *intenções* de verdade, crenças de verdade, apenas quase-crenças e quase-intenções. A intencionalidade desse sistema seria uma intencionalidade derivada, e não original, intrínseca.

Mas isso nos levaria a um problema, afinal, é exatamente isso que nós somos, máquinas de preservação dos genes nos quais estamos, equipadas pela mãe natureza com desejos, vontades e noções conforme a melhor sabedoria acessível no espaço de design. Não apenas nós, mas todos os seres vivos são veículos dos genes. Então, nossa intencionalidade igualmente é derivada, da mesma maneira que a intencionalidade do robô. Genes não podem reagir adequadamente a um problema, mas podem gerar indivíduos a seu redor que fazem isso por eles³¹. Então, nossa intencionalidade é

³¹ Muito depois da escrita de *A perigosa ideia de Darwin*, houve dentro da biologia um refinamento de diversos dos conceitos trabalhados por Dennett, inclusive o conceito de fontes de nossa hereditariedade.

derivada dos nossos genes egoístas, eles são os significadores não-significados, não nós!

Esse é o alerta de Dennett e o *intuition pump* através da qual ele nos traz essa ideia.

As várias etapas de desenvolvimento desse robô nos levarão ao próximo capítulo, no qual veremos os *tipos de mentes* que existem por aí, a partir de nosso projeto de robô.

Filosoficamente, o resultado não difere tanto do modelo com genes e memes que Dennett enfatiza, mas a maior precisão biológica dos processos hereditários se desenvolveu para além do que se pode encontrar no livro dele. Àqueles que se interessarem pelos múltiplos processos hereditários que determinam o significado de animais e humanos, recomendo o livro de Jablonka & Lamb (2005).

6. Tipos de mentes

Vamos agora fazer quatro diferentes tipos de cápsulas³², cada uma mais completa que sua antecessora, e colocar 10 pessoas em cada tipo. Isto servirá para entendermos melhor as capacidades das mentes no nosso mundo. Nos basearemos em larga medida no livro *Tipos de mentes* (1996a).

De início construímos apenas um tipo de cápsula com algumas patas programadas para caminhar ininterruptamente, como fazem alguns carrinhos eletrônicos de criança. Cada uma das 10 anda numa certa velocidade e com um certo padrão de movimento.

Cápsulas 1: robôs darwinianos

Soltamos essas dez máquinas no mundo, e com o tempo provavelmente alguma intempérie vai acontecer à maioria delas. É possível que o padrão de movimentação de uma seja melhor do que o padrão de movimentação de outra, então a mais hábil terá maior probabilidade de sobreviver. Essas máquinas são *criaturas darwinianas*, elas estão sujeitas ao processo de seleção no seu nível mais bruto. Aqueles que são piores morrem. Esse é o primeiro mecanismo de aprendizado que surgiu nos seres vivos. A morte. Sim, porque a morte de 8 indivíduos, em 10 indivíduos que se reproduzem, gera um aumento de frequência de uma forma boa de se locomover. Cada indivíduo não aprendeu nada em nenhum sentido, mas a nova geração é melhor em média que a anterior, e portanto aprendeu alguma coisa no processo.

Nossas máquinas dificilmente sobreviveriam por 400 anos se fossem tão simples, e de fato, os seres vivos cuja estratégia de aprendizado é baseado na morte, as criaturas darwinianas, são seres vivos que vivem bem menos do que as demais versões que discutiremos adiante.

As bactérias podem estar muito bem adaptadas e espalhadas, mas isso se deve principalmente a manterem um ciclo reprodutivo de pouco tempo e um número muito grande de descendentes, permitindo sobrevivência no longo prazo. No início da vida, esse era o único sistema de aprendizado disponível na vizinhança dos seres vivos no espaço de design. Todo ser vivo, há 4 bilhões de anos, só possuía vizinhos no espaço de

³² Esse capítulo trabalha uma mistura de temas de inteligência artificial com temas de Dennett (1996d).

vizinhança em que a morte é uma estratégia de aprendizado, ninguém havia atingido a proximidade de outro tipo de aprender. Mas conforme os seres vivos se multiplicam e se adaptam, novos nichos são criados, e portanto novas vizinhanças, que antes pareciam ruins, passam a ser boas, e mais espaço se torna acessível a *alguém*. Então ocorreu uma grande mudança nesse processo de aprendizado. Mas estou divagando... voltemos às nossas cápsulas.

Cápsulas 2: robôs skinnerianos

Na segunda tentativa de construção de cápsulas vamos dotá-las de um sistema de locomoção, um sistema sensorial simples (pelos bem grandes nas extremidades) e uma resposta automática (qualquer excitação maior que um limiar nos pelos gera uma movimentação no sentido contrário). Digamos que cada uma tenha uma variedade de movimentos possíveis, e que sempre que um movimento funciona, ele é favorecido como estratégia na próxima vez que um pelo for excitado. Essas cápsulas são *criaturas skinnerianas*, elas funcionam como o psicólogo arcano Skinner acreditava que os pombos e as pessoas funcionassem. Comportam-se aleatoriamente de início (dentro de um espaço de projeto determinado por nós, seus programadores) e possuem processos de reforçamento que os levam a manter a estratégia de maior sucesso. É possível, embora muito pouco provável, que alguma de nossas máquinas sobrevivesse por 400 anos, talvez na eventual extinção da espécie humana ela tivesse chances um pouco maiores. Mas de qualquer maneira, são negligenciáveis.

No mundo biológico, os seres com sistemas nervosos mais simples são skinnerianos. A complexidade comportamental dos pombos e ratos supera bastante a necessária para que eles fossem classificados aqui, como se supunha no século passado, mas alguns vermes e anelídeos podem muito bem ser o tipo de criaturas que se comportam principalmente de maneira skinneriana. Como Dennett (1996a, p. 88) pontua: “O condicionamento skinneriano é uma coisa boa desde que você não seja morto por um de seus erros anteriores.”

O que ocorreu foi que *alguém* estava no lugar certo, na hora certa, e com a mutação certa para que surgisse esse novo mecanismo de aprendizado, e como ele é bastante melhor que a anterior estratégia, ou seja, ele se replica mais eficientemente que os vizinhos, essa estratégia “pegou” e surgiu toda uma série de galhos e ramos naquela região do espaço de design. Provavelmente mais de um galho evolutivo estava nas

vizinhanças daquela região do espaço de design, e diferentes indivíduos devem ter evoluído separadamente a capacidade de se direcionar àquele espaço. Se de fato essa é a história, é um caso de evolução convergente, como o caso dos olhos. A única maneira de se compreender a evolução convergente é observá-la da perspectiva adaptacionista, usando o vocabulário adaptacionista. É só com um olhar adaptacionista, em termos de quais modificações é *melhor* ou *pior* possuir, que podemos ver que é uma estratégia boa tornar-se skinneriano. O edifício dennettiano vai se encaixando aos poucos.

Cápsulas 3: robôs popperianos

Chegamos a nossa terceira tentativa, e agora equiparemos nossas máquinas com sensores mais sofisticados, capazes de detectar fótons, e portanto de obter informação sobre eventos bastante mais distantes que os de seus antecessoras experimentais. Além disso, eles são capazes de computar, dentre seus movimentos possíveis, aqueles que têm maior chance de dar certo *antes* de tentar. Se um perigo iminente se aproxima, parte de sua energia é gasta em avaliar quais das soluções com as quais já vêm equipadas têm mais chance de funcionar naquele caso particular. Seu repertório é limitado, mas ao menos existe um sistema de busca que avalia qual ferramenta usar. A sobrevivência de uma dessas criaturas ao longo de 400 anos dependeria principalmente do tamanho de seu repertório e de sua precisão e velocidade na busca de uma boa solução para problemas como “A montanha na qual tenho vivido está soltando fumaça”. ¿Haveria algum tipo de conhecimento interno pré-programado do risco de vulcões?

Se soltássemos dez delas por aí, com graus exponencialmente maiores de “conhecimento” impresso em suas arquiteturas internas, é bastante possível que alguma sobrevivesse. Essas criaturas são chamadas de *criaturas popperianas*. O filósofo Karl Popper pontuava que esse tipo de capacidade “permite que nossas hipóteses morram em nosso lugar”.

No espaço de design é bastante interessante ser uma criatura popperiana, mas o gasto de energia também é bem grande. É necessário ter algum sistema de simulação ou computação dentro de si para avaliar a melhor solução de um problema, e isso requer energia. A evolução opera em muitos nichos diversos, desde as fossas sulfurosas e muito quentes do fundo do oceano até lagos lamacentos. A evolução cria muitos nichos diversos, como a copa de uma árvore amazônica ou o intestino delgado de uma baleia

azul. Era de se esperar que dado tempo suficiente, assim como alguém venceria 40 cara ou coroas seguidos (jogando uma vez ao dia), alguém encontraria o caminho para tornar-se uma criatura popperiana.

Evidente, nada disso é antevisto, apenas uma sequência mecânica de processos vai se desenrolando, e alguém acaba caindo na vizinhança no espaço de design desta bela mansão, e ocupa-a sem demora. O mundo animal está repleto de criaturas Popperianas, dos pombos aos manatis, passando pelos gnus e condores; por todos os lados há criaturas de médio porte, com sistema nervoso desenvolvido, que passam o dia computando hipóteses pré-programadas em seu genoma, e evitando perigos. ¿Essas criaturas sabem, com a mesma precisão que Popper, que estão deixando suas hipóteses morrerem em seu lugar? Não. Os racionais da evolução não se importam muito em quem está ou não informado de algo, a não ser que isso gere mais replicação. Um alce não precisa compreender, num nível alemão de meta-análise e sofisticação, o porquê escolhe comer numa área plana e não numa área inclinada. Ele só precisa estar programado para isso.

Para produzir uma boa cápsula *popperiana*, teríamos de inserir algum modelo do mundo em nossas cápsulas. Elas precisam ter acesso à informação do mundo para poderem escolher dentre seu repertório de respostas. É necessário poder simular ao menos algumas propriedades do mundo num ambiente interno para poder fazer uma escolha melhor do que a sorte. Essas regularidades não precisam de maneira alguma estar internamente espelhando o mundo exterior, mas precisam existir isomorfismos, possibilidade de conversão etc. O modelo do mundo não precisa ser a cara do mundo, apenas ser útil *no* mundo.

Cápsulas 4: robôs gregorianos

Na nossa quarta tentativa, vamos equipar nossos robôs com uma capacidade de aprendizado a partir do ambiente, e de reformular suas estratégias possíveis para lidar com algo. Eles serão capazes de ler, aprender línguas, utilizar livros como guias para ações futuras, observar eventos, encontrar regularidades nesses eventos e encontrar caminhos através dos quais eles possam evitar eventos indesejáveis e maximizar eventos desejáveis, mesmo que suas observações não sejam de eventos que lhes ocorreram. São robôs culturais, por assim dizer. E já que já têm todas essas propriedades, ¿que tal colocarmos uma fome epistêmica, uma curiosidade?

Esses sim são bons robôs para a sobrevivência a longo prazo, máquinas com alta probabilidade de cumprir seus objetivos, e portanto de estar ainda vivas daqui a 400 anos. Esses robôs culturais possuem dentro de sua programação o necessário para aproveitarem uma descoberta fundamental da economia do sudeste asiático, a de que *copiar* um projeto é mais barato do que *criar* um projeto. A possibilidade de aprendizado cultural acelera muito a velocidade de desenvolvimento de estratégias de evitar perigos.

As *criaturas skinnerianas* são um sub-conjunto das *criaturas darwinianas*, e as *criaturas popperianas* são um subconjunto de ambas. Nossa sequência continua com as *criaturas gregorianas*³³, que são criaturas como esses nossos robôs. São criaturas que têm capacidade para melhorar sua própria adaptabilidade e velocidade de solução de problemas quando apresentados a designs anteriores criados por outros indivíduos para fazer algo. Nenhum de nós inventou o machado, mas vê-lo auxilia em muito o tempo que demoraríamos para inventar um espontaneamente. Essas capacidades podem aumentar a velocidade de encontro com vizinhanças boas, no espaço de design dos comportamentos possíveis. Copiar tecnologia de informação não apenas torna um animal mais inteligente, mas acelera o passo no qual se fica mais inteligente.

E com alta probabilidade: ¿O vencedor é? Gregory!

Evidentemente ao menos uma criatura no mundo natural é uma *criatura gregoriana*, Karl Popper. Além dessa, seguramente vários homínídeos anteriores também tinham capacidades análogas, o que indica que estavam na mesma área do espaço de design (algo esperado, já que temos ancestrais em comum muito próximos). Quanto aos animais, muita disputa ocorre a respeito de quanta transmissão de informação cultural há em diferentes culturas. Sabe-se que alguns bandos, mas não todos, de macacos-japoneses lavam sua comida. Em algumas áreas, chimpanzés usam varinhas para capturar cupins. Alguns orangotangos (dificilmente coincidência, aqueles que passam mais tempo em grupos) usam cascas de árvore para retirar a parte comestível de uma fruta. Todos esses comportamentos regionais, no qual se igualam todos os indivíduos de uma região, mas não de outra sem contato informacional/cultural com esta, apontam para um grau de *gregorianidade* entre alguns outros primatas.

³³ Richard Gregory é um psicólogo que estuda os movimentos bons no espaço de design, permitidos pela informação cultural. Daí o nome.

Também primos muito distantes se enquadram entre os animais com tradições culturais. Golfinhos mantêm junto a humanos uma tradição, há mais de cem anos, de seguir seus barcos e avisá-los (saltando) da quantidade de peixes disponíveis, facilitando o processo de pesca. Não bastasse essa atividade incrível, ela ocorreu *duas vezes separadamente*, no sul do Brasil e na ilha de Sumatra, na Ásia. Baleias cachalote tem um espalhamento de seus cânticos que muito lembra o espalhamento a partir de Paris e Milão da moda fashion no mundo (Cantor & Whitehead, 2013). O epicentro é o leste australiano, e de lá, ao longo dos anos subsequentes, se espalham as canções em direção ao Havaí e além. Assim como a fronteira entre chimpanzés e o ancestral comum do chimpanzé com o humano não é uma fronteira clara, mas uma sucessão de infindáveis mutações e indivíduos mais ou menos sortudos, da mesma maneira não existe uma fronteira clara entre os diferentes tipos de criaturas. Das *criaturas darwinianas*, no espaço de design, até nós (atualmente as criaturas mais *gregorianas*), existe uma infinidade de caminhos possíveis, muitos dos quais ainda possuem seres vivos até hoje. Não há fronteiras muito claras no mundo da biologia porque, assim como a pele indiana e a japonesa parecem um degradê, tudo que se distancia muito das bases nitrogenadas perde a cara de discretização que tinha antes. No mundo da biologia há poucos “tudo ou nada”, e as bases nitrogenadas que constituem os genes são um deles, a ativação de um neurônio em particular é outro, e alguns comportamentos “imprintados”³⁴ (darwiniana ou skinnerianamente) são outros. Mas, via de regra, a vida no nosso planeta parece um evento bem analógico e contínuo.

É interessante ver a alternância entre processos analógicos e digitais na natureza, embora Dennett não fale muito sobre isso em sua obra; a evolução constantemente alterna esses dois tipos de processamento de informação, para possibilitar novas camadas de computação no aparato sintático do mundo. Bases nitrogenadas digitais codificam órgãos analógicos, incluindo um cérebro analógico cujos neurônios digitais agem em coalizão gerando uma computação analógica que por vezes se evidencia em sentenças compostas por palavras digitais compostas por letras digitais. Palavras que em sua composição permitem reflexões de qualquer nível de profundidade, aproximando-se novamente do degradê típico de processos analógicos. Um caso particular de

³⁴ *Comportamentos imprintados* (do inglês *imprint*) são comportamentos que possuem algum tipo de gatilho simples e são determinados geneticamente, antes de relacionar-se com o mundo. Por exemplo, patinhos, quando nascem, seguem qualquer objeto que se mova e que esteja em seu campo de visão. Bebês humanos, por um curto tempo, respondem “entusiasticamente” a um *smile* ☺, *mesmo que ele tenha 6 bolinhas no lugar dos olhos*.

redigitalização importante é o surgimento da nossa linguagem, com suas letras, sílabas e palavras, unidades discretas populando um mundo analógico. Uma das mais grandiosas ferramentas que uma *criatura gregoriana* pode acessar é aquilo que Gregory chama de ferramentas mentais, as palavras.

Palavras permitem processos sofisticados de testes de hipótese e análises minuciosas. Linguagens sofisticadas como a humana são recursivas: elas permitem que se coloque uma classe de assuntos dentro de outra, dentro de outra indefinidamente. Uma sentença inteira pode ser sujeito de outra sentença, e essa super-sentença pode ser o sujeito de ainda uma terceira, e assim por diante. Recursividade, em termos computacionais, é uma carta branca de acesso profundo, é uma propriedade organizacional que equivale a acessar níveis de sofisticação indeterminados. Tanto quanto alguém quiser se aprofundar, pode-se aprofundar. Evidente que por muitos outros fatores, desde o tempo de memória de curto prazo até nossas adaptações específicas, não podemos analisar todo e qualquer objeto com um nível qualquer de sofisticação. Mas sem dúvida o fazemos muito melhor do que um animal inteligente sem linguagem, como o elefante.

Então concordaremos que é uma ótima ideia fazer cápsulas do quarto tipo, que elas serão as melhores sobreviventes, e que nossa tarefa está garantida. Não exatamente... Talvez tenhamos sofisticado *demais* nossas criaturas, podemos ter criado *concorrência*.

7. Os outros: a memética

Nota ao simulador: Embora Dennett haja escrito um bom tanto sobre o assunto desse capítulo, e seja um tema que bastante lhe interessa, existem diversas ideias interessantes na vizinhança daquelas que ele advogou *explicitamente* em livros e artigos que são de bastante interesse. Saindo um pouco da perspectiva de observador externo da mente dennettiana, pretendo aqui tratar mescladamente as ideias que ele defende e as ideias que são *compatíveis* com sua visão sobre o tema. Dentre essas ideias haverá ideias presentes na literatura técnica sobre o assunto, e também algumas ideias minhas. O presente texto foi escrito como parte de um projeto de pesquisa que focava tanto a filosofia de Dennett quanto a *memética*, e seria um desperdício de informação não me utilizar de algumas páginas para expressar interrelações interessantes entre essa área da memética e o pensamento dennettiano. Outra razão para fazê-lo é a honestidade intelectual de admitir que não há fronteiras bastantes claras para mim entre aquilo que foi explicitamente dito por Dennett e que *faria sentido* de acordo com Dennett a respeito de memes. É justamente esse tipo de sensação, de não conseguir mais delimitar fronteiras entre as ideias de Dennett, e as compatíveis com elas, que eu pretendo passar através desse livro.

Ancient FM

Enquanto escrevo essas palavras, estou ouvindo uma rádio que encontrei na internet hoje que apenas toca música medieval. Especificamente, li em algum lugar que toca apenas música do ano 1100 ao ano 1600. Pontuo que “li em algum lugar” porque isso será parte importante da discussão desse capítulo. Iremos do ano 1100 até hoje, e de hoje até o ano 2401, e no caminho compreenderemos como é possível que nossas criaturas gregorianas possam ter criado sua própria concorrência e se tornado sofisticadas demais.

Suponha que você esteja interessado em música medieval, *realmente interessado*, e está disposto a fazer uma pesquisa a respeito. ¿Como começar? Existe uma infinidade de maneiras; seu espaço de busca é imenso e as possibilidades em aberto no espaço de design de planos de ação são muitas. Diversos caminhos levariam você, de seu projeto atual, “uma vaga noção de o que pretende fazer”, à satisfação desse anseio.

Uma maneira de começar seria procurando no Google.³⁵ Mas a maioria das pessoas não consideraria os primeiros sites uma boa fonte de informação. Mesmo não sendo uma boa fonte de informação, eles podem ser uma boa *fonte de fontes* de informação. A internet é um veículo interessante, entre outras coisas pela *transponibilidade das hierarquias* do que nela se encontra. Quase sempre o conhecimento de mais baixo nível sobre um fenômeno está na internet. Conhecimento impreciso, mal memorizado, misturado com outras informações e sobre temas que apenas tangenciam seu tema. Provavelmente há por aí milhares de entradas de blogs em que alguém comenta algo como “Música de igreja é algo super antiquado, dá vontade de dormir!”, um exemplo perfeito de o que *não* queremos encontrar ao fazer uma pesquisa séria sobre música medieval. Mas nem tudo na internet é assim, e a empresa que mais tem atraído a atenção do mundo, o Google, ganha seu dinheiro basicamente ajudando-nos a “refinar” nosso gosto, e nossas procuras. ¿Mas refinar até que ponto? Não queremos um tratado sobre a música medieval escrito em francês do século XVII, queremos algo intermediário, talvez um artigo sobre isso. Se pensarmos no grau de refinamento de um conhecimento como uma pirâmide, onde no topo situa-se o mais arcano conhecimento acadêmico, e na base o mais frugal e leviano pseudo-conhecimento de um blog adolescente, podemos apreciar claramente a vantagem da internet. Podemos nos movimentar entre os níveis muito rapidamente: transparência de hierarquias. Talvez os “andares” extremos da pirâmide não estarão acessíveis, mas existem várias ordens hierárquicas que estão, e podemos atravessá-las facilmente (o autor suspeita que essa seja a razão pela qual tanto se defende que a internet seja um espaço mais democrático por aqueles que defendem a democracia, mais anárquico pelos anarquistas, e mais socialista pelos socialistas).

Após um pouco de tempo passeando por categorias intermediárias, encontramos a página que declara que a rádio que estou ouvindo toca apenas 1100 a 1600. ¿Como saber se podemos confiar nessa informação? De início temos razões para suspeitar de uma série de erros. ¿Teriam os organizadores da rádio um *sistema de seleção* para bloquear músicas de 1037? ¿E de 1650? ¿Existe algum sistema de detecção capaz de fazer isso? ¿Como saber se alguém errou a digitação e na verdade são músicas apenas de 1200 em diante? ¿Existe uma notação musical que permaneceu todo esse tempo?

³⁵ Recentemente encontrei no Google um site que iniciava citando o Google, e fazia uma menção à possibilidade de que no futuro alguém que estivesse lendo aquilo poderia não saber o que é o Google. Meu pensamento é em geral bastante científico e um tanto quanto cético, não acredito que tal tempo chegará.

¿Temos certeza que esse era o tempo das músicas? Todas essas são perguntas legítimas, e ao ligarmos a rádio encontraríamos mais algumas ¿Se as músicas são cantadas em inglês moderno, serão elas as *mesmas* músicas? ¿Quanta liberdade poética o tradutor considerou aceitável?

Provavelmente, para um interesse mais acadêmico, já temos perguntas o suficiente para perceber que nosso erro foi parar numa hierarquia muito baixa. Mas talvez nossa curiosidade tenha sido despertada por essas perguntas, que agora nos interessam mais do que a original (¿Qual era mesmo?). Sem dúvida nosso uso de palavras refinou nosso interesse, não necessariamente para algo melhor, mas para algo mais preciso. Um desejo vago de busca, expresso por mais e mais palavras, torna-se um desejo mais preciso.

¿Mas afinal, é legítimo esse objeto?

Todas essas perguntas são perguntas sobre autenticidade: queremos saber se as músicas realmente são o que achamos que são. Mas agora voltamos a um interesse mais filosófico e vamos nos meta-perguntar a respeito das perguntas sobre autenticidade.

¿Porque suspeitamos de alguém que nos garanta que uma música foi composta entre 1100 e 1600? Porque sabemos que, se o que ele diz for verdade, existiriam oito séculos de separação entre a música e o momento no qual estamos recebendo essa informação, e que quanto mais tempo houver entre uma informação e seu assunto, menor a chance de que ela esteja correta. Pela mesma razão, jamais desconfiaríamos de alguém que nos dissesse que horas são: a informação é muito nova para estar manchada. A informação se torna menos confiável ao longo do tempo porque existe sempre uma possibilidade de interferência no meio que a carrega. Uma música que foi transcrita em papel encontraria muitos problemas ao longo de oito séculos para chegar até nós. Papel não é uma boa cápsula do tempo: manchas, traças, incêndios, ar, pressão, tensão, microorganismos, músicos com uma tendência a “revisar” obras do passado a seu gosto, cães. Nomear todas as tragédias que poderiam acometer um conjunto de traços numa folha, ao redor de uma fogueira, amedrontaria crianças e nos faria repensar nossa sorte relativa enquanto seres do universo. E não só a folha está ameaçada: mesmo que a música estivesse inscrita em pedra, toda a população que conhecia aquela notação poderia morrer, ou abandoná-la.

Uma música, como uma palavra, não é uma inscrição específica, mas sim uma estrutura abstrata, que pode ser representada de diversas maneiras em diferentes meios. O exercício de pensar sobre a cápsula do tempo é muito interessante, quando estamos olhando para frente, e só se adequa ao estudo de músicas a partir de suas origens. Mas nós não estamos lá, estamos aqui, no século XXI. Nosso problema não é preservar nada, nosso problema é verificar autenticidade, e o exercício da cápsula do tempo só pode nos ajudar até um certo ponto. Temos de virar-nos para trás: a maioria das músicas que já existiram não existem hoje. *Mas essa existe*, e portanto devemos pensá-la de maneira diferente.

São poucos os casos em que temos uma intuição boa para pensar a evolução. A intuição está preocupada com perigos imediatos, e não pode se dar ao luxo de observar detalhadamente por anos o que quer que seja. “Intuição” afinal é o nome que damos para nossas cognições mais automáticas. Mas quando pensamos sobre autenticidade de objetos culturais, nos tornamos ótimos detetives.

Você, esse detector de trapaça

Faça o teste, nomeie algumas razões pelas quais a música que estou ouvindo agora possa não ser autêntica. Se só pensou em duas, pense um pouco mais; se já passou de cinco, pise no freio, é apenas um teste!

- Alguém pode *deliberadamente* ter inventado essa música.
- Alguém pode ter *deliberadamente alterado* a música.
- Um copista pode ter desejado imprimir sua marca na história, mudando algum trecho da música ou simplesmente substituindo-a por uma música similar de sua autoria e mantendo o nome.
- Alguém pode ter mudado o sistema de notação.
- A música pode ter permanecido apenas como parte da cultura popular, sem nenhuma cópia escrita por muito tempo, e por muitas vezes alguém preencheu um esquecimento como lhe pareceu melhor no momento.

Essas são as hipóteses que têm maior chance de ter aparecido em sua mente. Se você pensou em alguma delas, estava usando a *perspectiva intencional*. Somos muito bons nisso, porque é importante detectar trapaças se você é um animal que vive em grandes grupos. É tão natural conceber que pessoas estão trapaceando que

frequentemente achamos que alguém está com “um ar de trapaça”. Descobrir amigos ou inimigos é uma questão de vida ou morte entre os índios, aborígenes, !Kung e outras pequenas sociedades.

Muitas vezes, não existe resposta para *¿quem é o culpado?*

No entanto, além da alteração intencional da música, também há outras possibilidades, menos acessíveis de imediato:

- Um copista pode ter errado a cópia.
- O sistema de notação pode ter mudado sem uma interferência clara de nenhum indivíduo, gradualmente.
- A música pode ter estado na cultura popular, e alguém não apenas esqueceu um pedaço dela, mas lembrou de outra sequência de notas no lugar!

Agora estamos na *perspectiva do design*: algum pequeno erro no design de algum copiadador (humano ou não) pode ter sido a causa da não autenticidade da peça musical.

Finalmente chegamos à *perspectiva física*:

- Caiu uma gota de tinta numa cópia ancestral que foi interpretada como uma nota.
- Um pássaro emitiu um som que um ouvinte inseriu na música, achando que era parte dela.
- Um ouvinte estava muito longe do músico e, por uma série de coincidências de reflexão, movimento e posição do ouvido, ele de fato ouviu uma nota diferente, e passou essa informação adiante.

Todas essas hipóteses são possíveis, a maioria delas é pensável também.

Acabamos de descrever uma série de processos de *mutação* de elementos culturais. Ora, músicas sofrem mutação, palavras também. Se o leitor está começando a aprender a pensar como Dennett, agora é uma boa hora para se fazer a pergunta: ¿O que meu sistema de informação sobre Dennett sugere que ele se perguntaria nesse ponto?

Algumas hipóteses interessantes:

¿Será que esses elementos possuem as outras propriedades necessárias para a evolução? ¿Teremos encontrado outras estruturas que caminham pelo espaço de design? ¿Se um pássaro não precisa *saber* a função de seu comportamento de fingir que está

com a asa quebrada, para que esse comportamento exista, alguém precisa saber alguma coisa para que uma música exista? ¿Seria boa parte das músicas uma *idiotice completa*?

Por mais tentados que estejamos a examinar primeiramente a última questão, a primeira é mais importante, e dela dependem as demais.

Sim, músicas possuem hereditariedade e também possuem um meio ambiente limitado (as mentes humanas e papéis em 1200). Temos o necessário para que nosso algoritmo evolutivo entre em ação.

¿Isso significa que todas as músicas hão de ter evoluído? Essa questão pode ser postergada por ora. Sabemos que *algumas* músicas evoluíram, e que elas estavam explorando o espaço de design que lhes é acessível. E não por acaso, temos estudos de como um gênero musical influencia outro, sobre como uma música mudou através do tempo etc. Esse é um momento importante: encontramos *outros replicadores* que não estão codificados em DNA ou RNA, *e que são parte do nosso mundo*.

Não, simplesmente não existe!

Existe uma tendência na atualidade de chamar de *natural* aquilo que não envolve interferência humana, e *artificial* o que envolve. Mas não há como traçar claramente a fronteira entre uma coisa e outra, pois humanos são parte da natureza, bem como suas atividades. Computadores e bisontes são coisas com design, e como tais, demandam uma explicação; se as leis da física tomadas de maneira simplória não dão conta, podemos nos servir do algoritmo evolutivo. Essa linha divisória entre artificial e natural nunca existiu, não mais de um parágrafo deveria ser usado para dizer isto.

A evolução não vai parar por nada!

Um pouco mais de tempo deve ser dedicado a explicar por que devemos levar a sério a ideia de que o algoritmo evolutivo também está presente nas coisas artificiais. Sim, a evolução é capaz de ultrapassar essa barreira inexistente, uma habilidade incrível.

Quando Richard Dawkins cunhou o termo “meme”, em seu livro *O gene egoísta* (1976), ele pretendia imaginar uma unidade cultural análoga ao gene, um replicador que agisse por conta própria. Mal sabia ele que o ideia do “meme” se multiplicaria descontroladamente, gerando ondas de revolta da parte de antropólogos, sociólogos e até mesmo biólogos, ao mesmo tempo que criando uma legião de seguidores, livros,

artigos e até uma revista própria. O meme “meme” ganhou vida própria, e apesar das constantes batalhas entre aqueles que defendem e os que demonizam os memes, há ao menos um consenso: a ideia tem força e precisa ser examinada minuciosamente.

Um meme pode ser uma frase, uma ideologia, um jogo, ou até mesmo uma equação. São ideias que podem ser transmitidas. A perspectiva memética é justamente olhar esse tipo de transmissão *do ponto de vista do meme*. Algumas perguntas que se podem fazer para observar a multiplicação de ideias dessa perspectiva são: ¿O faz com que pessoas que nunca estudaram física conheçam a equação $E=mc^2$? ¿Por que todos lembram como morreu a princesa Diana, e quase ninguém lembra a razão da morte de Frank Sinatra, pouco mais recente? ¿O que faz com que não consigamos esquecer aquela música infernal que detestamos, e a faz ressurgir na cabeça nos piores momentos? Todas essas questões podem ser vistas pensando da perspectiva do meme. A questão sobre a morte de Diana se tornaria então: ¿Que propriedades da ideia da morte de Diana fazem com que ela se reproduza com mais fidelidade e mais durabilidade do que a morte de Sinatra?

No caso do acidente de Diana, por exemplo, a ideia permanece em nossa cabeça porque as memórias de tragédias, de mortes inesperadas, são mais fortes do que as mortes ligadas a envelhecimento e causas naturais. Ou seja, um meme que seja a respeito de uma morte trágica tem maior chance de persistir temporalmente na mente de seus portadores. OK, eu admito que não fiz estudos duplamente cegos e meta-estudos históricos de análise comparativa para chegar a essa afirmação, ou seja, é um chute, mas concedam que é um chute bem informado.

¿Se você fosse um meme, o que você faria?

Olhando da perspectiva memética, podemos pensar que nossos interesses psicológicos naturais funcionam como pressões seletivas no espaço de design. O espaço de design dos memes funciona de maneira parecida com o espaço de design genotípico, ocorrendo mudanças, transformações, hereditariedade, cópias, vizinhanças, atratores, repulsores etc. Quando um meme está na vizinhança da sexualidade, é uma ótima estratégia tornar-se mais sexual, porque adoramos sexo e falar de sexo. A julgar pela frequência, adoramos mais falar do que fazer! Tente imaginar o racional evolutivo para isso...

Memes chutam de volta e sem eles ninguém devolve algumas bolas

Uma questão que filósofos (particularmente os reducionistas) se alegram muito ao poder usar contra seus oponentes é: ¿Mas isso é necessário para explicar alguma coisa? ¿é condição *sine qua non* para uma visão de mundo mais completa e científica? Já sabemos que o design de toda a vida animal que conhecemos parece ter sido produto da replicação com variação cega de DNA e RNA. ¿Precisamos explicar algum design do mundo com memes, ou podemos nos contentar com genes e simplesmente assumir que a cultura também evolui para o benefício dos nossos genes, em última instância? Parece, com isso, que o ônus da prova foi jogado para o nosso lado. É hora de averiguar se de fato os memes são causa última de algum design, ou se apenas são um derivado dos genes, no mesmo sentido que a resposta automática de fuga de uma lesma marinha é derivada dos genes.

Partamos para mais um *intuition pump*, continuemos a bombardear nossas mentes com analogias filosóficas interessantes que possam ajudar a pensar ideias relevantes. Dessa vez, no entanto, retiraremos nosso intuition pump da realidade, contando uma intrigante história real.

Imagine que você está caminhando pela floresta e percebe uma formiga se comportando de uma maneira muito particular. Ela sobe um pedaço de grama, até o topo, e cai, e mais uma vez sobe e cai, e faz isso sequencialmente, indeterminadamente, sem aparentemente nenhum objetivo. ¿O que estaria fazendo essa formiga? ¿O que será que ela objetiva (ou os genes dela objetivam) com esse comportamento aparentemente inútil? Um exame mais minucioso, feito por um biólogo especializado, mostrará a mais profunda resposta: absolutamente nada.

Não existe nada ali que a formiga possa usar em benefício próprio. Em verdade, o que ocorre é que o cérebro da formiga foi, literalmente, raptado. Um parasita, um verme achatado parasítico (*Dicrocoelium dendriticum*), entra no corpo da formiga, e vai até o cérebro dela, gerando, através de girar um ou dois manípulos de seu simplificado cérebro, um comando que faz com que ela se comporte dessa maneira. Ele faz isso porque parte de seu ciclo de vida requer ser comido por uma vaca ou uma ovelha. É infrutífero procurar um benefício para a formiga aí, porque é o código genético do parasita que está no comando.

Mas é interessante ver como o comportamento está ocorrendo *na formiga*. Mudar o comportamento de um indivíduo é um movimento adequado em algumas

vizinhanças do espaço de design, essa estratégia foi inventada independentemente por outros seres vivos, sequestro cerebral é algo que ocorre! A natureza não é exatamente boazinha, como diria Bertrand Russell (1957, p. 92): “O universo pode ter um propósito, mas não há nada no nosso conhecimento que indique que, se ele tiver, esse propósito tenha qualquer semelhança com os nossos”.

Pensemos agora nas seguintes ideias: ‘celibato’, ‘vontade de se matar por sua religião’, ‘comer a carne dos mortos’ (um costume recém-extinto de uma comunidade tribal). ¿O que uma pessoa com essas ideias tem na cabeça? Em outras palavras, ¿quem está saindo beneficiado por elas?

Música ruim existe, mas é possível que de algum jeito ela beneficie nossos genes. Ideias repulsivas para nós também, mesmo o anti-semitismo ou o racismo podem ser entendidos como detectores de tribos inimigas³⁶, mesmo sendo ideias consideradas maléficas.

Mas o celibato, o desejo de morrer por uma religião ou alimentar-se dos mortos são claramente ruins para a reprodução de nossos genes, e nos parecem más ideias (ao menos para a maioria de nós). Aliás, nos parecem más ideias *justamente por isso*.

Mas não possuem uma explicação evolutiva, pelo contrário, elas *desafiam tanto o nosso senso estético e moral quanto a seleção natural*. Essas ideias são parasitas, não interessa de que perspectiva olhemos para elas, elas desafiam nosso senso de autonomia por todos os lados. É esse tipo de meme que, na visão dos memeticistas, garante a existência dos memes. A razão é simples: se não houvesse memes, não haveria como explicar o surgimento desses comportamentos, não haveria beneficiário. Estaríamos presos como um biólogo que ainda não encontrou o parasita na formiga, tentando entender a questão da perspectiva da formiga, o que nunca seria resolvido. Por essa razão, os defensores da memética como ciência afirmam que seu objeto de estudo está lá, é necessário e tem poder explicativo.

³⁶ Não porque hoje sejamos inimigos desses grupos, mas porque, no ambiente no qual ocorreu a evolução dos módulos mentais que lidam com outras tribos, surgiu a capacidade de detectar amigos e inimigos, e supor que todos os membros do grupo de um inimigo é inimigo, e todos os membros do grupo de um amigo é amigo. Detectores naturais explicam por que ser preconceituoso é em geral um processo passivo, e não ser preconceituoso requer educação e resistência a impulsos internos visíveis em qualquer criança. A natureza realmente não é boazinha.

¿Cui bono?

¿Quem se beneficia com isso? ¿Cui bono? Essa questão não é apenas importante no direito, mas também na economia. E no espaço de design dos replicadores, é uma ótima ideia ser econômico; comprar na baixa e vender na alta é o mandamento não só da bolsa de valores, mas da sobrevivência a longo prazo. *Alguém* está saindo beneficiado, do contrário, ninguém estaria *investindo* design. A natureza faz poucas coisas que exibem design e não têm um beneficiário. Bolhas têm um pouco de design, planetas também, pérolas talvez sejam um caso mais interessante, mas no geral, se algo exhibe design, existe um designer (Não! Alarme! Não acreditem nisso, esse é o raciocínio dos criacionistas.) Se algo exhibe design, existe algum *beneficiário*, e pode apostar que ele será um replicador. A ciência vem apostando cada vez mais nisso desde 1860, e nunca deu zebra.

Pois bem, estamos diante de design ao pensarmos nessas ideias, elas são portanto um indicativo forte de que haja um replicador por aí que não é DNA. O celibato não se multiplica pelo bem de ninguém, ele se multiplica porque *consegue*. A AIDS se multiplica porque *consegue*. Com efeito, o mote dos memes e genes poderia muito bem ser aquele da campanha presidencial de Barack Obama em 2008: “Yes, we can!” (Sim, nós podemos!).

Temos uma razão epistêmica para acreditar que os memes tenham poder explicativo, e isso implica que o projeto de pesquisa da memética está justificado. De agora em diante, faremos ajustes, diferenciando o funcionamento da evolução de replicadores culturais e replicadores biológicos, delimitando as fronteiras de até onde essa ciência cultural se justifica.

Voltemos ao nosso teste de detecção de autenticidade, para considerarmos uma possibilidade omitida até agora. Nos idos de 1420, tornou-se moda numa certa comunidade um assovio muito parecido com uma parte da música, e eventualmente a parte da música que se parecia com o assovio tornou-se igual a ele. Ninguém percebeu.

¿Quem se beneficiou aqui? A sequência de notas do assovio. Por alguma razão aquela década viu um grande verdejar na grama das casas que ali se encontravam no espaço de design dos memes. A música, que morava numa casa próxima, viu que a grama parecia mais verde do outro lado, e mudou-se para aquela região. Não, ninguém se mudou conscientemente, eu estou falando apenas de padrões de informação abstrata se reproduzindo no meio de uma cultura humana. Nossos detectores de trapaça não

costumam raciocinar nesse nível.

O que chamamos de “moda” são pressões seletivas no mundo dos memes; se algo se torna moda, torna-se bastante produtivo estar por perto. Ninguém precisa apreciar uma razão para que ela esteja lá. Contrariamente à visão popular, a moda não existe para ser vista, a moda existe porque consegue.

Paremos para pensar nas ideias que existem por aí como replicadores, um exercício mental interessante. Mais interessante ainda é ir e voltar: adotamos a perspectiva dos memes, unificamos esses conceitos com outros e voltamos a pensar em memes da *nossa perspectiva*. Mas pensá-los da *nossa perspectiva enquanto* memes. É pouco controverso e bastante razoável por exemplo pensar que os seguintes memes são bons, olhando de nossa perspectiva: ‘simpatia’, ‘pedir por favor’, ‘música’, ‘paz’, ‘democracia’, ‘escrita’, ‘multiplicação’, ‘não faças a outros o que não queres que te façam’, ‘escolaridade’. Esses memes, além de existirem como replicadores, multiplicando-se pelas mentes humanas simplesmente por terem a propriedade de se reproduzir, isto é, por serem passáveis de pessoa para pessoa, também são *bons para nós*. Queremos que essas ideias continuem aí, se espalhem e não se percam no tempo.

Outros memes já são mais controversos: ‘comerciais na TV’, ‘ativismo anti-aborto’, ‘os fins justificam os meios’, ‘ativismo pró-aborto’, ‘editoriais de jornal’, ‘censura de internet’, ‘Jesus te ama’, ‘Quem não está comigo está contra mim’, ‘propaganda política’, ‘prisão perpétua’. Não são memes considerados genericamente bons por toda a população, ou genericamente ruins. São memes controversos, e grande parte das convicções que defendemos com mais força estão provavelmente nesse grupo. Os memes pró e contra o aborto, por exemplo, têm uma propriedade chamada ‘adaptabilidade por diferencial de frequência’. Isso significa que o meme é mais apto (fica mais “forte” na cabeça, é pronunciado mais vezes, ocorre em mais mentes) quanto mais equilibrada for a frequência entre ele e seu antagonista no ambiente. Se todas as pessoas fossem pró-aborto, dificilmente o tema seria discutido, mas, *justamente por ser controverso*, o meme pró-aborto continua ressurgindo na mente das pessoas. Hoje em dia não se discute a respeito da terra ser redonda ou não, ou seja, esse meme não é muito citado ou pensado. Nos tempos em que essa questão era profundamente controversa e havia bons argumentos de ambas as partes, esse meme tinha muito mais chances de se reproduzir. O mesmo fenômeno ocorrerá, Dennett dirá em sua autobiografia, com o ateísmo. Ele tem certeza que o termo se tornará tão obsoleto quanto terra-redondismo.

Alguns memes, embora considerados imensamente daninhos, são bastante

difíceis de eliminar: ‘anti-semitismo’, ‘neo-nazismo’, ‘guerra’, ‘humilhação dos mais fracos’, ‘repressão da liberdade’, ‘vírus de computador’.

Quando buscamos uma explicação no nível cultural de por que um meme foi selecionado, geralmente pensamos em algo como “porque ele é verdadeiro”, ou “porque ele é belo”, ou ainda por alguma utilidade prática que ele possa ter. É hora de lutar contra essa intuição. Memes podem ser bons, ruins ou neutros para nós, exatamente como, dentre as trilhões de bactérias no nosso corpo, há aquelas que nos estão ajudando, atrapalhando, e aquelas que só estão de passagem, sem nos incomodar.

Não gostamos dos memes, ou melhor, a maioria de nós não gosta do meme “meme”. A indigestão não é apenas um produto do senso comum, é sentida também por sociólogos, antropólogos e intelectualoides pós-modernistas, e não sem razão, pois a principal característica do conceito de meme é ser definido como algo que está agindo *em benefício próprio*. Ora, se age em benefício próprio, se ele tem autonomia, ¿onde fica a minha autonomia, a minha individualidade? O incômodo gerado pela ideia do meme é semelhante ao gerado pela ideia do gene egoísta: é um golpe no nosso antropocentrismo. ¿Quem está disposto de bom grado a aceitar que é apenas um veículo de genes? A ideia é estranha, e não parece fazer sentido. Da mesma maneira, quando concedemos autonomia aos memes, estamos lhe dando o estatuto de entidades autônomas, os donos da bola, e isso significa que estamos tirando a autonomia de nós, pessoas. É famosa a descrição das feridas narcísicas de Freud. Copérnico nos tirou do centro do universo, Darwin nos deu primos peludos, e ele Freud, nos tirou a responsabilidade pela totalidade da nossa mente, ao postular a existência de um inconsciente. A memética pode ser pensada como a quarta ferida narcísica, pois ela nos tira do controle até mesmo de alguns processos mentais conscientes. As ideias estão se replicando em benefício próprio, e nós somos apenas os veículos de transmissão no processo, os fios condutores que garantem a reprodução e evolução delas.

E é por isso que criar cápsulas robôs *gregorianas* talvez não seja uma ideia tão brilhante quanto pareça a primeira vista. ¿Não era óbvio?

Os significadores não-significados são os genes e memes, não nós!

Provavelmente é uma ótima estratégia para um meme estar numa frase separada entre linhas num livro. Como eu havia *falsamente* escrito que são apenas os genes que produzem significado anteriormente, achei que esse belo meme acima merecia destaque.

Genes *não têm* fronteiras definidas (é, eu não deixei isso explícito até aqui, meu objetivo era dar destaque para o caráter discreto do código genético), mas um gene é uma estrutura grande, que pode determinar várias características ao mesmo tempo e interagir com outros genes. Com efeito, essa é a regra e não a exceção em seres vivos grandes e complicados como nós, os salmões e as macieiras. Mendel, um personagem fundamental na Síntese Darwinista Moderna, ficou famoso justamente porque conseguiu encontrar exceções, ervilhas cuja determinação de algumas características era bastante clara. Através de cruzamento de algumas gerações encontrou a existência de unidades discretas subjacentes à hereditariedade. Não que todos os genes das ervilhas sejam facilmente traduzíveis em características de ervilhas adultas, mas pelo menos alguns o eram, e isso foi suficiente para fazer as tabelas de proporções matemáticas e concluir algo interessante.

Genes e memes: ¿parentes distantes ou próximos? Trapolins e coleiras da natureza

Ninguém deseja abandonar o termo “gene”, logo ausência de fronteiras claramente definidas não pode ser uma razão para abandonar o termo “meme” e seus usos, diz Dennett.

Memes se replicam horizontalmente e verticalmente, genes apenas verticalmente³⁷ (de pais para filhos), e isso confere uma vantagem de velocidade grande para os memes, que faz com que possam, através da engenharia que é a biologia, encontrar espaços no espaço de design que não são benéficos a nenhum gene, e ainda assim permanecer por lá. Vários autores dizem que os memes estão no cabresto dos genes, já que o espaço onde se replicam é determinado pelos genes, ou seja, no fim das contas, os genes estão no controle. Dennett faz uma analogia interessante que é contrária a esse ponto. Os genes na verdade funcionam como um trampolim, permitindo que a replicação se desenvolva e acelere e chegue a espaços antes não exploráveis no espaço de design. Simpatizo com ambas as analogias, o que exige uma explicação. É evidente que como forças replicadoras disputando energia, haverá algum tipo de disputa

³⁷ Em organismos pluricelulares ao menos.

entre memes e genes... ao mesmo tempo, pode haver coevolução, e não há dúvida de que os genes possibilitam grandes oportunidades aos memes, e os propulsionam. ¿Que tal um trampolim e uma coleira? Os genes permitem um avanço considerável, aumentam a velocidade, dão tudo que um adolescente poderia desejar, mas também limites, e existem níveis de malefício genético que nem os memes são capazes de produzir sem se matar no processo. Comer a carne dos próprios mortos é uma exposição muito grave a doenças, e talvez esse seja um caso em que os memes pediram demais. O canibalismo, o suicídio em massa, e o sacrifício religioso por outro lado continuam por aí, firmes e fortes, porque se reproduzem rápido demais para serem destruídos. Seus portadores morrem apenas depois de tê-los espalhado.

Até aqui meu simulador de Dennett estava controlando e restringindo meu espaço de escrita, ou melhor, meu espaço de design de escrita. Agora estou ativando apenas o compatibilizador de Dennett, e introduzirei algumas ideias pelas quais ele claramente não pode ser responsabilizado, que estão principalmente representadas em meu artigo “Why truthful memes have a bigger survival chance” (“Por que os memes verdadeiros tem maior chance de sobreviver”) (Caleiro, 2008).

O quarto golpe

Um reducionista perspicaz, que esteja observando todos os movimentos até agora tomados, provavelmente estaria com alguns incômodos persistentes que foram se ampliando até nosso ponto atual. ¿Se somos produtos da replicação de genes e memes, onde está nosso livre-arbítrio? ¿Se os memes se reproduzem para seu próprio benefício, ou seja, se eles se reproduzem porque conseguem, como podemos saber que o que achamos saber sobre o mundo é de fato conhecimento e não apenas um conjunto bem sucedido de memes?

Há muitas outras questões, mas essas duas já são duras o suficiente. Sobre o livre-arbítrio falaremos no capítulo 9, sobre como recuperar conceitos antigos de seus vazamentos e putrefações internos. Estamos então diante do problema do conhecimento, ou seja, especificamente nosso problema é justificar o conhecimento humano, dado o paradigma reducionista segundo o qual todo o design sobre a Terra (e fora dela) se deve à ação de replicadores interagindo e sendo meticulosamente filtrados pela seleção natural.

Sim, se deve

Existem muitas maneiras de classificar a biblioteca de memes do mundo. A mais simples, dadas nossas classificações anteriores, seria colocar ao longo da primeira dimensão todos os memes que se iniciam na letra A, na primeira posição da segunda dimensão todos os que têm a segunda letra como A, e assim por diante. Há apenas um problema nessa maneira, ela é péssima. Nem todos os memes são tradutíveis em palavras (algumas danças e mantras são importantes justamente por serem algo além do mundo linguístico que habitamos diariamente). Escolheremos por ora, no entanto, ocupar-nos apenas com um subconjunto específico daqueles que de fato podem ser expressos em palavras, aqueles que *são* sentenças gramaticais e com valor de verdade. São aqueles que se poderia classificar como V ou F em uma prova. A razão para nos ocuparmos desses em particular é porque é dentre eles que se escondem as proposições científicas, aquelas que versam a respeito da realidade e que podem estar envolvidas numa epistemologia. Gráficos e outras formas de expressão também poderiam estar incluídos num bom retrato científico do mundo, como a palavra “retrato” já denuncia...

Foquemo-nos em proposições: sentenças que admitem interpretação como verdadeiras ou falsas. O único critério estável que parece ter sido cravado na pedra pelo filósofo Bertrand Russell como guia para a verdade proposicional é a coerência interna. Não sabemos nada que garanta que algo seja necessariamente verdadeiro (para alegria dos em breve deconstruídos pós-modernos e cia.), apenas sabemos que aquilo que não é coerente internamente não o pode ser.

Teorias da verdade se espalharam por todo o século XX, primeiro procurando garantir uma relação de correspondência entre elementos linguísticos e os elementos do mundo: se houvesse uma mesma “estrutura” na proposição e no mundo, estaríamos diante de uma verdade. Depois com teorias de aproximação, por exemplo dizendo que a verdade é o ponto para o qual as teorias científicas se aproximam lentamente, da mesma maneira que o limite mínimo de tempo que um ser humano pode correr os 100 metros rasos é aquilo para o qual o histórico de recordes se aproxima lentamente. Um pouco da visão de Dennett sobre esse assunto foi exposto na Introdução, nas parcelas de seu texto “Belief in the truth” (“Fé na verdade”, Dennett, 1997). A análise de memes que estou prestes a fazer não pressupõe nenhum conceito sofisticado e trabalhado por pensadores da verdade, pelo contrário, parece haver boas razões para acreditar que nenhuma das tentativas pontuais filosóficas de se chegar a uma boa maneira de entender a verdade é

consistente, e nesse caso devemos nos manter mais próximos de uma concepção mais inocente e natural de verdade, aquela que utilizamos para aceitar como verdadeira a resposta de alguém confiável que nos diga que dia é hoje.³⁸

Não utilizaremos a sequência de letras para classificar os memes, mesmo tendo restrito nosso espaço aos memes que são de fato descritíveis em sentenças de linguagem, pois não poderemos obter nenhuma informação nova e interessante daí. Mas podemos tomar essa vasta quantidade de memes e espalhá-los de outra maneira. Por exemplo, na primeira dimensão classificaremos os memes por seu tamanho escrito em português, na segunda, por quão mais provável é que esse meme seja dito por um monge budista, na terceira, por quão mais esse meme influenciou a criação do Google etc. O ponto desse raciocínio é perceber que podemos usar quaisquer propriedades arbitrárias para traçar um mapa, no espaço de design, dos objetos que queremos classificar. Como estamos lidando com ideias abstratas, temos de conceder que qualquer forma de classificá-las dependerá de uma tradução, de uma maneira de representá-las. A mesma ideia pode ter representações em português e em inglês que são diferentes, ou pode ser representada em português de diversas maneiras. E por vezes pode até mesmo ser desenhada. Sempre que quisermos transmitir uma ideia para outra pessoa, seremos obrigados a traduzir o conceito mental que fazemos dela numa forma qualquer de representação. E não é diferente quando estamos traçando um espaço de design para as ideias que se replicam, para os memes. Os memes podem sim ser representados de diversas maneiras, mas é necessário fixar-se em alguma específica para poder *ordená-los*. Como estamos tratando de memes e nosso objetivo é compreender melhor como a verdade pode emergir de replicadores “auto-interessados”, vamos utilizar uma dimensão particularmente relevante: a dimensão de *quão mais verdadeiros* os memes são, aquela na qual astronomia e astrologia estão muito distantes. É também o eixo que ordena segundo o grau de precisão, no qual uma afirmação sobre uma distância avaliada num microscópio tem um valor numérico maior do ponto que nos interessa do que uma afirmação sobre a mesma distância com a precisão do olho nu.

³⁸ Um acessível texto sobre essa recuperação de um sentido comum de verdade pode ser encontrada em Yudkovsky (2008b).

O vasto espaço de memes

Estamos olhando para o espaço de todos os memes proposicionais possíveis, um gigante N-dimensional. Vamos imaginá-lo, como antes, como um gigante tridimensional e dentro dele, assim como nos espaços genotípico e fenotípico havia ramificações de animais e seres vivos, aqui existem troncos, galhos e ramificações que representam os memes que de fato existem hoje e seus ancestrais. Os memes também têm sua árvore da vida. Nossa segunda dimensão então será ordenada de acordo com a idade dos memes, exatamente como uma árvore da vida. Nesse momento escolheremos deixar vazio o espaço ocupado por memes que nunca foram atuais, ou seja, nunca estiveram representados na mente de ninguém, na fala de ninguém, em nenhum pedaço de papiro, em nenhuma linguagem que tivesse um sistema capaz de decodificá-la, para sermos o mais genéricos possível. Com efeito, nossa árvore começa a parecer bastante aquela que liga as amebas ao milho, as zebras e a meu primo Thiago.

Mas a estrutura tronco-galhos-ramos não é exatamente igual, afinal, alguns memes se unem com outros para gerarem novos memes, de fato esse processo de união é extremamente comum. Darwin juntou conhecimentos sobre animais e a teoria econômica de Malthus para pensar na seleção natural. John Maynard Smith, o biólogo, utilizou teoria dos jogos da economia para explicar a distribuição de alguns genes na população, por exemplo, o do sistema ABO de sangue. Não vemos lagostilos ou moscofantes pelas ruas frequentemente, embora cruzamento inter-espécies ocorra raramente entre parentes próximos como o tigre e o leão, o cavalo e a zebra. ¿Os memes então se organizam de maneira distinta dos genes nesse sentido?

Não particularmente. No próximo capítulo falaremos sobre como é necessário compreender (e passar) a parte correta de uma analogia para podermos raciocinar sobre um assunto que nos foi introduzido através de uma analogia ou *intuition pump*. Se estamos pensando na árvore da vida da perspectiva dos genes, toda reprodução sexuada é um caso de fusão de uma grande quantidade de genes de dois indivíduos, gerando um descendente, então, ao menos dentre os seres sexuados, existe esse tipo de fusão que parece típico da reprodução cultural, da esfera memética. O que ocorre é que normalmente pensamos na árvore da vida da perspectiva dos indivíduos, e é nesse nível que não existe fusão inter-espécies.

Agora sabemos que, em ambas as árvores, há duas maneiras de se observar. No caso da biologia, se estamos estudando zoologia, pensar no nível dos indivíduos é

melhor do que no nível do gene. Quando pensamos em evolução de memes também devemos encontrar o nível adequado: ¿seria o nível dos memes individuais ou de conjuntos de memes? Aqui percebemos que a fronteira clara entre os diferentes níveis hierárquicos de organização de memes não são tão bem definidos como no caso dos genes. Ninguém confunde um cão com seu genoma. Mas não existe uma distinção clara entre um meme mínimo (o menor tamanho que conseguiria se replicar) e um meme composto, ou *memplexo*. Ao falar da internet, citei uma propriedade interessante que é a transparência de hierarquias, a possibilidade de atravessá-las sem maiores problemas. Agora temos aqui um novo conceito interessante sobre hierarquias, que é o conceito de hierarquias emaranhadas.³⁹

Certo, se não temos uma distinção nítida entre o meme e o *memplexo*, isto é, o conjunto deles que se reproduz em uníssono, podemos dirigir nossa atenção da maneira que parecer-nos melhor, ora aos memes como unidades indivisíveis, *quanta* mínimos de informação necessária para uma reprodução fiel, e ora aos *memplexos*, conjuntos organizados de memes que tendem a caminhar juntos.

No espaço de design da memética, independentemente da profundidade do recorte ao qual escolhamos nos atentar, os galhos podem se unir novamente, os *memplexos* podem compor novas ideias, e por isso devemos abandonar nossa analogia com uma árvore e substituí-la por algo mais minuciosamente relacionado. O delta de um rio parece uma boa escolha: antes de desembocar nos oceanos, alguns rios formam uma grande abertura com córregos e canais, que se separam em vários pontos e se reúnem noutros tantos antes de atingirem seu salgado destino. A evolução memética funciona da mesma maneira. *Memplexos* vão deixando seus filhotes por aí, por vezes com mutações, por vezes sem, as vezes fundem com outros *memplexos* e formam uma nova ideia, um insight, que de fato não é tão “in” quanto poderia parecer a primeira vista.

Insights brilhantes não surgem da luz interior

O uso que se faz da palavra inglesa ‘insight’ é interessante, vejamos as tentativas de definição da Wikipédia brasileira: “Insight pode referir-se a: O ato ou resultado de aprender a verdadeira natureza das coisas; enxergar intuitivamente; capacidade da observação profunda e de dedução; discernimento; percepção; introspecção”.

³⁹ Em *Gödel, Escher, Bach*, Douglas Hofstadter (1979) trabalha o conceito vastamente.

Essas definições escondem aquilo que Minsky, Dennett e outros pensadores sobre a mente trariam à tona imediatamente ao pensar sobre o termo. Insight é o nome que damos a uma ideia que, embora nos pareça ter surgido do nada, é simplesmente produto de sub-sistemas, rotinas, processamentos e cognições que ocorrem em níveis mais baixos do que o nível no qual nos conscientizamos de um processo de pensamento. ‘Insight’ é uma capa mágica que o senso comum usa para vestir uma ideia quando ela é produto do que Minsky chama de ilusão de imanência, a ilusão constante que perpassa a vida de um humano, fadado a estar consciente apenas dos processos de alto nível que ocorrem em sua mente, mas incapaz de escrutinar profundamente e minuciosamente os processos de níveis inferiores que a compõe. Dentre esses processos, encontra-se a fusão de memes. Nesse sentido, memes são mais como o Ganges do que como um carvalho.

Atratores universais estão na memosfera, mas não na biosfera

¿Existiria alguma tendência geral a ser seguida pelos memeplexos? Acredito que sim, para examinar o porquê, pensemos primeiramente na biologia, para depois pensarmos nos memes. No espaço de design da biologia existem vários tipos de movimentos forçados, existem vários *bons truques*, e volta e meia os seres que se aproximam daquele bom pedaço de vizinhança “descobrem” seus caminhos para atingir mais aptidão. Como o universo biológico é vasto e dinâmico, são poucas as tendências universais que nos fariam acreditar numa tendência geral. Bons candidatos são eliminados numa análise simples, supor universais no mundo da biologia é bastante difícil. Supor a ubiquidade da luz do sol desconsidera os peixes abissais, supor a gravidade desconsidera o ambiente aquático e pequenas cavidades em rochedos nos quais algumas bactérias super-resistentes vivem seu dia a dia. Alguns míopes examinadores da história da vida argumentam que sempre existe um atrator no espaço de design biológico para mais inteligência, que dizem que inteligência é sempre evolutivamente boa. Mas a inteligência não é exatamente a habilidade pela qual as minhocas são mais conhecidas, e ainda assim elas estão aí, tão aptas a sobreviver quanto o mais competente cientista, arquiteto ou lutador de kick-box.

O universo memético, por outro lado, é muito mais específico; o meio ambiente dos memes, por ora, é quase só dependente de mentes humanas, que têm interesses largamente biologicamente determinados, maneiras específicas de raciocinar etc. É

bastante mais razoável esperarmos uma regularidade universal nessa área do que na biologia. E de fato, como eu havia citado antes, sabemos que uma linguagem só poderia se sustentar a longo prazo numa população se ela fosse principalmente uma linguagem a respeito de informação precisa, previsível, ordenadora, em uma palavra, verdadeira. Minha defesa principal é que a razão pela qual podemos confiar na nossa visão de mundo atual é que é da natureza de um replicador como os memes que os mais verdadeiros sejam em média mais aptos. Isso não significa que todos os memes verdadeiros serão os mais aptos, afinal isso tornaria inexplicável a venda de bíblias em todo o mundo e o título de “Rei” dado a Roberto Carlos pela Globo, ou por ela adotado (qual será a origem desse meme?). Há alguns nichos particulares nos quais esse fato não se manifesta.

Não é apenas a linguagem que está do lado dessa tese, existem mais critérios de verificação operando sobre nossas proposições que as tornam mais precisas e que eliminam as falsas do que que promovem a falsidade. Se ensino para meus filhos que cavalos podem ter asas, o mundo será critério de verificação o suficiente para que eles contestem. Se digo que o Rio é a capital do Brasil, o tempo os fará descobrir (através de contato com os memes dos colegas e professores de escola) que não se deve crer nisso. Se isso ainda não parecer suficiente, podemos considerar se temos, dentro das nossas disposições para dizer e pensar coisas, mais coisas nas quais acreditamos ou não acreditamos. Como somos animais moldados pela evolução, em grande parte para economizar energia nesse mundo tão disputado, sabemos que nossas atitudes tendem a ser parcialmente racionais e nossos sistemas perceptivos nos dão informação tão precisa quanto precisamos (ou quanto precisaríamos no ambiente para o qual evoluímos). É custoso viver, e por isso fazemos algumas economias, como não ter visão colorida fora do ponto de foco. Podemos não possuir o grau de nitidez desejado em nossas percepções, mas elas sem dúvida estão adaptadas para nos guiar para um caminho certo (não necessariamente *bom*), um caminho que leva a crenças verdadeiras que por sua vez levam a alimentos e parceiros verdadeiros e filhos verdadeiros.

Existe uma pressão seletiva na memosfera que puxa o delta memético em direção a memes mais verdadeiros. O espaço de design dos memes é curvado. Mas alto lá, não é apenas devido a minha vontade consciente de passar uma ideia para você e sua vontade de conhecê-la que a replicação de um meme se dá. A série de desenho South Park já se utilizou de um contra-exemplo muito bom para se fazer piadas. Num determinado episódio o garoto mais doentio e maléfico da escola, Eric Cartman,

descobre a existência da síndrome de Tourette. Os portadores dessa rara síndrome têm descontroles neurais que os fazem muito propensos a dizer coisas sem querer, formam frases completas e gritam por várias vezes, sem ter nenhum desejo consciente de fazê-lo. O que torna a síndrome de Tourette particularmente fascinante para humoristas é que ela zoa a fiação mental de tal maneira que o portador tende a falar muito mais palavrões em seus pequenos lapsos.

No nosso contexto a síndrome de Tourette serve para ilustrar que por vezes mecanismos inconscientes e não muito claros nos fazem passar memes adiante, sendo esta síndrome um caso extremo desse tipo. Também somos particularmente propensos a entender quaisquer memes pronunciados a nós em nossa língua nativa. Algumas pessoas (eu pelo menos) têm um dispositivo de liga-desliga para outras línguas, de modo que podemos ouvir conversas sem nada entender até que decidamos entender algo, assim como podemos controlar nossa respiração embaixo d'água. Mesmo em casos aparentemente pouco ortodoxos de transmissão como esses, podemos confiar em nossa configuração neural para garantir que a maioria das transmissões será fiel aos critérios de verificação do ambiente. A maioria do que ouvimos será confirmado num embate com o mundo, nossos conhecimentos servem para caminhar por ele, encontrá-lo, entendê-lo e enfrentá-lo. Uma pessoa com Tourette, apesar de estar todo o tempo xingando, não abandona o uso de sentenças gramaticais e seus xingamentos muitas vezes são bastante adequados ao contexto. E na maioria dos casos em que se fala sem perceber, sejam casos de pessoas normais ou com algum tipo de problema, podemos confiar que a maquinaria cognitiva estará subrepticamente verificando o que será dito, mesmo que com precisão inferior ao escrutínio consciente que por vezes fazemos em nossas sentenças.

Dentre as objeções da sugestão de que a veracidade é um atrator de memes, não há apenas Tourette, mas há outros casos de anomalia que podem ser encontrados. No mesmo espírito ficam as objeções de que eu tenho/contenho em mim (em algum sentido) o meme “Uma vaca é menor que um esquilo”, mesmo que eu não acredite nele. Mas no sentido relevante, quando estamos falando de memes *que se manifestaram*, são muito raras as oportunidades que um meme falso e inútil como esse aparece, como aqui nesse parágrafo. No geral, os memes que mais têm oportunidade de aparecer são aqueles que melhor condizem com seu contexto de elocução, com a perspectiva do falante sobre o mundo etc. Os memes estão ganhando veracidade com o tempo, e esse tipo de pensamento é o que dá origem à possibilidade de unificação do paradigma reducionista

com uma epistemologia consistente.

Mude um pouquinho aqui, um pouquinho ali, e voilà: temos uma abobrinha!

É hora de voltarmos às nossas máquinas contendo cápsulas *gregorianas*, desenhadas por nós para sobreviver por 400 anos. Como elas têm capacidade de revisar seus objetivos, sua visão de mundo e suas formas de aprendizado e transmissão de aprendizado, haverá bastante mudança ao longo dos 400 anos entre essas criaturas, supondo que haveria milhões, senão bilhões delas espalhadas pelo mundo. Ao longo do tempo perceberíamos algumas diferenças entre as diversas máquinas e suas “visões de mundo”, por assim dizer. É possível que alguma delas tenha uma visão de algum tipo estranho de entidade, por exemplo algo que tenha propriedades de animal e também de pedras. Animais podem se mover e nós os avaliamos de acordo com a perspectiva intencional. Pedras, ao menos na cognição aprendida de nossas máquinas *gregorianas*, são eternas (já que duram mais de 400 anos). Suponha que uma máquina diga que encontrou um animal eterno num beco escuro, e consiga convencer outras deste “fato”; a lenda pode se espalhar, e eventualmente aquele animal que tem uma propriedade típica dos minerais pode adquirir uma propriedade própria dos gases também (flutuar e atravessar objetos). Temos aí uma possível rota para o surgimento de proto-religiões entre nossas máquinas. ¿Isso significa que quaisquer criaturas gregorianas abandonadas a seu cuidado necessariamente criariam algum tipo de proto-religião? Não, e sequer significa que essa é a história da origem de nossas religiões, mas essa conjectura nos abre duas perguntas interessantes ¿Serão as religiões produto de evolução memética, artefatos que encontraram um nicho do espaço de design no qual conseguem florescer na mente de milhões? E ¿os memplexos das religiões são mutualistas, comensais ou parasitas dos seres que habitam, os humanos?

Para responder à primeira pergunta podemos continuar em nossa perspectiva das cápsulas *gregorianas*. Existem muitas maneiras de se espalhar uma crença não veraz entre essas máquinas, a começar por estratégias para enganar adversários, ou ideologias falsas que sirvam para manter uma elite. Simples erro de percepção, ou de transmissão, também podem ocorrer, e pode ser o caso que um robô-cápsula revise sua programação de modo a piorar seu sistema de avaliação de verdades. Todos esses métodos levariam ao florescimento de um meme, por pequeno e inocente que seja, o qual não se basearia em sentidos apurados e desejo de saber, mas em alguma forma de

banalidade histórica, confiança na autoridade ou erro de percepção. Uma vez que esse nicho haja sido criado, um meme pode sofrer mutações dentro dele, tornando-se por alguma razão mais interessante às mentes as quais habita. Um machado de lava é uma ideia atraente, mais atraente do que um machado invisível, mas nem de perto tão atraente quanto um homem-cavalo com um arco e flecha de fogo que é invisível e está representado por um grupo de estrelas no céu. Sagitário seguramente não é o produto de uma única mutação, nem de uma única mutação que se desvia de um objeto da realidade, mas de uma sucessão delas. O recente livro de Dennett (2006), *Quebrando o encanto*, tem como tema principal a religião, e a visualiza de uma perspectiva memética. Em particular, ele procura compreender os mecanismos que *possivelmente* podem ter gerado as religiões organizadas que temos ao nosso redor hoje, e traça diversas hipóteses a respeito de como se dá o fenômeno da religião. Além disso ele recomenda que ensinemos sim religião a nossas crianças, mas ensinemos desde cedo todas as religiões, e a existência de pessoas não religiosas. Em última instância ele faz um apelo a uma espécie de democracia memética, na qual todas as pessoas tenham acesso a todos os memes religiosos, ou todos os memes religiosos tenham acesso a todas as pessoas (notem como em algumas construções as hierarquias ficam emaranhadas).

Não pretendo adentrar aqui mais profundamente o conteúdo desse livro, mas deixo no ar a ideia de uma análise memética do surgimento, desenvolvimento e transmissão das religiões como um exercício mental ao leitor interessado. Por ora, apenas pontuo uma das características mais interessantes da transmissão de memeplexos religiosos. Quase sempre as pessoas herdaram as religiões de seus pais, então diferentemente de boa parte dos memes, cuja transmissão é largamente horizontal, os memes religiosos (sejam eles hindus, budistas, cristãos, mórmons, judeus) são passados com fidelidade principalmente de maneira vertical, assim como os genes. Não estou insinuando (e não seria verdade) que haja genes para o candomblé, genes para crer em Jesus ou para adorar o Flying Spaghetti Monster; a verticalidade desses memes não está sozinha, e talvez seja interessante ao leitor justamente tentar encontrar alguns memeplexos que também se transmitem verticalmente e procurar propriedades em comum entre a maioria deles. Eis aí um não ortodoxo “segundo teste” de simulação dennettiana.

Finalmente, antes de encerrar o capítulo, pensemos no mundo futuro repleto de cápsulas robôs com memes super avançados de crenças religiosas, ¿seriam esses memes bons ou ruins da perspectiva das pessoas criogenadas dentro das máquinas? Depende,

seria necessária uma análise comparativa entre os aderentes às diferentes religiões que surgissem e aqueles que não adotassem nenhuma para ver suas chances de sobrevivência.

Uma pergunta mais importante: Seriam esses memplexos bons ou ruins da perspectiva dos robôs-cápsula? O livro de Dennett é uma chamada para que a ciência responda essa pergunta a respeito de nossas religiões. Não sabemos, mas não há porque continuarmos ignorantes. O território do estudo das religiões foi renegado pela ciências duras, em particular aquelas que levam em conta a perspectiva darwinista, e o objetivo de Dennett é ajudar esse hiato acadêmico a acabar. É chegado o momento de estudarmos cientificamente a religião, assim como um dia descobrimos que poderíamos estudar cientificamente a psicologia.

8. Intuition pumps

Intuition pumps, literalmente: bombas de intuição; conotativamente: experimento mental, *Gedankenexperiment*, situação imaginária, analogia, metáfora, bengala do pensamento, ferramenta do pensar, projetor de ideias.

Vejam os que Dennett (in Brockman, 1995, p. 182) diz a respeito:

Se olharmos a história da filosofia, veremos que todas as coisas influentes e grandiosas eram tecnicamente esburacadas, mas incrivelmente memoráveis e vividas. São o que chamo de ‘intuition pumps’ – adoráveis experimentos mentais. [...] Eles não são argumentos, são histórias. Ao invés de ter uma conclusão, eles bombeiam uma intuição. Eles lhe fazem dizer “Ahá! Agora entendi!”

Há boas e más maneiras de bombear nossas intuições, o que é esperado, na medida em que fomos traçados para nos inclinar intuitivamente ao que quer que fosse preciso para os genes e memes se replicarem. Nossos processos de pensamento tendem criticamente para direções que muitas vezes não podemos antever, ou que nos parecem irracionais, ilógicos, e em larga medida de fato o são.

Também é de se esperar, dada nossa breve incursão pelo mundo dos memes e suas replicações, que a história da filosofia esteja pontuada por dezenas, senão milhares de experimentos mentais desse tipo, intuitivos, memoráveis etc. Afinal, ¿não é o mesmo que se passa com todos os nossos memes? Em outras palavras, ¿não é o caso que, no vasto espaço de design das máquinas gregorianas, é quase sempre um *bom truque* travestir uma ideia com uma roupagem de algo em que a criatura se interessa?

Os intuition pumps em geral focam nossa atenção para um ou outro aspecto de um argumento, e portanto diminuem a proporção relativa de outros. Assim sendo, eles são memes muito perigosos. Mesmo sendo perigosos, por conta de serem memes muito profícuos e poderosos no mundo no qual nossas ideias evoluem, devemos saber que utilizá-los não é uma afronta, uma ameaça, uma violência linguística ou retórica barata. No grau de sofisticação que esse tipo de argumento possui hoje em dia, depois de milênios de seleção, é uma retórica bem cara, e sabemos que o beneficiário não é necessariamente o pensador, mas o pensamento.

Do demônio de Laplace ao moinho de Leibniz, a história da filosofia está repleta de argumentos mentais, experimentos mentais, bombas intuitivas, argumentos por analogia etc. Uma simulação do pensamento dennettiano é em si mesmo uma espécie de

intuition pump, ao declarar e tornar explícito o objetivo de permitir ao leitor a capacidade de pensar nos mesmos moldes de Dennett. No mesmo sentido que um software pode ‘copiar’ outro software dentro dele, em sua própria linguagem, estou também induzindo a intuição do leitor a se formatar de uma maneira, a conduzir seu aprendizado de acordo com essa ideia. Mas não nos deixemos levar por meta-intuições: apresentarei uma sequência de intuition pumps que Dennett apresenta em sua obra. Talvez seja o caso de o leitor ficar mais atento ao nível mais alto do processamento cognitivo desses próximos parágrafos, do que ao nível da explicação ela mesma. Isto é, mantenha seu olho aberto para a *forma de introdução de um argumento*, mais do que para o argumento em si. Não porque os argumentos não sejam interessantes, mas porque estarão sempre comprimidos num espaço pequeno, dada a natureza resumida de qualquer compêndio que pretenda sumarizar a obra de um autor.

Alguns dos temas de intuition pumps já foram tratados anteriormente (significado, sintaxe e semântica) e outros ainda estão por vir (padrões e principalmente consciência). Isso é mais uma razão para não se desesperar com um entendimento obtuso dessa torrente de bombas de intuição que vou apresentar. Elas ganharão a devida atenção no devido tempo, se tudo correr bem. Começemos por aquelas que Dennett critica:

A) Memória como um estoque de coisas escritas

Muitas vezes somos convidados a pensar na memória humana como um estoque ou biblioteca cheia de coisas escritas. Esse intuition pump dirige nossa atenção em particular a três aspectos da estocagem. O primeiro é que a estocagem é feita em forma de sentenças (em uma ou mais línguas), o segundo é que ela é grande e espaçosa, e o terceiro que ela é estática: uma vez lá, lá fica.

Os três aspectos são bastante ruins e nos dão uma visão da memória como ela *não é*. O cérebro humano é uma estrutura muito complexa, mas em todas suas complexidades ela não poderia armazenar todas as sentenças que somos capazes de emitir. Esta sentença auto-referente que estou agora escrevendo, por exemplo, tem doze palavras. Uma sequência de doze palavras num idioma complexo inclui provavelmente um número de combinações possíveis que não permitiriam que elas fossem todas ditas em sequência desde o surgimento da vida na Terra. A memória não pode estar estocada em forma de sentenças. Em particular, nossas línguas funcionam de uma maneira que

depende de generalizar um uso (por exemplo colocando “ão” no final de uma palavra para criar um aumentativo) e existem demonstrações de que essas regras não poderiam ter sido aprendidas. Isso aponta para um modelo de memória no qual o armazenamento está em um nível mais profundo do que o nível das palavras ou sentenças. É mais simples para a evolução criar um aparelho de internalizar conceitos e um aparelho de transduzi-los em sentenças, do que apenas um aparelho muito grande de linguagem. Além disso, como Noam Chomsky ficou famoso descobrindo, parte do nosso registro mental de linguagem já está conosco desde o nascimento (ou melhor, está codificado no nascimento e vai se expressando conforme crescemos).

A memória também não é infinitamente grande, o que seria um pré-requisito para que de fato mantivéssemos somente as sub-estruturas linguísticas nela, num massivo estoque. Oitenta bilhões de neurônios são muitos neurônios, mas é importante lembrarmos que cada neurônio não é um indivíduo pensante, mas um sistema físico de ação e reação. Requer-se muita maquinaria neural para registrar informação, e sabemos hoje que essa informação está codificada muito mais na forma como os neurônios estão conectados do que dentro deles. Essa maquinaria neural ainda se divide entre tarefas motoras, perceptivas, de equilíbrio etc. Outra propriedade fundamental do cérebro, sobre a qual trataremos mais longa e precisamente mais adiante, é a redundância. Qualquer neurônio (ou até centenas deles) destruído de nossos cérebros, até onde sabemos, é inessencial para as atividades que fazemos e para a continuidade de nossa vida mental e de nossa personalidade, como qualquer adolescente de ressaca já percebeu.

B) O argumento do quarto chinês

Houve um tempo em que se dizia que a função dos filósofos que se utilizam de inteligência artificial era refutar o argumento do Quarto Chinês de John Searle. Dennett tomou um caminho diferenciado, e sugeriu que o Quarto Chinês era uma bomba de intuição mal projetada, que nos leva a uma compreensão falsa sobre os poderes de um computador.

O argumento defende que um computador, sendo apenas uma máquina de manipulação de ordens, só realiza tarefas puramente sintáticas. As mentes, por outro lado, são semânticas, no sentido de que possuem mais do que uma estrutura formal, pois possuem um conteúdo. Como ilustração, sugere-se imaginar que um sujeito que não fala chinês é posto num quarto com um livro de manipulação de ideogramas chineses. O

livro está escrito em português, e descreve regras formais de como operar aqueles ideogramas, sem explicar o que significam. Do lado de fora do quarto, um chinês está escrevendo como se estivesse conversando com alguém dentro do quarto. Searle defende que mesmo que a manipulação mecânica de símbolos seja boa o suficiente para imitar perfeitamente um falante de chinês para o chinês do lado de fora, o brasileiro do lado de dentro não entendeu nada, e isto por sua vez demonstraria que apenas instanciar um programa (conjunto ordenado de regras sintáticas) não seria suficiente para compreender algo.

Segue que computadores não são capazes de semântica.

Dennett aponta que esse quase-argumento tira de nosso foco de atenção uma série de processos que precisam ocorrer para um entendimento consciente de alguma coisa. Em particular, memória, lembrar, emoções, conhecimento do mundo e racionalidade são aspectos de qualquer sistema que fosse passar num teste como esse. Searle sabe, pontua Dennett, que um programa que fosse preencher o pré-requisito de responder em chinês teria de ter todas essas propriedades, mas a forma como ele introduz o argumento nos desencoraja a pensar nesses aspectos, mais ou menos da mesma maneira que ao escolher termos filhos tendemos a ignorar a adolescência nos cálculos.

C) Ideias como cópias fracas de fotos

David Hume foi dos primeiros a perceber que para explicar o pensamento era necessário livrar-se do pensador. Introduziu algumas teorias interessantes, por exemplo a de que as ideias na mente são como uma cópia enfraquecida da experiência e que as ideias puxam umas as outras, como se fossem tensionadas por cordas ou algo assim. Para seu tempo, era um avanço interessante, mas a analogia que mais faz sentido hoje em dia, a que direciona nossa intuição a um caminho menos tortuoso, é a de que a estrutura do mental é composta de vários recursos, cada um deles menores que o todo e capazes de executar menos tarefas. Esses recursos (chamados também de agentes ou demônios) são divisíveis em outros ainda menores, com menores capacidades, e assim por diante até chegarmos ao nível já discutido da idiotice completa e mecânica de um neurônio.

D) Buscar na memória como o aviário de Platão

Platão nos traz a imagem de um aviário como a mente, e vamos adquirindo novos pássaros conforme aprendemos coisas. Se quisermos algum conhecimento temos de apanhar o pássaro referente e apresentá-lo. Uma pessoa que diga que $5+5=12$, por exemplo, pegou o pássaro errado na hora de responder.

A essa altura o estilo de raciocínio de Dennett já deve estar suficientemente claro para que o leitor perceba o que é criticável nessa teoria sobre como pensamos. Após terminar esse exercício de simulação, iremos agora para as bombas de intuição que Dennett *favorece*. As primeiras estavam com letras, estas estarão com numerais romanos, para evitar confusão.

I) Mente como pequenos demônios com pequenas tarefas

Imagine-se na sala de controle do cérebro, de um lado existe uma parede enorme cheia de luzes vermelhas que piscam de maneira que, para você, parece completamente desordenada. Na outra parede, existem botões. Não há mais nada no quarto. Sua tarefa é sobreviver. Poderíamos tentar apertar os botões aleatoriamente, mas provavelmente essa não seria uma estratégia de sucesso. “Ahá, já sei, eu posso colocar *rótulos* nas luzes e assim saberei quando apertar que botão”. Dificilmente essa é a estratégia adotada pela evolução... Essa não pode ser a estratégia do cérebro porque precisaríamos novamente de outra “sala de comando” dentro de nós, numa recursão sem sentido. ¿Então o que podemos saber? Podemos saber que a solução tem de ser mecânica e automática, por exemplo conectando com fios as luzes e botões, assim não é necessário um controlador, e sequer uma sala de controle. É só existir a seleção natural beneficiando aqueles que têm as conexões certas e impedindo a continuidade dos que têm as erradas. Se as tarefas forem muito complexas, as conexões terão de possuir subsistemas mecanizáveis (demônios, recursos, agentes) que fazem algum tipo de computação.

II) Consciência como múltiplos rascunhos

As vezes erramos uma letra, misturamos sentenças em nossa fala, as vezes fazemos um gesto com a mão e dizemos o contrário com a boca sem perceber, cometemos atos falios, etc etc.... ¿Como modelá esses pequenos problemas dentro de uma

teoria da consciência?

Temos de imaginar que existam várias rotas de informação caminhando dentro de nossos sistemas perceptuais e nossos sistemas motores, cada parte só estando em contato com suas vizinhas, de forma que existem diferentes fluxos de histórias correndo ao longo de nossos pensamentos. Alguns têm mais ou menos força, mas é possível que haja duas ou três ou mais legiões de neurônios nos dirigindo para caminhos diferentes *ao mesmo tempo*. Se várias crianças começarem a rascunhar algo num papel, teremos vários quase-desenhos em formação até que um deles domina o cenário ou eles se unem de maneira medianamente harmônica e se tornam um único desenho. Os rascunhos que não foram plenamente incorporados ainda estão lá, no papel, mas tornou-se claro que há um desenho que tomou espaço dos outros. Essa é a ideia da consciência como múltiplos rascunhos, que há diversas histórias transcorrendo ao mesmo tempo pelo seu sistema de pensamento. No momento, a corrente de leitura e processamento da informação semântica está dominando, e por isso você está entendendo essas palavras, mas existem também rascunhos que não chegaram ao ponto de exercer alguma diferença em seu comportamento visível, com partes do processamento visual de sua visão periférica, ou sons de fundo aos quais você não estava dando atenção até eu falar sobre eles. Um certo tanto de atenção o seu organismo estava dando para esses estímulos, ele só não era suficiente para tornar-se parte do desenho dominante.

III) Consciência como a fama no cérebro

Na formulação mais atual de Dennett sobre como pensar a consciência, trocamos nosso intuition pump e nossa nova analogia é com a fama. Temos um povoado enorme de neurônios, correntes corticais e subsistemas operando em nosso sistema nervoso. Quanto mais famoso você é nessa população, mais você influencia o comportamento do dono do corpo, e mais você é consciente. As vezes alguma coisa rouba nossa atenção por um tempo fugaz, o que poderíamos pensar como uma eleição para presidente: alguém se torna imensamente famoso por 4 anos e depois deixa de ser tão famoso. Outras coisas estão sempre operantes em menor nível, têm menos fama, mas podem durar mais, como a fama de Beethoven por exemplo. Um zunido de mosca é como uma eleição presidencial, modificando radicalmente as estruturas de fama no sistema nervoso, capaz de nos tirar a atenção de quase qualquer coisa, enquanto ficar em pé, equilibrado sem cair, é mais como Beethoven, algo que o sistema está sempre ocupado

em fazer, mas que raramente atrai tanta ‘gente’ a ponto de controlar o foco de atenção.

Notem que esse modelo é uma versão mais sofisticada do anterior; o nível de profundidade da analogia anterior é mantido, mas são acrescentados alguns novos aspectos que fazem dessa uma analogia um pouco melhor. É bastante intuitivamente claro, por exemplo, que quando a fama de uma pessoa diminui ela ainda existe, e menos claro que ela ainda tem influência na fama das demais pessoas. Já conhecemos os memes, então podemos falar da fama de memes. Com a morte do cantor, o meme “Michael Jackson” atingiu um pico de fama que há muito não tinha; ainda é frequente ver a influência desse período em lugares tão diversos quanto discotecas, conversas em restaurantes e livros de filosofia. É bastante claro que fama é dinâmica, efêmera, e não como um rascunho numa folha de papel. Por todas essas razões, ela é um guia de intuição melhor.⁴⁰

IV) Percepção de tempo como a descoberta do Império Inglês no tratado de Ghent

Suponhamos que você haja lido estes dois últimos intuition pumps a respeito da consciência e não esteja nada convencido dessa ideia. Você poderia por exemplo dizer o seguinte. “Está bem, entendo que isso explica meu comportamento num certo sentido, e que isso se determina através de milhões de pequenos fluxos ganhando e perdendo força até que haja alguma força dominante, mas existe um momento introspectivo no qual eu *sinto* que entendi alguma coisa, por exemplo. Quando você me diz ‘São duas horas’ de início há um processamento de som acontecendo, depois alguns fluxos informacionais começam a trilhar seu caminho, mas eventualmente sou *eu* quem entende sua fala, e é a partir desse momento que eu passo a poder fazer algo com essa informação.”

Essa é uma objeção interessante, e talvez caiba uma história para traçarmos uma comparação (Dennett, 1996b). Houve uma guerra, chamada de ‘Guerra de 1812’ entre ingleses e norte-americanos. Em 1814, em Ghent, hoje em dia na Bélgica, foi assinado um acordo que encerrava a guerra. A notícia começou a correr o mundo. Os

⁴⁰ Uma analogia que não foi defendida por Dennett enfaticamente, mas que encontrei em trabalhos mais recentes e que parece ainda melhor, é pensar a consciência como política. Uma vez que seu partido elege um candidato, os demais partidos ainda mantêm uma força parlamentar, poder de persuasão etc., mesmo que sejam os memes do seu partido que estejam dominando o cenário político naquele momento. O segundo partido que teve mais votos provavelmente será o segundo mais influente durante o período em que o seu partido estiver dominando, e assim por diante. A vantagem da analogia política é que já estamos automaticamente acostumados a pensar na influência que os políticos exercem sobre os outros, e não costumamos levar em consideração intuitiva a influência da fama da rainha da Inglaterra na fama do escritor J.R.R. Tolkien, autor de *O senhor dos anéis*.

embaixadores, duques, soldados e generais do Império Inglês iam se familiarizando com a informação conforme os cavalos, navios e pombos chegavam com as notícias. Suponhamos agora que alguém fizesse a inusitada pergunta: “¿Quando o *Império* ficou sabendo da assinatura do tratado?” Ora, depende, em Londres, em menos de um dia, nas ilhas inglesas do Pacífico, muito tempo depois, em Nova Orleans não foi a tempo suficiente para evitar uma batalha entre as nações. Por mais de um semestre, batalhas foram vencidas e perdidas por soldados desinformados de que a guerra já havia acabado. A pergunta que fizemos não faz sentido, ela não é apenas inusitada, ela é *mal colocada*. Da mesma maneira, é mal colocada a pergunta “¿quando foi que *ele* entendeu aquela equação?” O *ele* aí é como um império, composto de milhões de partes menores, e cada pedaço foi dominado por essa informação numa hora diferente. O meme da equação vai tomando espaço e se aconchegando na estrutura mental de um sujeito, conforme essa informação vai sendo tacitamente armazenada (não através de regras explícitas, mas de aproximações *suficientemente* boas). Eventualmente a pessoa consegue fazer todas as atividades que associamos a compreender uma equação, como chegar a resultados, reproduzi-la, explicá-la etc.

V) Semântica como aquilo que os cientistas nunca arrancariam das caixas pretas com botões e lâmpadas

Antes daquele parágrafo bastante complicado no capítulo 1 sobre reducionismo (ao qual voltaremos no final do texto), eu citei um intuition pump interessante de Dennett (1995a, pp. 412-5) sobre a necessidade de continuarmos com a semântica, no sentido de usar o vocabulário que diz respeito a ela. Duas caixas pretas conectadas por um fio passam informação uma para outra através do único fio de cobre que as une. Dentre três lâmpadas numa das caixas, apenas duas se acendem, uma quando na outra caixa um botão é apertado, e a outra quando o outro botão é apertado. A informação no entanto está correndo por apenas um fio. Mesmo avaliando os padrões físicos que atravessam o fio, toda vez que um botão é apertado, não podemos saber porque aquela luz se acendeu e não a outra. A cada apertada, o padrão que atravessa é diferente, e isso não nos permite generalizar um mecanismo simples. Eventualmente cortamos o fio e começamos a enviar códigos (quaisquer) de mesmo comprimento para a caixa com lâmpadas, e surpreendentemente a terceira lâmpada se acende. O mesmo ocorre repetidas vezes.

Eventualmente os criadores das caixas nos explicam que uma delas recebeu input de grande quantidade de 'conhecimento' do mundo, e a outra recebeu de outra pessoa grandes quantidades de informação a respeito do mundo, um dos botões é um produtor de sentenças verdadeiras, o outro de sentenças falsas. Uma das luzes acende quando uma verdade é detectada, a outra acende quando uma mentira é detectada. O código é um código comum de computação, seguido de algarismos aleatórios até completar um tamanho padrão. Quando os cientistas curiosos (ou nós mesmos, já que o experimento é mental) cortam o fio e mandam códigos de mesmo tamanho, eles nunca produziriam uma sentença verdadeira ou falsa, mas algo como “Hasdgsdik30sohso” ou “000)))Suf998---)”, e esse é o caso no qual a terceira lâmpada acendia.

O ponto de Dennett com essa intuição é voltar nossa atenção para a *inescrutabilidade* da semântica a partir de um vocabulário totalmente sintático. O idioma da física não era suficiente para compreendermos o que estava sendo transmitido e entendermos aquele padrão. A perspectiva física ali não poderia resolver nossos problemas de previsão, precisávamos adotar a perspectiva do design, e isso é justamente o que nos forneceram os criadores das caixas, ao nos informar com que finalidade foram criadas.

VI) Calculadora como demonstração de que não temos representações internas das regras

Também no início do presente texto citei o caso de que existem várias maneiras através das quais uma calculadora pode fazer uma operação como 7×6 , e que ela não precisa conter uma representação da regra lógica em si para executar a tarefa. Da mesma maneira, não é necessário que nós tenhamos uma representação bem definida específica das regras lógicas às quais nosso comportamento está constantemente submetido. Aqui a ideia do intuition pump é trazer a atenção do leitor aos estágios intermediários do processamento de informação, que em um caso trariam os números 6 12 18 24 30 36 42 e em outro caso 7 14 21 28 35 42, no sistema de armazenamento (ou cache).

VII) Escala “máquina de refrigerante, sapo, humano” como demonstração da inescrutabilidade da referência dos objetos nocionais

Seja uma máquina de refrigerante. Sejam a moeda de 25 centavos norte-americana e o *quarter balboa*, uma moeda utilizada antigamente no Panamá que tem o mesmo peso, grossura e circunferência da versão americana. A máquina de refrigerante foi produzida nos EUA, e passou seus primeiros dez anos recebendo moedas americanas. Eventualmente foi levada ao Panamá, e agora (estamos no passado) recebe *quarter balboas*.

Somos tentados a imaginar que a máquina recebe os balboas *como se fossem* 25 centavos americanos. Mas essa intuição não pode resistir a uma análise mais minuciosa: se tivéssemos desde o início inserido balboas na máquina, ela reagiria do mesmo jeito, se inserirmos um quarto de dólar agora, ela fará o mesmo. A máquina não serve para receber 25 centavos de dólar, ela serve para receber objetos de peso X, tamanho Y e forma Z.

Seja um sapo. Sejam moscas que caminham por seu pântano. Quando uma mosca passa suficientemente perto do sapo, sua língua é ejetada com minuciosa precisão para uma deliciosa refeição de fim de tarde. Agora imagine que desenvolvemos pequenas catapultas de carne moída. O sapo provavelmente teria exatamente a mesma reação à passagem de uma carinha voadora que à passagem de uma mosca. ¿Estaria o sapo comendo aquela carinha *como se fosse* uma mosca, ou seja, existiria aí alguma significação do objeto desejado da parte do sapo? Também não. Não existe uma verdade última sobre o que é o objeto nocional que o sistema do sapo deseja. O nível de sofisticação da retina e do cérebro do sapo simplesmente não tem essa informação. O sapo está simplesmente detectando um certo padrão e reagindo de acordo com o que a mãe natureza mandou.

Imagine que você é transportado para um planeta idêntico ao nosso, como ocorre ao personagem de *Guia do mochileiro das galáxias* (Adams, 2004), mas você não sabe que isso aconteceu. Tudo se passou durante seu sono, colocaram você numa cama igual a sua no planeta Z. A única diferença entre os planetas é que o líquido transparente que as Zpessoas tomam por lá é composto de XYZ e não de H₂O. Você acorda e dá bom dia para seu Znamorado, Zprimo, Zpai ou Zco-habitante.

Há um conjunto de filósofos que defende que suas crenças a respeito do mundo agora são falsas, porque suas crenças são da forma: “A respeito de ‘pai’, que pai toma café da manhã comigo”, ou “A respeito de ‘meu quarto’, que ele fica no andar de baixo da casa”.

Mas isso não pode ser verdade, o grau de determinação de nossas crenças não inclui esse tipo de sofisticação, isso não é necessário para nossa vida. Não temos um “A respeito de” que determina os objetos de maneira última em nossa cognição. Assim como o sapo pode comer a carne *sem* ser *enganado*, porque sua crença não é tão determinada, nós podemos também interagir com pessoas sem que nossas crenças se tornem falsas. O significado das nossas crenças é inescrutável, olhando apenas para nosso corpo como referência. Como vimos anteriormente, a quantidade de informação a respeito das circunstâncias modifica o significado de nossas atitudes proposicionais, e a extensão do agente que escolhermos também. Existe uma opacidade na informação a respeito do que cremos ou não. Nossas crenças não são a respeito do mundo, são a respeito de nossos mundos nocionais.

¿Porque então parece tão difícil imaginar situações desse tipo? Porque nossos sistemas de detecção são tão sofisticados, tão bons mímicos da semântica, que só uma situação bastante esdrúxula pensada por Hilary Putnam poderia passar despercebida por eles. Criaturas gregorianas são incrivelmente poderosas, mas não podemos assumir que seus poderes cheguem ao grau de determinação exigido para crenças mais determinadas do que seu nível máximo de percepção. Não temos acesso epistêmico imediato aos objetos que nossas crenças delineiam de maneira aproximada, e por isso nossas crenças *não são exatamente sobre aquele objeto do mundo atual*.

Aqui encerramos nossa torrente de intuition pumps⁴¹. Terminando citando Dennett, a respeito de como as simulações computacionais alteraram a indústria de produção de intuition pumps:

É importante reconhecer que essas simulações computacionais são de fato experimentos mentais filosóficos, intuition pumps, não experimentos empíricos. Elas sistematicamente exploram as implicações de um conjunto de suposições. Filósofos costumavam ter de conduzir seus experimentos manualmente, um de cada vez. Agora eles podem conduzir milhares de variações em uma hora, uma boa maneira de checar para ter certeza que as intuições que eles bombeiam não são artefatos de uma propriedade arbitrária do cenário. (Dennett, 2003b, p. 218.)

⁴¹ Leituras adicionais: para uma introdução mais aprofundada sobre as razões para afirmarmos a falta de acesso epistêmico imediato ao mundo, ver o artigo “Beyond belief” (Dennett, 1982b). Para reflexões sobre esse tema, conectando-o com a defesa de David Chalmers do “problema difícil da consciência”, ver meu artigo “Fishing the content and epistemology of phenomenal belief” (Caleiro, 2012). Para reflexões conectando a ideia de mundos nocionais com os designadores rígidos de Kripke e a semântica bidimensional da filosofia australiana contemporânea, ver o artigo “Humble primary intensions” (Fabiano, 2013).

9. Tornando conceitos antigos palatáveis

Os filósofos, mas não as pessoas comuns, têm o estranho hábito de acreditar que existe um objeto no mundo para cada palavra que consigam inventar. Esse mote já gasto nos círculos de filosofia carrega muito de verdade. Em primeiro lugar, temos uma grande abundância de palavras imprecisas, e uma forte intuição de que elas devem significar alguma coisa, devem se referir a algo específico. Já vimos que nossas intuições sobre a existência ou não de referência e significado específicos são muito enganadoras. Há várias maneiras de lidar com isso, mas as duas mais utilizadas na filosofia são as seguintes. Primeiro, declarar que algo é inconsistente, absurdo ou logicamente impossível: “Deus não pode ser causa primeira porque isso fere o pressuposto de que tudo tem uma causa”. A segunda estratégia é dizer que talvez o conceito em questão não seja exatamente como pensamos, mas que podemos modificar isso e aquilo nele e teremos algo que pode ser trabalhado: “Podemos pensar Deus como sendo as forças que regem o movimento do universo, ao invés de como sendo o *criador* dessas forças”. Ambas as estratégias possuem defeitos e vantagens. Dois principais defeitos de declarar inconsistência são a persistência da intuição de que algo permanece inexplicado e a falta de um foco claro do que estudar depois. “Ok, isso não existia, mas isso não me dá o que aprender de novo”. Modificar o significado nos possibilita encontrar uma nova área de escrutínio, e creio que essa seja a razão pela qual é preferida por Dennett.

Muitas vezes conceitos não são totalmente vazios, mas apenas imprecisos, nunca chegaram a passar por uma afronta cara a cara com a realidade. Quando nos permitimos modificar conceitos, também estamos nos permitindo *refiná-los* a um grau de determinação ao qual não estavam submetidos antes, e com isso podemos obter algo de novo e interessante. Evidente que nem todas as maneiras de refinar serão boas maneiras de refinar, mas todas as boas maneiras são modificações, o que é mais um argumento em favor da modificação como estratégia de ataque a problemas filosóficos.

A desvantagem principal desta estratégia é que nos dá a mesma sensação do público que foi a um show de strip-tease para o qual contrataram um palhaço para fazer a abertura: “Pode ser divertido, mas não paguei por isso.”

Nesse capítulo falaremos de três conceitos clássicos da filosofia, *self*, livre-arbítrio e consciência, e veremos quais desconstruções, reconstruções e refinamentos

são oferecidas por Dennett para que nos sintamos o mais satisfeitos possível com o que possamos querer.

Self

A palavra “self” não é tão frequentemente usada em português quanto em inglês, mas pode ser ouvida ocasionalmente em algumas conversas. Alguém que conheça ambas as línguas provavelmente deve ter uma ideia mais ou menos clara de quando ela deve ser utilizada, mas a ideia pode não ser tão imediata a um falante apenas de português. O melhor auxílio para pensar sobre isso é imaginar uma situação no qual você está falando de si ou de outros na terceira pessoa, usando a palavra “eu”. Por exemplo: “Meu eu é incrivelmente irritante quando se trata de arrumar a casa”; “Ele parece uma pessoa sincera, mas o eu dele é mau”; “Tenho que melhorar meu eu para atingir minhas metas para o ano”. De certa maneira, você está fixando o que é persistente naquela pessoa, mesmo que ela mude de roupas, cabelos, sotaque etc.

Nem sempre é imediata a problematidade da ideia de um Eu, ou de um alguém em geral. Tente por você mesmo atender à intuição de que evidentemente existe um você (provavelmente em algum lugar entre suas orelhas e atrás de seus olhos) e escrever em um parágrafo o que é o tal Eu, ou Self.

As respostas em geral vão se dividir entre aquelas que não são muito explicativas, como “Meu Eu, a pessoa que está aqui agora nesse corpo”, caso no qual a palavra “pessoa” é uma bengala invisível para disfarçar o problema sem resolvê-lo, e aquelas que não sobreviveriam a uma análise minuciosa, como “Eu sou a soma de tudo o que esse corpo faz”, que pode ser aniquilada diante de questões como ¿e se eu transplantasse seu cérebro para outro corpo?, ou ¿inclusive aquilo do qual você não tem nenhuma memória que ocorreu no seu segundo dia de vida?

É bastante mais difícil do que parece entender o que somos, caso realmente nos proponhamos a essa tarefa, e há aqueles que defenderam que, em última instância não há um Self, da mesma maneira que não há uma massa de gás pensante habitando meu corpo de maneira invisível, como pensavam os povos antigos. A questão a respeito de nós pode ser refinada de várias maneiras. A sequência que talvez foi mais profícua na filosofia pode ter sido a substituição da pergunta “¿Existe um Self?” por “¿O que é um Self?”, desta para “¿O que determina a identidade pessoal?”, e finalmente para “¿É a identidade pessoal *o que interessa?*”. Essa última questão foi vastamente discutida pelo

filósofo Derek Parfit ao longo de mais de trinta e cinco anos, sendo que a compilação mais atual de sua resposta a questão pode ser encontrada no artigo de 2007, “Is personal identity what matters?” (Parfit, 2007).

O que Dennett nos oferecerá nesse campo é uma maneira de pensar o Self que seja consistente com o que conhecemos da realidade. Uma noção de Self que é compatível ao mesmo tempo com os usos que fazemos da palavra e com aquilo que sabemos sobre nós, desde sermos constituídos por recursos simples que se organizam mecanicamente em hierarquias de fama, até não sermos *nós*, num sentido metafísico triunfal, que executamos o *entendimento* ou que fazemos, *acima e além* das partes que nos constituem, os processos de significação.

Na *intuition pump* dos múltiplos rascunhos, sugeri que se dermos uma folha a várias crianças para fazer um desenho, provavelmente emergirá um desenho principal a partir dos inícios de desenho, que ficarão talvez um pouco fora de contexto no todo. Quando falamos de alguém como uma pessoa, ou como um Self, estamos falando dessa pessoa como algo que permanece o mesmo, tem convicções similares, desejos, formas de agir etc. Estamos de certa maneira tomando todas as manifestações de um certo sujeito e fixando-as como pertencendo a uma só pessoa. Mas sabemos (e os psico-economistas nos mostram cada vez mais) que mudamos muito nossas decisões o tempo todo, que podemos ser influenciados por praticamente tudo ao nosso redor, de maneiras imprevisíveis. Uma placa de “Promoção 50% do preço” comprovadamente é capaz de vender um cano de metal torto que simplesmente não pode ter uma função caso você não possua um carrinho de golfe (é uma peça do carrinho). Quando as pessoas são perguntadas a respeito de porque fizeram a compra, prontamente estão dispostas a dar razões, mais ou menos inventadas. Se ordenamos a uma pessoa hipnotizada que, ao sair do transe, abra a janela, e depois que ela o faz perguntamos o porquê, ela inventará uma história conforme as condições do momento lhe permitam. Temos um hábito de nos justificar que muitas vezes ultrapassa nossa consciência do que estávamos de fato fazendo. O seguinte parágrafo ilustra um caso particularmente fascinante desse tipo de ato, no qual tentamos fazer sentido de nós mesmos (Dennett, 1981c).

Em alguns casos de epilepsia, é necessário separar os dois hemisférios do cérebro, cortando o corpo caloso, que os conecta. Isso faz com que haja muito pouco intercâmbio de informação entre eles; impressionantemente, *isso não é perceptível* em condições normais. Mas cientistas são capciosos e descobriram que fazendo com que a pessoa olhe para um ponto fixo central, toda a informação visual que chega a um lado

dos olhos é conduzida apenas ao hemisfério cerebral oposto. Na maioria das pessoas o hemisfério esquerdo é o principal processador da fala, e isso significa que, num experimento, o sujeito não consegue falar a respeito daquilo que fica do lado esquerdo do ponto de foco (porque o controle é cruzado). O mesmo vale para as mãos. Se colocarmos uma palavra do lado que não pode ser acessado pelo centro de fala e pedirmos para que a pessoa faça um desenho só com informação visual, a mão faz um desenho. A pessoa, no entanto, não sabe explicar porque fez aquele desenho, então prontamente *inventa* alguma razão. É possível misturar palavras e imagens de diversas maneiras para demonstrar tanto a quase independência dos hemisférios, quanto nossa compulsão por, *a posteriori*, inventar razões e justificativas para aquilo que achamos que fomos *nós* que fizemos. Torna-se difícil não querer concordar com Dennett, ao ver esses experimentos, em dizer que o Self é determinado *posteriormente*, de acordo com esses atos de justificativas que constituímos sobre nossas ações, e não anteriormente, como um ectoplasma racional capaz de tomar decisões unas e com nexos. Aos mais curiosos, cabe procurar na internet por experimentos usando as palavras chaves “*split-brain experiment*”, que provavelmente renderá alguns vídeos interessantes. O nome do cientista que estuda esses casos a mais tempo, Michael Gazzaniga, da Universidade da Califórnia, talvez seja útil também.

Há diversas outras evidências, de diferentes áreas do conhecimento, de como muitas de nossas ações são pouco conscientes, e de como temos uma tendência a nos justificar mesmo quando fazemos algo totalmente sem sentido.

Temos então uma dificuldade a mais na hora de criar uma noção de Self para nós mesmos. O Self não é tão senhor de si quanto possa parecer. Com efeito, ele não parece ser o responsável por nossos atos, mas mais uma *forma de falar de nós mesmos* que encontramos, através da qual nos entendemos melhor, nos justificamos melhor, uma espécie de figura construída para nos auxiliar a ser consistentes. Digamos, para uma figuração intuitiva, que um monte de formigas esteja caminhando por um gramado, com seus detectores de cheiros, fótons (olhos), e outros estímulos; eventualmente algumas encontram uma árvore e começam a trazer folhas de volta para o formigueiro. Elas não precisam saber o que estão fazendo, apenas o fazem. Elas tem competência sem compreensão. Com o tempo, outras tomarão o mesmo rumo, e surgirá na grama entre o formigueiro e a árvore um caminho. Esse caminho não foi planejado por nenhuma formiga mestre, mas foi emergindo conforme mais e mais formigas usavam a mesma rota para o mesmo objetivo. Cada vez que uma formiga passava por ali, estava criando a

pequena história daquele caminho. O caminho como um todo não é nenhuma daquelas histórias particulares. Também não é a soma daquelas histórias, porque as áreas que foram usadas poucas vezes não ficam marcadas na grama. A área marcada na grama é aquela que ficou, de alguma maneira, marcando a parte mais importante da soma das várias histórias das formigas. O único problema dessa figuração é que o caminho ainda é uma coisa muito física e causal para se adequar ao papel do Self que Dennett pretende. Dennett se utiliza da metáfora do Self como *centro de gravidade* das narrativas que nos constituem; isso porque os centros de gravidade são entidades existentes, mas são *abstracta*, são ferramentas de auxílio de cálculo para entender alguma coisa, mas não são constituídos de matéria ou energia. Falaremos mais sobre os *abstracta* adiante. O Self, para Dennett, é o centro de gravidade narrativa das múltiplas histórias que constituem um indivíduo. É a abstração facilitadora de cálculo que utilizamos para pensar aquele indivíduo, e que ele usa para pensar a si mesmo.

Vejam como essa noção de Self proposta por Dennett, do Self como centro de gravidade narrativa (Dennett, 1992), é capaz de lidar com as dificuldades que possam surgir. Toda concepção de conceitos abstratos está fadada a levar a alguma dificuldade de clareza, alguma nuance de intuição ou entendimento que pode, por assim dizer, levar-nos a duvidar do que está sendo dito. Uma boa metáfora não é apenas representativa e ilustrativa de um conceito, mas é capaz de defletir as possíveis críticas ou incompatibilidades que o conceito possa apresentar. Listarei na sequência aquilo que pareça problemático ou incompatível na noção de Self da qual nos servimos usualmente (afora a vagueza do conceito), e indicarei aquilo que é solucionado ou simplesmente evapora enquanto problema quando adotamos essa visão do Self como *abstracta*.

Problema Um) A falsidade do Homo economicus: Desde o século dezanove, uma meme particular a respeito do que é um agente econômico permeia estudantes de economia pelo mundo. Nessa concepção, o indivíduo é pensado como um ser monolítico e puramente auto-interessado, um Self que é senhor de suas razões, podendo escolher livremente, dentro de seus conhecimentos, entre diferentes decisões (e produtos), maximizando seus benefícios. Mas não é assim que de fato os agentes operam.

Resposta Um: Uma vez que passarmos a olhar o Self como um centro de gravidade de várias histórias de hierarquias de fama, que disputam na história de um indivíduo, tanto no espaço de funcionamento de seu sistema cognitivo, quanto espalhados no tempo em suas várias anedotas, podemos ver duas coisas. A primeira é que mesmo que cada anedota fosse internamente consistente, ainda haveria espaço para contradição entre elas, então poderíamos fazer dois raciocínios incompatíveis que demonstrariam uma “falta de razão” do todo. De fato é isso que observamos, o comportamento humano é sistematicamente inconsistente e variável (Kahneman 2011). A segunda é que, mesmo dentro de cada uma dessas anedotas cognitivas que são as pequenas histórias que constituem nossa vida mental, existem coalizões neurais, hierarquias “disputando” espaço no nosso inventário mental, e cada uma das anedotas pode ela mesma sair aquém das expectativas de uma Razão Pura e Perfeita.

Dois) O Self é aquilo que nos unifica em diferentes momentos; eu sou o mesmo hoje, ontem e daqui a dois anos *porque* meu Self se mantém. Ainda assim, as opiniões daquele sujeito que tinha meu nome, caligrafia e impressões digitais dois anos atrás são bastante idiotas – algo parece estar errado.

Resposta Dois: Um centro de gravidade organiza, *lato sensu*, a matéria ao seu redor. Pensemos o Self como uma espécie de adaptação organizacional do organismo humano para que surja um grau superveniente de univocidade entre seus diferentes momentos. Pensemos que a própria ideia de Self permanece justamente para que *nos obriguemos* a centralizar nossos acontecimentos sob um único emblema, etiquetar-nos como aquela uma pessoa. Existem várias razões para que uma adaptação como essa surgisse, evolutivamente; a principal delas é que um Self como sujeito persistente no tempo é algo no que se pode confiar: “Essa pessoa na minha frente agora é a mesma que me prometeu há um ano me devolver um favor, e agora que eu, que sou o mesmo homem de um ano atrás, estou precisando de um favor, posso ir até ela e pedir que pague sua dívida”. Tornar-se confiável, por sua vez, é estrategicamente bom quando sua espécie vive em grupos, e é um meme “bola de neve”, que se sustenta e se fortifica conforme se manifesta. De maneiras muito diferentes, antropólogos como Richard Boyd e filósofos como Adorno & Horkheimer (1985) discutiram essa questão da confiança intertemporal como determinantes do sujeito.

Três) Roubo aqui um problema e uma resposta do livro *Freedom evolves* (Dennett, 2003b, p. 252): “Por que mesmo estou em pé na cozinha, em frente ao armário de porcelana? Eu sei que estou onde eu deveria estar, mas o que eu vim pegar mesmo?”

Resposta Três: “Num momento como esse, *Eu* perdi o contexto, e portanto a *raison d’être* (razão de ser) desse pensamento, essa experiência consciente, e esse significado (e isso é o mais importante), estão temporariamente não mais acessíveis para *mim* – o grande mim que faz planos – do que estão acessíveis a qualquer terceiro, qualquer observador ‘externo’ que viesse observar. De fato, algum observador poderia muito bem estar apto a me lembrar o que eu queria fazer. Minha capacidade de me recordar é crucial, já que é apenas ela que me convenceria que esse observador está certo, que isso era algo que *Eu* estava fazendo. Se o pensamento ou plano pertencia a alguém, ele era meu – ele pertence ao eu que colocou em movimento e proviu o contexto no qual esse pensamento faz sentido; ocorre apenas que a parte de mim que está bestificada está temporariamente inepta a ganhar acesso a outra parte de mim que é o autor desse pensamento.”

Com essas linhas gerais, podemos delinear onde o Self de Dennett, isto é, aquilo que unifica a obra de Dennett ao longo do tempo, quer chegar com essa ideia (¿ideias?) a respeito de como podemos pensar a noção de “Self”. Partimos agora, por livre e espontânea inevitabilidade – ¿ou você consegue mudar as palavras desse texto? – à discussão sobre o livre-arbítrio.

Livre-arbítrio

Você não pode mudar as palavras desse texto, mas pode decidir ignorá-las se quiser, pode não olhar para elas. ¿Será? Há quem defenda que não, que todas as nossas decisões já estão mecanicamente pré-determinadas desde o início dos tempos (ou desde alguns milissegundos depois do Big Bang, ao gosto do freguês). Há aqueles que dizem que nossas ações são deliberadas no cérebro alguns décimos de segundos antes de estarmos conscientes de nossas decisões, e que portanto não somos nós que decidimos nada. Há os que simplesmente defendem que temos sim o poder de decisão (mesmo que constrita em alguns sentidos) sobre aquilo que fazemos; e finalmente há alguns que

creem que a solução da questão esteja pendendo sobre a mecânica quântica, que salvaria o *indeterminismo*, e portanto nos daria a possibilidade de decidir por nós mesmos o que, doutra maneira, seria sujeito às leis da motricidade newtoniana, tal qual uma bola de capotão.

Uma série de memes foram associados à palavra ‘livre-arbítrio’ ao longo do tempo. Muito parece depender de termos ou não livre-arbítrio, das mais profundas questões morais humanas aos mais severos atos de punição, o livre-arbítrio (como se isso fosse um só e não dezenas de memes!) é sempre um ator presente na discussão, e sua existência é sistematicamente posta em cheque. Precisamos separar as definições de livre-arbítrio para podermos eliminar aquelas que são apenas fantasmas da imaginação e aquelas que de fato podem auxiliar nosso entendimento de nós mesmos. Uma vez que tenhamos feito isso, passaremos a nos perguntar a respeito *desses* livre-arbítrios que vale a pena perseguir, se eles nos dão ou não a sustentação moral que costumamos esperar e desejar do conceito, e também se são compatíveis com nosso sistema de responsabilização e punição de agentes da sociedade.

Antes de falarmos separadamente de diversas vertentes do uso desse termo, disponho aqui de uma transcrição feita por Dennett (2003b, p. 134) de um argumento muito comum que defende a incompatibilidade do livre-arbítrio com o determinismo:

[1] Se o determinismo é verdade, se eu Vou ou Fico é completamente fixado pelas leis naturais e eventos no passado distante.

[2] As leis naturais não dependem de mim, nem o passado distante.

[3] Portanto, se eu Vou ou Fico é completamente fixado por circunstâncias independentes de mim.

[4] Se uma ação não depende de mim, ela não é livre (no sentido moralmente relevante).

[5] Portanto, minha ação de Ir ou Ficar não é livre.

Esse argumento é bastante forte para aqueles de inclinação determinista, e por essa razão mesmo raramente costumam engajar-se em demasiadamente extensos debates sobre o tema. Pessoalmente, eu adoraria poder discutir extensamente se o mundo é ou não determinista, mas isso fugiria bastante dos propósitos desta dissertação. Por ora, peço a gentileza ao leitor de assumir que o mundo é deterministicamente determinado para que possamos seguir viagem. O argumento de Dennett, com efeito, pretende sugerir que *mesmo que o mundo seja determinista*, ainda assim existe espaço para variedades de livre-arbítrio desejáveis. Ou seja, estamos assumindo o

determinismo para que a brincadeira fique mais difícil. Como um exercício ao simulador, sugiro procurar dentre as concepções a seguir aquela que mais pareça vir do pensamento dennettiano; este será nosso último exercício antes de uma tentativa mais explícita de simulação, que sugerirei mais adiante:

[1] Imagine algum tipo de oscilador, como um pêndulo. Digamos que nossa vontade seja determinada pelas leis da física, e que tenhamos de decidir entre ir para a esquerda e para a direita. Se nossa vontade estiver variando da mesma forma que o pêndulo, então nossa decisão foi livre caso tenhamos tido tempo de decisão o suficiente para que o “pêndulo da vontade” estivesse rumando em cada um dos sentidos ao longo desse tempo decisório.

[2] O livre-arbítrio é a vontade livre e normalmente racional da alma humana, um ectoplasma semi-gasoso que habita nossos corpos e interfere no mundo material.

[3] A liberdade humana está na nossa capacidade de interferir no mundo material com nossos pensamentos e deliberações.

[4] Estamos determinados a fazer tudo o que fazemos e o livre-arbítrio é um conceito mal formulado.

[5] O livre-arbítrio só é possível se existir indeterminação quântica, que interfira de maneira significativa no nosso processo de pensamento, a ponto de deixar em aberto qual dentre as opções acabaremos por executar.

[6] Nossa liberdade é maior do que a dos demais seres vivos porque fomos arquitetados para conseguir evitar mais coisas; nossos dispositivos de evitabilidade são os responsáveis pela existência do nosso livre-arbítrio.

[7] Se nossa experiência consciente acontecesse antes das tomadas de decisão, teríamos livre-arbítrio, mas como a neurociência consegue prever *antes* dos relatos conscientes das pessoas para que lado elas vão mover uma alavanca, por exemplo, não há livre-arbítrio.

A memória de trabalho média do ser humano é de sete dígitos, e talvez seja demasiado demandar do leitor que mantenha em mente sete *definições* de alguma coisa, por isso paro por aqui.

Espero que nesse ponto já seja bastante fácil eliminar as opções [2] e [3], que presumem nossa não materialidade, bem como a opção [5], pois até agora não vimos nenhum ponto em que a física quântica interferiu no pensamento de Dennett, e não podemos nos esquecer que estamos *presumindo* que o determinismo seja verdadeiro. Genericamente falando, vale pontuar que, diferentemente das pessoas comuns, parece que a maioria dos filósofos não se satisfaz mais e melhor com a ideia de que seus atos sejam determinados por resultados aleatórios do que por resultados mecanicamente determinados. É difícil, aos de nossa profissão, entender de que maneira ser determinado por resultados de dados seria diferente, moralmente falando, de ser determinado por puxões e empurrões newtonianos, mas talvez isso seja um defeito de mercado desses profissionais.

Sobraram-nos a opção do pêndulo variando no tempo [1], do conceito mal-formulado [4], da inevitabilidade que se desenvolve em seres complexos [6] e do livre-arbítrio somente quando consciente [7]. Convido mais uma vez o simulador a pensar sobre qual dessas alternativas se incorpora mais na forma de pensar que temos visto até agora, e que não se esqueça de que, no geral, Dennett tem por hábito modificar os conceitos e refiná-los, não eliminá-los por completo.

Como existem muito mais do que 7 conceitos de livre-arbítrio, é bastante estranho, e porque não dizer *arbitrário*, dispensar completamente o conceito como sendo mal formulado. Sem dúvida há versões mal formuladas dele, versões que são por excelência inconsistentes, mas há também versões que podem ser exploradas, e, ao menos com o material que vimos até agora, vale a pena explorar as três remanescentes de nossa lista. Façamo-lo.

Suponhamos que a vontade de um sujeito esteja distribuída no tempo, oscilando entre as decisões de ir e ficar, sendo que o momento em que ele agiu seja o momento 5. Se tivesse decidido no momento 4, teria tomado outra decisão, no momento 2, a mesma. Desenhe o gráfico como queira. O que há de importante para esse tipo de visão, que não precisa ser simétrica como um pêndulo, é que haja uma janela de decisão na qual existam pelo menos dois momentos em que o que a pessoa faria seria diferente. É uma visão interessante, e sem dúvida compatível com o determinismo físico. O problema é que padece de uma arbitrariedade grande na decisão de qual seria o tempo de janela permissível. A tendência natural de alguém que adota essa posição é dizer que as duas opções precisam estar disponíveis *ao mesmo tempo*, do contrário não existe escolha verdadeira, apenas algum tipo de pseudo-escolha. Então quando somos obrigados,

dentro de um paradigma determinista, a criar uma janela, vamos querer criar a menor janela possível, talvez algo como o tempo que a luz demora para percorrer um metro. É evidente que esse tipo de estratégia não funcionará, dado que os processos relevantes, os ‘alístamentos’ de legiões de neurônios que tomam forma, duram muitas milhares de vezes mais tempo do que isso. Ainda que possamos olhar para essa concepção e ver o que há de interessante nela, temos de contentar-nos com grandes janelas de escolha para que realmente estejamos falando de pessoas, fazendo escolhas reais, em tempo real. Tanto quanto Descartes desejava colapsar todo o *espaço* do centro de comando do cérebro para a glândula pineal, existem aqueles hoje em dia que desejam colapsar todo o *tempo* de decisão para *um instante*, o que quer que isso seja.

Como eu nunca canso de pontuar, todo o trabalho feito pelo imaginado homúnculo no teatro cartesiano tem de ser dividido e distribuído no espaço *e no tempo* no cérebro. É mais uma vez hora de repetir meu mote irônico: Se você se fizer realmente pequeno, você pode externalizar virtualmente tudo [*externalize virtually anything*]. (Dennett, 2003b, p. 122.)

O mote de Dennett é importante porque destaca uma forma geral de erro ao pensar a filosofia da mente: essa estratégia de fazer-se cada vez menor. Vimos como essa estratégia falha ao abordarmos a opacidade dos significados, isto é, como a quantidade de contexto, circunstâncias, e o tamanho aceito como parte do agente modificam os significados de suas atitudes proposicionais. Vimos como a noção de Self, como um centro puro da razão, colapsa vontades demais sob uma mesma bandeira. Vimos como encapsular a inteligência num centro de comando como a glândula pineal não pode funcionar. Essa estratégia em geral não é boa, e é por isso que Dennett insiste que não devemos nos fazer realmente pequenos, não quando temos propósitos explicativos que se pretendam frutíferos e compatíveis com nosso conhecimento da ciência atual.

Suponhamos então que escapamos da micro-janela que impede uma visão interessante sobre o livre-arbítrio. Ou seja, estamos livres agora para distribuir um pouco da vontade no tempo, e ganhamos talvez 100 milissegundos como nosso tempo de decisão. Um décimo de segundo parece até razoável, num primeiro olhar, para que tomemos uma ou outra decisão. Mas aí encontraremos problemas com a neurociência, ¿ou aparentes problemas?

Vários experimentos em neurociência, em particular os de Benjamin Libet⁴², ficaram famosos por demonstrar que não temos liberdade nas nossas escolhas. Isso pelo seguinte: se colocarmos escaneadores de cérebro (eletroencefalograma, imagem de ressonância magnética funcional) numa pessoa e dissermos para ela mexer uma alavanca *quando ela quiser*, e registrarmos as mudanças que ocorrem no cérebro durante o processo, podemos utilizar essas informações para, futuramente, *prever* com antecedência, apenas a partir do padrão informacional, a hora em que ela vai decidir mexer a alavanca.

Uma metodologia de contagem do tempo bastante popular é uma na qual uma bolinha gira (como se fosse ao longo de um relógio) numa tela de computador. Desde o início a pessoa está acompanhando a bolinha na tela. Quando a pessoa decide conscientemente, ela está olhando essa bolinha, e diz para o pesquisador *onde ela estava no momento da decisão*, após apertar a alavanca. Com isso, se pretendia dizer que o momento em que a bolinha estava naquela posição é o momento no qual o sujeito se tornou consciente da decisão. Ficamos (de acordo com a evidência experimental) com uma linha do tempo assim.

Início → informação suficiente para a máquina dizer que “é agora” → momento em que o sujeito diz (após o experimento) que decidiu, medido pela percepção da bolinha → movimento da alavanca → fim do experimento.

Por conta de a informação suficiente para a decisão da máquina ocorrer antes do momento declarado pelo sujeito como a posição da bolinha, quando ele decidiu conscientemente, parece ser o caso que nosso cérebro determina nossas decisões antes da tomada consciente da decisão, e nós apenas “assistimos”, por assim dizer.

O mesmo fenômeno pode ser observado quando a máquina utiliza seus algoritmos baseados no passado para descobrir (dois ou três décimos de segundos antes) se a pessoa vai virar a alavanca para *frente ou para trás*. Não apenas é possível prever *que* a pessoa vai decidir, mas também *para que direção* ela decidirá.

Más notícias para os libertários, talvez não possamos dar a eles o que eles desejam. Afinal, qual o poder da filosofia diante de uma evidência *experimental* de que não possuímos livre-arbítrio. O poder de correção. O experimento na realidade tem erros estruturais importantes, e trago à tona novamente um *intuition pump* já utilizado antes, para depois traçar um comparativo.

⁴² O artigo principal, que encabeça uma cascata de equívocos há três décadas, é Libet (1985).

Percepção de tempo na consciência

Recordemo-nos do intuition pump IV, do capítulo 8, que trata do momento exato em que o Império Britânico ficou sabendo do fim da Guerra de 1812. Trata-se de um paralelo com os experimentos que pretendem determinar o *momento de consciência*. O problema com esses experimentos é que eles partem do pressuposto que existe um momento em que surge a consciência, assim como haveria um momento em que o Império Inglês, espalhado pelo mundo, teria ficado sabendo do fim da guerra. Se você se fizer realmente pequeno, pode externalizar virtualmente tudo. Não é possível *ao mesmo tempo* se fazer pequeno e manter dentro de você o momento de aquisição da consciência particular de onde a bolinha estava.

Vale pontuar, de início, que mesmo que a luz não demore quase nada para chegar ao olho do sujeito, existe um tempo significativo entre o momento no qual ela bate na retina e o momento no qual o processamento da informação visual se dá no córtex visual. Outro tempo significativo entre a informação estar no córtex visual e correntes de informação se direcionarem para outras áreas do cérebro, como o lobo frontal, que é correlacionado com processos de raciocínio e decisão. Continuando a batalha por fama, as correntes neuronais que estão carregando essa informação vão para diversas áreas, inclusive a área motora, que possibilita que o sujeito *diga* (com a boca!) que experienciou a bolinha no ponto X.

Notavelmente há uma grande quantidade de possibilidades de má informação nesse processo. Faremos um tour pelo cérebro examinando como essa informação declarada pode ser diferente da dada pelo sujeito. Dennett está mais preocupado com o aspecto geral da possibilidade de informação cruzada, e não foi até a literatura de neurociência cognitiva (que ainda não existia em 1991, quando ele a escreveu) para ter certeza de que sua hipótese estava certa. Eu fui, então apresento aqui um resumo da literatura mais atualizada (Horton & Sincich, 2004; Romo et al., 2004).

O córtex visual (a parte de trás do cérebro, onde a visão é processada depois que a informação passa pelos nervos ópticos) é dividido em várias áreas, e cada uma delas tende a disparar em coalizão, em função de um aspecto do que se vê. Na área inicial, V1, qualidades mais superficiais, de base na hierarquia, tendem a correlacionar melhor com o disparo; por exemplo, a informação de uma linha inclinada ou reta, horizontal, a cor, e o movimento.

Quando utilizar a expressão “responsável por x”, o que quero dizer é “cujos disparos em média correlacionam com x”.

Há um grupo de células (konio) responsável pelo eixo azul/amarelo e um grupo (parvo) responsável pelo eixo verde/vermelho. Desses dois grupos saem extensões (axônios) que se dirigem a camadas dentro da V1. Todas as áreas do córtex se dividem em seis camadas, das quais a 4 é sempre a mais importante, e seus braços polícticos são mais compridos. Acreditava-se que áreas diferentes dentro da V1 recebessem informação de tipos diferentes separadas, e enviassem separadas, mas verificou-se que os três tipos canônicos de informação, cor, movimento e forma, são projetados para quase todas as camadas da área V1. Delas, existem projeções para a área V2, que constituem a maioria, e umas poucas, somente da área 4 na V1 para a área MT, motora.

A área V2 se divide entre faixas finas, pálidas, e grossas. As faixas fina e pálida enviam informação para a área V4, onde o processo continua, enquanto a faixa grossa envia informação para a mesma área MT, motora.

O zig-zag de alistamento neuronal continua, agora já demasiado difícil de traçar e não mais considerado como processamento do córtex visual. Um ponto interessante sobre essa mudança de perspectiva sobre o sistema visual é que o modelo anterior, dividido em partes discretas e cada um com uma função, era mais fácil de decorar, ao passo que o novo tem projeções demais. Faço aqui uma previsão memética que até 2020 a maioria dos neurocientistas ainda falará no modelo antigo, bem dividido, como atual.

Tudo isso tem de ocorrer depois que o sujeito vê a bolinha. Dennett sugere que a informação sobre o tempo de algo não precisa ocorrer ao mesmo tempo do que o algo. Por exemplo, uma carta do Império pode vir com *data* e portanto o receptor da carta sabe que (caso o emissor seja sincero) ela foi escrita naquele dia, mesmo que haja chegado a ele muito tempo depois. Assim sendo, é possível que a informação de horário esteja codificada num padrão de ativação, e portanto o tempo real e o tempo experienciado sejam totalmente diferentes. Quando Dennett escreveu, em 1991, só se tinha certeza disso sobre a percepção, por exemplo a diferença de percepção de tempo de algo que nos toca nos pés e algo que nos toca na bochecha. Sabia-se que o cérebro não “espera” a informação do pé chegar para reagir, caso haja algo a se reagir na bochecha. Ele não precisa centralizar a informação em um único tempo newtoniano e juntar todas as áreas sensoriais. A evolução não faria um atraso desses com um indivíduo, porque os outros indivíduos, que não tivessem esse atraso para computar tudo ao mesmo tempo, e que fossem mais rápidos em reagir a estímulos na cabeça, por

exemplo, seriam mais aptos a sobreviver. Nossa experiência do tempo é portanto desconexa temporalmente.

É interessante notar que bastante depois de Dennett se utilizar dessa ideia para falar sobre a experiência de *sentir* o tempo, ele utiliza da mesma ideia para falar da experiência de *decidir no tempo*. No livro *Freedom evolves* (Dennett, 2003b), ele desenvolve um argumento semelhante de como suas percepções e decisão não ficam numa fila no cérebro para passar pela *central de comando* do teatro cartesiano, mas que cada parte reage conforme a informação que lhe é dada, assim que lhe é possível. Não faz sentido nos reduzirmos a um ponto cartesiano no meio (no canto, ou oscilando ao longo) do cérebro, pelo qual hão de passar nem nossas memórias, nem nossa experiência de tempo, nem nossa faculdade racional. Tudo isso ocorre simultaneamente em grandes coalizões que não sabem umas das outras, mas que estão entrelaçadas da maneira certa (pela mãe natureza) para fazer-nos agir exatamente como agimos, muito bem para focar, e não tão bem para cálculo algébrico. Se você se fizer muito pequeno, pode externalizar virtualmente tudo, e isso não é desejável.

A sugestão de Dennett, de que a informação sobre um *quando* poderia ser enviada (ao invés de retroativamente calculada a partir da distância do estímulo) como numa carta, recebeu confirmação experimental em Romo et al. (2004). Nesse artigo, os autores confirmam que as áreas sensoriais de macacos alteram a forma de processamento da informação de frequência ao longo da jornada pelo cérebro. Os macacos são treinados para diferenciar frequências diferentes de toques nos seus dedos. De início, como é de se esperar, há um estímulo físico na extremidade do macaco, isto é, no dedo; vemos então que a frequência gera impulsos nervosos de mesma frequência, que percorrem sua jornada até o cérebro, onde chegam à área S1. Passam várias vezes de um neurônio para outro, ao longo da coluna, e também numa área chamada núcleo posterior ventrolateral do tálamo. Até a área S1, os disparos ainda estão ajustados na mesma frequência do estímulo, mas no processo de transmissão entre a área S1 e S2, a informação é *codificada*, transformada em outra linguagem, e a partir daí o que é correlacionado com a frequência do estímulo é o número absoluto de disparos, e não mais a frequência de disparos. Houve uma modificação em como a informação é processada, deixando o registro de tempo tornar-se independente do tempo, e dependente do número absoluto de disparos. CDQD (Como Dennett Queria Demonstrar).

Talvez ainda reste algum chauvinista que diria que isso só demonstra que os macacos não poderiam saber o ponto no qual a bolinha está no experimento de Libet; humanos conseguiriam, mas o sistema de codificação sensorial é bastante parecido em nós e em nossos primos, e não temos razões para crer que, ao longo do tempo evolutivo, a mãe natureza tenha decidido *abandonar* essa ótima estratégia de codificar informação, depois de já tê-la descoberto. Noutras palavras, a vizinhança no espaço de design de um indivíduo (humano, macaco ou intermediário) que codifica o tempo é uma ótima vizinhança, e qualquer recuo ao estágio no qual isso não ocorre tende a ser uma piora. A similaridade entre nós e os macacos e esse racional evolucionário são suficientes para nos retirar essa preocupação.

Isso elimina a hipótese de que a neurociência contemporânea refutou o livre arbítrio. Somos deixados com apenas uma última hipótese:

[6] Nossa liberdade é maior do que a dos demais seres vivos porque fomos arquitetados para conseguir evitar mais coisas, nossos dispositivos de evitabilidade são os responsáveis pela existência do nosso livre-arbítrio.

Esse é nosso grande vencedor, essa é visão defendida por Dennett em *Freedom Evolves*. Note que a única razão pela qual só sobrou essa alternativa é porque eu escolhi colocar apenas essas sete. Eliminar todas as outras não é razão suficiente para concordar com essa, a ciência e a filosofia não são uma prova de teste, e sempre há a possibilidade de que todas as hipóteses já apresentadas estejam erradas. Devemos agora considerar nossa hipótese final. Precisamos ver como ela é capaz de lidar com o problema. Em primeiro lugar, sabemos que essa visão não colapsa o indivíduo nem no tempo nem no espaço, pecados mortais se quisermos ter uma teorização da liberdade compatível com um mundo em que duas coisas não podem estar ao mesmo tempo no mesmo lugar (entrelaçamento quântico à parte). Esse sujeito “espalhado”, por assim dizer, pode fazer tudo aquilo que queremos, tomada de decisão, raciocínio, deliberação, emocionar-se; ao menos ele tem espaço para tais tarefas, porque é uma mente incorporada, dispersa no espaço e portanto capaz de realizar tarefas complexas.

¿Qual é o tipo de coisa que nos faria reprovar alguém num teste de liberdade? Um caso seria o de uma zebra caindo num poço de elevador. Não costumamos achar que uma zebra caminhando no escuro em direção a um poço de elevador tenha tido escolha de ir para cima ou para baixo quando, finalmente, provavelmente contra sua vontade, ela atingiu o elevador. Esse é um caso de falta de escolha, mas muito simples.

Mais interessante é o caso da mosca *Sphex*, que traz um inseto até a entrada de sua toca, deixa-o parado do lado de fora, entra para fazer uma verificação de rotina no interior, sai, e arrasta-o para dentro. Experimentadores deslocaram o inseto um pouco para o lado, enquanto ela entra para “verificar” dentro de sua casa se tudo vai bem. Quando ela sai e o encontra, vem até a entrada da toca novamente, deixa-o, e faz uma verificação de rotina dentro da toca. Ao sair e encontrar o inseto no lugar deixado pelo experimentador novamente, ela arrasta-o para a entrada da toca, e entra para fazer uma verificação, e assim por diante, até incríveis 40 repetições do mesmo comportamento. Isso não é o que chamaríamos de uma liberdade que valha a pena possuir. O grau de determinismo genético agindo sobre o ato da mosca é demasiado grande, mesmo que de início pudesse parecer inteligente. Ela está simplesmente programada, seu cérebro está entrelaçado para agir de uma maneira mecânica e ritualizada, não existe a perspectiva de agir de outra forma, ela não pode evitar entrar na toca, mesmo que já o tenha feito 20 vezes – um comportamento que lembra um pouco o hábito de alguns humanos com relação a geladeiras.

Existem alguns comportamentos ritualizados em humanos, e somos em algum sentido obrigados geneticamente a fazer algumas coisas, mas elas são poucas em relação àquilo que consideramos moralmente relevante ao discutir a questão da liberdade. Nossa vontade nos permite, por exemplo, sermos mais ou menos egoístas, mais ou menos contemplativos, e mais ou menos conhecedores da última coleção de inverno Dolce Gabbana. Uma das mais falsas concepções dos economistas é a ideia de que somos agentes que buscam maximizar apenas o próprio prazer, ou utilidade, marionetes de uma racionalidade cuja única função seria fazer com que nosso carro seja mais luxuoso do que o do vizinho. Isso pode ser verdade no mundo dos genes, mas não no ambiente vasto e complexo construído pelas alianças de genes e memes que é o mundo que habitamos.

O altruísmo evoluiu muitas vezes, de maneira radical em diversas delas. Insetos sociais são um caso clássico em que todas as irmãs se dedicam a uma irmã especial a tal ponto que sequer chegam a tornar-se reprodutoras. Mas muito antes disso, com a invenção da pluricelularidade, um sistema análogo fora descoberto como estratégia. Todas as células de nossos corpos são descendentes do primeiro ser vivo que houve na Terra, muito provavelmente. Cada uma delas viveu até hoje das repartições constantes e incontáveis daquele primeiro replicador. No entanto, a esmagadora maioria delas está fadada à morte ainda no tempo de sua vida. As células somáticas, que não são da

linhagem reprodutiva, e não vão virar gametas, estão perto do fim do ciclo de suas vidas, tal qual as formigas operárias não deveriam ter perspectivas de deixar descendentes. Isso porque elas fazem parte de uma grande coalizão, que ao longo de milhões de anos especializou-se em permitir que você tenha filhos, reiniciando o processo. A ideia de ser altruísta parece biologicamente inaceitável à primeira vista, mas bastante razoável ao pensarmos que o que há de ser preservado é um código genético, e que esse código está copiado em cada uma das células (75% dos genes são iguais entre formigas irmãs, 100% entre células do mesmo humano⁴³). Quando isso acontece, torna-se uma boa ideia ajudar aqueles que parecem com você ao seu redor, inclusive especializando-se para tornar-se, por exemplo, tecido cerebral.

Quando existe algum grau de viscosidade numa sociedade de agentes, ou seja, quando nem todos têm contato com todos, vale a pena ser altruísta com aqueles que estão próximos a você também por outra razão, porque ele pode ser altruísta de volta. Ou seja, temos altruísmo por *kin selection* (seleção parental, devido ao compartilhamento de genes) e também altruísmo recíproco, devido à alta probabilidade de que aquele que eu alimento hoje tenha a comida da qual preciso amanhã. Isso, no entanto, provoca a possibilidade do nascimento de traidores, aqueles que aceitam o altruísmo alheio sem nunca fornecer nada em troca. Os modelos de evolução de agentes computacionais que foram desenvolvidos de início chegavam nessa barreira, não havia como compatibilizar o altruísmo real que existe no mundo com essa ideia de que qualquer gene para trapacear sairia vencedor, e em algumas gerações dominaria o comportamento do grupo. Até que surgiu a ideia de inserir na simulação a punição de agentes com altruísmo não-seletivo. Ou seja, aquele que não detecta a quem está fazendo bem é punido pelo grupo, de tal maneira que o altruísmo real volta a ser uma estratégia que vale a pena; podem surgir novos trapaceiros mais sofisticados, e surgirão altruístas ainda mais sofisticados, e em algum ponto nessa batalha sem fim estamos nós, talvez os mais altruístas entre os mamíferos (McElreath & Boyd, 2007).

Estamos livres, nesse sentido, do egoísmo biológico que leva algumas espécies a comerem seus parceiros, ou filhos recém nascidos de matarem seus irmãos antes que tenham a possibilidade de nascer.

⁴³ Recentemente li *The superorganism* (Hölldobler & Wilson, 2009), o mais importante compêndio do conhecimento humano a respeito de insetos sociais. Na realidade, nem toda espécie social tem um sistema de compartilhamento de genes 75% entre irmãs, e é possível que uma espécie evolua a chamada “eusocialidade”, mesmo tendo um esquema genético pouco distinto do esquema humano. A descrição completa das condições para a eusocialidade pode ser encontrada nesse livro, pp. 31-42.

Expus aqui, anedoticamente, algumas liberdades que temos, a liberdade de sermos ou não altruístas, a liberdade de decidir quantas vezes verificamos se há alguém em casa antes de trazer as compras de supermercado, e a liberdade de construir elevadores para evitar o colapso gravitacional. Nossa inteligência é em vários sentidos a capacidade de *antever problemas* e evitá-los, *antes que aconteçam*. Isso é algo fascinante. Enquanto as criaturas darwinianas, como as bactérias, estão morrendo, e as criaturas skinnerianas respondem a alguns estímulos simples apenas, nós, do topo de nosso gregorianismo podemos nos unir a Karl Popper sob a mesma bandeira e evitar que morramos, assassinando nossas hipóteses em nosso lugar. Estamos fadados, é claro, a ser atacados de surpresa por várias coisas, como o colapso econômico ou um meteoro. Mas ainda assim, somando os sistemas desenvolvidos pelos genes, e os sistemas desenvolvidos culturalmente pelo homem, temos segurança de que muitas das coisas que outrora eram inevitáveis hoje passam longe das preocupações da maioria das pessoas. Somos capazes de evitar a fome, a varíola, as rugas e em grande medida o ataque de tribos inimigas. Temos soluções internas para uma miríade de problemas individuais, familiares e sociais. Evitamos uma visão empobrecida com o uso de óculos, evitamos o tédio com a televisão (uma solução problemática), evitamos o medo com sistemas de segurança, evitamos a paralisia infantil com vacinas etc.

Em suma, o ponto defendido por Dennett é que liberdade vem em graus, e o grau de liberdade do qual gozamos tem poucos paralelos no mundo conhecido. Individualmente, somos capazes de evitar quantidades enormes de problemas de uma maneira que para um cavalo seria um sonho. Nossa capacidade de pensar nos confere a possibilidade de nos salvar de animais peçonhentos, do frio, e principalmente da programação de sábado à noite da tevê aberta, ¿o que mais poderíamos desejar?

Antes de adentrarmos a consciência – privilégio que compartilharemos com a pequena parte da informação que entra em nossos cérebros, que se torna foco de atenção e *broadcast* – talvez caiba aqui citar a resposta de Dennett à revolta de alguns opositores quanto a ele parecer estar modificando os conceitos, e não explicando-os. Isso porque, se antes eu avisei que talvez tivéssemos a sensação de assistir a um palhaço ao ir ao show de striptease, uma insatisfação, mas não propriamente uma revolta, agora a revolta que está por vir será ainda maior. No livro do qual vem a citação, “Conrad” é um personagem fictício que representa, mais ou menos, a opinião tradicional.

[...] O que fazer com que as variedades “genuínas” aquelas com as quais vale a pena se importar afinal? Eu concordo que margarina não seja manteiga, independente de quão bom seja o gosto, mas se você insiste em manteiga de verdade a qualquer preço, você tem de ter uma razão.

Conrad: Aha! Você admite. Está apenas jogando com palavras, e tentando passar margarina como manteiga. Eu conclamo todas as pessoas a exigirem livre-arbítrio de verdade; não aceite substitutos!

E você também aconselharia diabéticos a insistir em insulina “real”, ao invés da coisa “artificial”? Se seu coração falhar um dia, você iria ridicularizar um substituto artificial que possa fazer todas as funções do seu coração real? Em que ponto o amor à tradição se torna uma tola superstição? (Dennett, 2003b, p. 225)

Consciência

“The fundamental cause of the trouble is that in the modern world the stupid are cocksure while the intelligent are full of doubt”

Bertrand Russell ⁴⁴

Consciência é um tema difícil. Não apenas isso, mas consciências são vários temas difíceis, com a dificuldade adicional de que todos eles usam a mesma palavra para designar seu objeto. Pior ainda! Alguns desses objetos são totalmente vagos, alguns simplesmente não existem e alguém se enganou ao pensá-los, e outros são incognoscíveis por definição. Não surpreende, ao vermos a quantidade de graus no qual uma teoria da consciência poderia falhar, que muitos considerem o problema da consciência o último dos grandes problemas não resolvidos pelo homem. Nesse subtópico tentarei clarificar um pouco melhor quais desses problemas Dennett irá problematizar e discutir. No próximo capítulo, farei uma exposição mais completa de sua teoria da consciência.

É evidente que, havendo vários conceitos engendrados pela mesma palavra, há igualmente vários problemas da consciência, e assim por diante. Essa é uma razão pela qual esse pode ter sido considerado um dos mais difíceis problemas, mas não é a principal. Existem muitas concepções de consciência, e vou emprestar do filósofo Ned Block (1995) algumas subdivisões interessantes. Vale notar que não é necessário que apenas uma dessas concepções tenha um referente que existe de fato; além disso, também não é necessário de cada uma que o referente exista. Uma função de conceituar consciência de várias formas é justamente poder ver, dentre essas, quais temos mais ou

⁴⁴ “O problema fundamental do mundo moderno é que os idiotas estão preenchidos de certeza, e os inteligentes cheios de dúvida” (Russell, [1933] 1998, p. 28).

menos certeza de existirem, e que tipo de evidências poderiam diminuir ou ampliar nossas crenças nessas diferentes noções. Aos conceitos:

Consciência fenomênica. O uso do termo “consciência” significando consciência fenomênica refere-se à experiência, é a mais difícil de definir, é aquela a qual nos referimos ao pensarmos que existe algo como estar no estado em que estamos (*what is it like-ness*), uma perspectiva de primeira pessoa, subjetiva. Há algo que é como é ver o azul do mar, e há algo diferente que é como é ver uma montanha verde. Esse aspecto de experiência é o que é chamado de consciência fenomênica por Ned Block. A maioria das preocupações dos filósofos, em particular o dito “problema difícil da consciência”, se refere a como conectar esse aspecto da consciência com o mundo físico no qual acreditamos viver. Supõe-se que a consciência fenomênica é algo quase sempre representativo, isto é, eu tenho um aspecto da experiência, como um matiz de vermelho, e ele é ao mesmo tempo representação do sangue que estou vendo, ele possui um aspecto de ser a respeito de algo, intencionalidade. Numa conversa de bar podemos perguntar a alguém se “¿o verde que você vê é o mesmo verde que eu vejo?” ou “¿será que eu vejo verde onde você vê azul?”. É necessário, no entanto, garantir que estamos falando de algo diferente da questão “¿ele é daltônico?”, isso porque o daltonismo é uma propriedade *detectável*. Note que quando eu pergunto se o verde que você vê é o vermelho que eu vejo, o *interessante* da pergunta é justamente que isso não parece detectável de nenhuma maneira! Não estou falando de se você é ou não capaz de perceber esta ou aquela amplitude de onda e atribuir a ela o nome ‘verde’ ou ‘vermelho’, estou falando da propriedade experiencial que você tem, e que eu não posso detectar, de ver aquela amplitude de onda como verde. A propriedade da qual estou falando é uma propriedade fenomênica, e por isso pertence à consciência fenomênica.

Consciência de acesso. Esse uso da palavra “consciência” é referente àquilo que está presente na transmissão de informação em larga escala (*broadcast*) do nosso comportamento racional. Aquilo que se manifesta em nosso comportamento como algo que estava atraindo nossa atenção, ou direcionando nossas ações, e que condiz com o que queremos fazer. A reportabilidade é uma propriedade essencial. Digamos que você está conversando com um amigo, enquanto ao longe toca um sino de igreja, mas você não está prestando atenção nisso. Durante a conversa você está acesso-consciente do que ele está dizendo, mas não das badaladas. Se ele lhe perguntar “¿quantas vezes o

sino tocou?”, e você não souber responder, isso significa que você não esteve acesso-consciente das badaladas. Há um caso interessante, aquele no qual na hora você não estava acesso-consciente do sino, mas mesmo assim, como que olhando para trás na memória, consegue retroativamente obter a informação de quantas vezes o sino tocou. Quando isso acontece, você estava se *tornando* acesso consciente conforme lembrava do número de badaladas do sino.

A razão pela qual consciência de acesso é entendida como racional é para retirar de seu escopo casos de “visão cega” (*blindsight*), condição na qual pessoas declaram que uma área de seu campo visual não consegue ver absolutamente nada, mas que, se obrigadas por um experimentador a chutar onde está uma bolinha naquela região, conseguem resultados bem acima da média. Nesses casos, a informação está sendo transmitida, mas não há reportabilidade de estar fazendo isso conscientemente, então eles não se configuram como consciência de acesso. Estar bastante focado em algo nos daria segurança de que alguém está acesso-consciente daquilo.

Consciência de self (*self consciousness*). Essa noção se refere apenas a ter uma concepção de self, e a habilidade de usar esse conceito ao pensar sobre si mesmo. Diversos primatas, por exemplo, conseguem reconhecer manchas em sua testa no espelho, e tentam removê-las, como dito por Steven Pinker (1997); consciência de acesso consiste basicamente em possuir um lugar para si mesmo no seu inventário mental. Por ser um conceito simples e uma propriedade obviamente empiricamente testável, não é parte de nosso foco.

Consciência monitora (*monitoring consciousness*). É frequente, ao usarmos o termo consciência, estarmos nos referindo a um ato introspectivo no qual estamos tentando observar nosso processo de consciência, um tipo de meta-consciência, observar-se pensando, ou algo assim. Como se tivéssemos uma percepção interna que pode ser ativada ou desativada, para percebermos o que estamos percebendo. Essa consciência poderia ter várias naturezas, por exemplo, poderia ser funcionalmente descritível e também ter um aspecto fenomênico. Com efeito, é ainda uma concepção bastante confusa, e não devemos aqui nos preocupar com ela mais do que citando-a, já que não será de fundamental importância no que se segue.

Essa distinção do conceito em quatro subconceitos é porque esse conceito é um conceito *mongrel* (vira-lata)⁴⁵.

As concepções contemporâneas de consciência e seu estudo são um tema grande, e a abordagem de Dennett sobre o tema possui cerca de 900 páginas entre livros, artigos etc. Dada essa magnitude, o próximo capítulo será inteiramente dedicado ao tema.

⁴⁵ Vale reforçar um ponto que sistematicamente escapa mesmo aos filósofos de maior poder de abstração. Quando utilizamos uma palavra, ela não se comporta como um símbolo interpretado de uma lógica formal, isto é, podendo ser interpretada apenas de uma única maneira. As palavras são utilizadas de diversas maneiras, e isso é o que gera o que se chama *mongrel concepts*, conceitos que são equivalentes a uma disjunção “x XOR y XOR z” (onde XOR é o “ou exclusivo”, significando “um desses, mas não mais de um”). Livre-arbítrio, *self* e consciência são casos clássicos disso. Cada uso da palavra “consciência” se refere a um desses quatro significados (ou subdivisões dos mesmos), caso seja um uso preciso.

Não confundir *mongrel concepts* (conceitos vira-lata) com *cluster concepts* (conceitos de agrupamento), os segundos sendo a visão defendida por Searle de que o que *define* um objeto é possuir uma *parcela* das propriedades de um conjunto definido, sem que nenhuma seja essencial. Não é impossível que alguma das subdefinições de consciência seja um *cluster concept*, na medida em que ele pode ser algo que se determina por ter a maioria de propriedades de um certo conjunto. Por exemplo, *self-consciousness* pode ser entendido como: “Possuir a maioria dos seguintes atributos: saber se reconhecer no espelho, pensar sobre si mesmo, diferenciar as ações próprias das de outros, usar teoria da mente para si próprio, pensar as próprias ações de maneira diferente das de outros”.

10. Consciência

Até muito recentemente, acreditava-se que não havia nada a fazer a respeito de impostos e da morte. Os impostos permanecem, mas há uma crescente corrente de cientistas e filósofos que acreditam que não mais é necessário assumir que “todo homem é mortal” é uma sentença verdadeira. Essa mudança ocorreu porque ao longo do século passado nosso aprendizado sobre os processos que organizam a vida cresceu mais do que em toda a história humana somada, e temos razões para crer que esse processo continuará cada vez mais rápido, ao menos por algum tempo. Enquanto a morte era uma certeza óbvia e não disputada por ninguém, era bastante razoável postergar indefinidamente a preocupação “¿como não morrer?” Afinal, até o ponto do qual tínhamos conhecimento, isso simplesmente era uma má questão, com quase a mesma força de “¿como não estar sujeito a gravidade?” O tempo no entanto tem o poder de modificar questões, transformando antigas más questões em tópicos fascinantes, e por vezes urgentes da ciência. O caso da morte é interessante e bastante atual, e eventualmente nos ajudará a ilustrar uma diferença fundamental no pensamento de diferentes filósofos sobre consciência. Existem atualmente métodos de, a partir de células tronco de humanos adultos e uma carapaça de cartilagem em forma de coração, construir um coração novo, que se organiza de maneira funcional e começa a bater sozinho, fora do corpo de qualquer indivíduo. Sabemos modificar o genoma de alguns animais a tal ponto que podemos colocar a forma de uma orelha humana nas costas de um rato, produzir gatos e coelhos fosforescentes, trocar a posição das patas e antenas de uma mosca, controlar alguns comportamentos de uma mosca por controle remoto. É difícil dizer que alguém acreditaria nisso em 1995, que dirá em 1900. A revista online EDGE todo ano pergunta a cerca de 150 cientistas de destaque (e um ou outro filósofo como Dennett) uma pergunta diferente. Em 2009 a pergunta era⁴⁶:

¿Que ideias fantásticas e desenvolvimentos científicos você espera viver para ver?

Raramente duas respostas coincidem numa pergunta tão ampla, e por isso me foi particularmente chocante ver como mudou a crença das pessoas com relação à mortalidade. Seis diferentes cientistas sugeriram que eles possam viver para conhecer

⁴⁶ Ver http://www.edge.org/q2009/q09_index.html

um tempo em que não mais envelhecemos, de uma ou outra maneira. Essa onda está começando, e a ideia está se espalhando. O que antes era considerado logicamente contraditório, agora é visto como possível, e até provável.

O que Dennett sugere com relação à consciência é que usemos por hora o mesmo método que utilizamos por tantos anos com relação a morte. Hoje em dia o problema da consciência chamado de “problema difícil” parece simplesmente ser incognoscível, contraditório ou irrelevante. O que Dennett sugere é que continuemos criando uma ciência cognitiva da consciência até que isso deixe de acontecer. Isso de maneira alguma significa abandonar o problema que descreverei em seguida, mas é uma proposta de metodologia para nos aproximarmos das questões relevantes, e finalmente descobriremos se, conforme atualmente formulado, o “problema difícil” é o ou não difícil, e se é ou não um problema.

Heterofenomenologia

Dennett se interessa por entender o que é a consciência, e sugere que a melhor maneira que temos de fazer isso é através do que ele chama de Heterofenomenologia, ou o estudo em terceira pessoa da consciência de uma pessoa. Fundamentalmente, como ele resume em seu *Sweet dreams* (2005), devemos procurar aquilo que nós e cientistas marcianos concordaríamos a respeito da consciência humana, e estudar esse tipo de aspecto. Para isso, ele pressupõe o uso da perspectiva intencional. Quando um cientista cognitivo confia num sujeito, quando este declara que ele está rotacionando um objeto em sua mente, por exemplo, o cientista deve confiar no fato de que o sujeito *acredita* que o está girando, e se manter neutro a respeito de se ele de fato está ou não girando mentalmente. O procedimento é análogo ao de um antropólogo, que não deve aceitar a existência do Deus Feenoman, mas deve coletar minuciosamente os dados declarativos que uma sociedade tribal lhe dá sobre as características desse Deus. Acreditar no Deus seria problemático como maneira de entender as crenças de uma população, mas estudar a natureza dele não. Isso porque ele pode ser considerado um objeto nocional apenas, e nem por isso deixa de ter características, ou uma natureza. É possível que haja conflitos, conforme as perguntas fiquem mais específicas: alguns dirão que Feenoman tem olhos azuis, outros castanho, mas suas características principais estarão determinadas. Ele não precisa existir para ser um objeto do mundo nocional dessas pessoas, ele não precisa, se fizermos uma concessão razoável, sequer ser consistente! Digamos agora que esse

povoado se considera a última autoridade sobre Feenoman, o que não é um grande problema, dado que eles são os feenomanistas afinal! Ainda assim, sua crença de infalibilidade é problemática, por exemplo, o próprio Feenomenan poderia aparecer e explicar algumas coisas, tirar dúvidas sofisticadas etc. (¿qual dos dedos do pé de Feenoman é maior?). Em seu *Consciousness explained* (1991a, p. 83), Dennett afirma:

Um feenomanista nativo que simpatizou com os antropólogos visitantes e adotou sua perspectiva teria de adotar uma atitude de distanciamento ou neutralidade sobre suas próprias convicções (ou não devíamos dizer suas convicções *anteriores?*), e iria no processo sair dos quadros dos verdadeiramente devotos.

Podemos dizer, defende Dennett, o mesmo dos fenomenólogos, ou dos realistas fenomenais, aqueles que defendem que a consciência fenomênica existe. Devemos sem dúvida acreditar que eles acreditam possuir esses estados, mas isso não nos compromete a acreditar naquilo que eles dizem estar experienciando. Esse agnosticismo diante das crenças de primeira pessoa de um sujeito é a característica fundante da heterofenomenologia.

A sugestão de Dennett é que deixemos o ‘problema difícil’ para depois. Esse é o problema que consiste em perguntar, ¿afinal, porque esses movimentos talâmico-corticais produzem *esta* sensação, ao invés *daquela*?

Esse problema, a julgar por uma análise da filosofia da mente da segunda metade do século XX e início do XXI, não tem progredido muito. Por outro lado, o conhecimento sobre cognição e motivação cresce exponencialmente a cada dia, a psicologia evolutiva, teorias como a dupla herança, a memética, a computação e diversas outras áreas do conhecimento iluminam cada vez mais como é possível que conheçamos o mundo e de que maneira o fazemos. Teorias cognitivas estão progredindo à toque de caixa, enquanto que teorias filosóficas progridem em baixa velocidade por meios viscosos e confusos.

Falamos anteriormente sobre soluções cognitivas para vários tipos de problemas abordados na filosofia da mente, livre-arbítrio, referência e significação, percepção do tempo, identidade pessoal e alguns outros. Se a consciência real consiste nesses processos, em particular quando eles estão sob o foco da atenção e do controle, então já estivemos falando sobre consciência várias vezes ao longo desse texto. Até aqui, tenho tentado mostrar como as formas que Dennett dá para os conceitos servem para resolver alguns problemas científicos e filosóficos que surgiram com relação a esses conceitos.

Continuarei essa tarefa inserindo mais algumas explicações cognitivas interessantes de eventos conscientes peculiares. Todas essas explicações estão em algum ponto da obra de Dennett, que as retirou em grande parte das ciências, adicionando um ou outro toque pessoal. Os experimentos mentais em particular não vêm das ciências, e neste ponto já sabemos suas vantagens e desvantagens, e estamos vacinados para analisá-los.

Dennett sugere que existe um tipo de ataque contra sua posição que incorre num erro análogo ao seguinte. Suponhamos que você vá a corrida de cavalos, e o cavalo Charlie está correndo ao redor da pista na frente de Joel e Bismuto; eventualmente, Joel ultrapassa por um tempo, finalmente Bismuto por alguns breves instantes, Joel o ultrapassa novamente e todos começam a desacelerar até que não estão mais correndo em nenhuma direção em particular. Aí alguém lhe pergunta: ¿Quem ganhou a corrida? Não existe uma resposta razoável a essa questão, pois não havia uma linha de chegada, o máximo de informação que poderíamos dar a um desavisado é que a liderança foi de um cavalo, depois de outro, variou de tal e tal maneira, até que por fim os cavalos pararam de correr, não existe outro fato *acima e além* destes já citados. Dennett sugere então que o mesmo se passa com alguns pensadores ao pensarem a consciência: eles acham que há de existir uma verdade totalmente bem definida explicitando do que você se tornou consciente primeiro, no meio e depois (lembrem-se do tratado de Ghent). Mas o que de fato ocorre é que por algum tempo a informação A está dominando suas hierarquias neurais e, portanto, seu comportamento, depois noutro instante um outro estímulo B está mais influente, e outro C, novamente B, até que eventualmente todos eles deixam de estar em posições de destaque, simplesmente porque a atenção desviou para outro assunto. A, B e C, estímulos que dominaram a atenção por algum tempo, não estarão para sempre perdidos. Pois deixaram algum tipo de influência no funcionamento futuro da mente, por exemplo através de memórias, padrões de reativação sináptica, etc. O ponto, fortificado de novo e de novo ao longo da obra de Dennett, é que não existe um teatro cartesiano, um lugar único que define onde algo se torna consciente, um ponto central de unificação da informação que perpassa nosso organismo e nossa vida mental. Alguém que defenda que há uma verdade última que define se eu estive ou não consciente de algum estímulo tem que embasar essa verdade em uma propriedade que *faça alguma diferença*.

Para alguém que está fora do mundo da filosofia da mente, a aceitação da condição de *fazer alguma diferença* pode parecer trivial, mas aqueles que estão dentro desse mundo sabem que isso não é aceito por todos. Muitos filósofos (notavelmente

Ned Block e David Chalmers) defendem a possibilidade de que haja uma propriedade que não faz nenhuma diferença em nada que possamos saber, e que é possível que a consciência seja tal propriedade. Não é apenas uma defesa de que pode existir algo que não podemos saber agora, mas algo que não poderíamos nunca saber, algo que não modificaria em nada o funcionamento do mundo físico de nenhuma maneira. Não modifica nenhuma propriedade disposicional de agentes (aquelas que os fariam agir duma ou outra forma). Não modifica nenhum ato de fala, não modifica nenhuma afirmação como “Estou consciente!” ou “Claro que estou consciente da propriedade intrínseca que é ‘ser azul’”, e assim por diante. Em suma, uma cópia do mundo físico inteiro poderia simplesmente não ter essa propriedade, e o mundo físico ainda seria totalmente igual e não diferenciável. Pois bem, digamos que assim seja, a questão fundamental agora se torna: ¿Quem se importa?

De início, sabemos que ao menos Ned Block e David Chalmers, mas também, de maneira leiga, todas as pessoas que vão a bares para conversar sobre se o verde que veem é de fato o vermelho visto pelos outros (¿será que essas pessoas existem de fato ou são uma fantasia filosófica construída por filósofos que não têm uma vida social agitada? eu sempre me pergunto). É interessante perguntar porque eles chegariam a conclusão tão exótica de que existem tais tipos de coisas sem recorrer à óbvia resposta de que filósofos são pessoas que inventam coisas que não podem ser testadas para ter o prazer de discutir por horas a respeito disso (há uma verdade grande em alguns aspectos disso).

O argumento da concebibilidade contra o fisicalismo

O argumento sugerido é o seguinte. Tomemos as propriedades fenomênicas como dadas; é possível conceber um mundo igual ao nosso fisicamente (do nível molecular ao nível cosmológico) que no entanto não possui esse tipo de propriedade. Como isso é concebível, é metafisicamente possível (pode acontecer em outro universo, mesmo que não seja o caso no nosso). Isso mostraria que das propriedades físicas não se seguem as propriedades fenomênicas. Mas no nosso mundo há propriedades fenomênicas. Porém, acabamos de ver que elas não são deriváveis das propriedades materiais. Assim, há entidades e propriedades que não são materiais, e portanto o materialismo é falso.

O materialismo é a crença de que todas as verdades são verdades físicas, e de

fato, se admitirmos propriedades fenomênicas tal como as intuimos, podemos conceder que elas não parecem ser propriedades relacionais, estruturais, relativas à forma das coisas. Elas não poderiam por exemplo ser capturadas por equações que descrevam o comportamento, pois uma equação só descreve o comportamento de uma entidade *em função do tempo, ao longo do espaço*. Uma equação não nos dá informação a respeito de aspectos intrínsecos de nenhum objeto, apenas de suas propriedades relacionais. A física não diz o que é um elétron, por exemplo, apenas descreve algumas propriedades relacionais do elétron disfarçadas de propriedades instrínsecas. A carga negativa de um elétron, por exemplo, na verdade é uma propriedade que o elétron possui de repelir outros objetos de carga igual e atrair objetos de carga inversa. Tudo o que a física nos mostra (e será necessário confiar em mim nesse ponto, pois não poderia enumerar exemplos sem sair demasiado de nosso escopo de análise) diz respeito a propriedades relacionais, descrições estruturais, a ordem das coisas, como elas se ordenam e se orientam umas em relação às outras. O mesmo vale para as leis da química, da biologia, da psicologia experimental etc.

Se existem propriedades de outro tipo além das propriedades estruturais, temos de conceder que ao menos é concebível um mundo em que houvesse apenas as propriedades físicas. Esse é o ponto defendido por Chalmers, Block e outros realistas fenomênicos. Se podemos imaginar outro mundo no qual as propriedades fenomênicas não ocorrem, mas as físicas ocorrem, isso significa que as verdades sobre as propriedades fenomênicas não são uma consequência das verdades a respeito da física.

Há muitos filósofos criticam o argumento porque consideram absurdo que conceber algo seja suficiente para determinar possibilidade. Há também diversos críticos que defendem não ser realmente possível conceber as verdades da física ocorrendo num mundo sem que as verdades fenomênicas ocorram. Dennett está no segundo grupo certamente; eu não poderia afirmar que ele está também no primeiro, embora pareça bastante condizente com sua forma de pensar. Mas este não é o jogo de Dennett.

Dennett dá o drible

Essa problematização acima foi colocada em sua forma mais conceitual e pura não para mostrarmos como Dennett interage com ela, ou quais seriam seus argumentos contra a posição dos realistas fenomênicos, mas sim para mostrar um pouco sobre o

estilo de Dennett. Dennett não interage com ela.

Em um interessante e bastante engraçado livro sobre felicidade, *Stumbling on happiness* (2007), o professor de Harvard Daniel Gilbert apresenta evidências experimentais de que ao fazermos previsões a respeito do futuro, isto é, simulações mentais de situações futuras, o principal erro é esquecermos de considerar o que *não* estará presente. ¿O que você acharia de comer um espaguete amanhã à noite no jantar? Provavelmente sua mente se antecipou, pensou em um espaguete particularmente suculento. Mas seu espaguete não estava contido na minha pergunta simples, é possível que o seu tenha azeitonas, molho a bolonhesa, e seja bastante rico em queijo ralado. Também é possível que você tenha imaginado um espaguete particular que comeu recentemente e pensou em como seria comê-lo amanhã. O fato é que sua mente preenche minha sugestão com uma miríade de informações que não estavam lá, advindas de seus gostos, memórias, desejos etc., nada que não tenha sido advertido pelos flancos por Kant. Mas note que aí ocorre um grande problema, ao prevermos o futuro preenchemos ele de detalhes que não estão presentes, e portanto avaliamos de maneira errada o quanto desejamos aquela situação. Isso é um fato a respeito de felicidade, e tenho uma convicção análoga com respeito a simulações de pessoas. Quando tentamos imaginar a forma de pensar de alguém, eu defendo, preenchemos muitos detalhes com o que deve estar lá⁴⁷, mas incorremos no erro cognitivo de não salientar o que *não* deve estar lá. Por isso expus o argumento filosófico contra o materialismo, porque esse não é o tipo de jogo que Dennett joga. A informação a respeito do que não fazer ao simular Dennett é mais importante do que a do que fazer ao simulá-lo, se as previsões de Gilbert forem aliadas a minha hipótese. Então, devo clarificar aqui porque esse não é o tipo de jogo que Dennett se dispõe a jogar.

Há dois fatores relevantes em jogo, o primeiro é que a vida é finita e é necessário optar pelo que vale ou não a pena discutir e que tipo de controvérsia é ou não frutífera. O segundo fator é que Dennett acredita que esse tipo de discussão é pouco frutífero, na medida em que os argumentos dados são indecidíveis. A pergunta que ele nos sugere é “¿porque não postergar indeterminadamente este tipo de questão enquanto os outros

⁴⁷ Os psicólogos evolutivos talvez encontrem um embasamento para minha convicção na teoria da mente, já que o “*default*” (padrão) em nossa teoria da mente é que o outro é como eu. Os filósofos encontrarão embasamento no princípio da caridade, requerido para superar a indeterminação radical proposta por Quine (ver por exemplo Lewis, 1974). Os lógicos talvez percebam que a intuição é guiada por uma problematização análoga em forma ao teorema de Skölem-Lowenheim, já que temos de gerar um procedimento de decisão das condições do que vamos imaginar num tempo finito, a partir de uma informação com infinitos modelos.

campos de pesquisa a respeito da consciência estiverem dando resultados?” Afinal, apenas após longa investigação das novas descobertas a respeito de como a mente funciona poderemos estar convictos de que algo foi de fato deixado para trás, que existe um caráter fenomênico *acima e além* das propriedades disposicionais, das crenças, das convicções e das ações das pessoas. Isso pode ser uma explicação de porque Dennett rejeitaria essa discussão em particular sobre o materialismo metafísico, colocada nesses termos. Mas, além disso, eu gostaria de libertar o simulador de um erro mais amplo no qual ele poderia incorrer ao simular Dennett. Esse erro é o erro de entrar nesse *tipo* de discussão e assunto. Estou defendendo, em outras palavras, que existe uma *classe específica* e mais ou menos bem definida de *formas* de se analisar um problema nas quais Dennett não incorre, e este caso particular é um exemplo dessa forma. Caracterizar essa forma é mais ou menos difícil, e em parte a caracterização depende da visão de filosofia de Dennett apresentada no início do texto, como uma extensão da ciência, sem fronteiras definidas com a mesma. Tentarei dar algumas pistas gerais, apresentando tipos de argumentos que Dennett rejeita. Note que a rejeição é a argumentos que tenham todas essas propriedades, ou boa parte delas, não apenas uma.

- Argumentos bastante formalizados (traduzidos em lógica formal), complexos e dependentes de algum sistema particular de lógica para funcionarem.
- Argumentos entre posições sem nenhuma consequência observacional ou a respeito de entidades não observáveis.
- Argumentos que presumem uma idealização do conceito de “observador” ou “verdade”, a tal ponto que esses conceitos pareçam estar tomando o ponto de vista observacional de Deus.

Com relação a esses tipos de argumento, em geral, Dennett dá o drible. Uma outra forma de conhecer esse tipo de argumento é consultando Chalmers (2003) e Chalmers & Jackson (2001). Evitar engendrar-se nesse tipo de argumento é difícil por diversas razões. Em termos intelectuais, esse tipo de argumento permite movimentos espertos, e portanto uma pessoa inteligente pode se sentir atraída a brincar com eles, a despeito de sua falta de relevância. São casos em que é particularmente fácil destacar-se em termos de inteligência, sem necessariamente passar pelo trabalho de aprender o que é necessário para conseguir criar algo condizente com a realidade. Acredito que Dennett, principalmente ao longo da segunda metade de sua carreira, tenha passado a driblar esses argumentos porque eles dizem respeito a algo que ele não considera importante.

Um de seus últimos livros, a respeito de religião e sua evolução histórica (Dennett, 2006), demonstra bem como ele se importa em pensar a respeito de coisas que fazem modificações significativas no mundo, assim como seu estilo claro de escrita, que visa inclusive o público não filosófico. Enquanto alguns filósofos como David Lewis ou David Hume faziam filosofia para descobrir no que devem acreditar ou para se divertir, me parece razoável dizer que atualmente Dennett faz filosofia para fazer diferença. Um simulador de Dennett deve levar em conta essa informação sobre o que *não fazer*, ao adentrar um novo problema.

De argumentos podres a intuition pumps sobre a podridão

Seria contraditório da minha parte continuar discutindo o argumento que sugeri ignorar, e por isso seguimos com algumas maneiras propostas por Dennett para pensar a consciência. Tomo uma citação do artigo “Animal consciousness”:

Considere a hipótese de que, até onde posso saber, para um urubu-de-cabeça-vermelha a carcaça de galinha podre cheira igual ao que um peru assado cheira para mim. Pode a ciência jogar qualquer luz a favor ou contra essa hipótese? Sim, ela pode refutá-la quase sem esforço: já que a maneira como peru assado cheira para mim é composta (e exaurida) pelo grande conjunto de disposições reativas, efeitos de memória, e assim por diante, que são em princípio detectáveis em meu cérebro e comportamento, e como muitas destas estão em última instância além do maquinário do cérebro de qualquer urubu, é simplesmente impossível que algo possa cheirar para um urubu da maneira que peru assado cheira para mim. (Dennett, 1995b, p. 706)

Ele prossegue o exemplo dizendo que se você está interessado em saber então *como* afinal cheira uma galinha podre para o urubu, a resposta dependerá de quão inquisitivo você está sendo e qual o grau de sofisticação de explicação desejado. Podemos aprender o conjunto de disposições que o urubu aprenderá, podemos aprender o mesmo tipo de coisa que eu posso aprender sobre você, mas existem diversas coisas que *não* iremos aprender. Por exemplo, não iremos encontrar nenhum maravilhamento ou padrão complicado de associação, nenhuma memória do sítio da avó, nenhuma reminiscência. A ideia de que haja uma divisão clara entre os possuidores da propriedade “o que é ser como tal” e meros “autômatos” é para Dennett um artefato antiquado de nossas presunções. A consciência não é um fenômeno tudo ou nada, ela vem em graus, e perguntar-se a respeito de se um animal a tem é um equívoco, é o equivalente a perguntar se aves são sábias.

Quineando o inacessível à heterofenomenologia

No artigo “Quining qualia”, Dennett (1988) se propõe o objetivo subversivo de “quinear” os qualia. *Qualia* (plural de *quale*) são as qualidades particulares que constituem o *como é ser algo para mim*, como um tom particular de azul do céu em um ponto, num momento particular, ou o sabor de um leite num certo momento (um qualia gustativo). O verbo “quinear” inspira-se no filósofo Willard Quine, e vem do divertido dicionário satírico de epônimos criado por Dennett, e significa: negar resolutamente a existência ou importância de algo real ou significativo. Dennett concede que à primeira vista nada seria mais absurdo do que tentar convencer as pessoas de que não existe uma propriedade tal como os qualia, mas ele não está brincando.

Falarei apenas de duas das muitas objeções ao conceito de qualia como qualidade singular da qual a experiência é constituída.

Digamos que um neurocirurgião maléfico queira alterar seus caminhos neurais de tal maneira a realizar o clássico experimento filosófico da inversão do espectro, isto é, fazendo com que você acorde e perceba que de repente a grama está vermelha, o céu amarelo e assim por diante. Supor-se-ia que você pode dizer que sofreu uma inversão das propriedades visuais, ou seja, uma inversão de qualia. Nesse caso, parece que os qualia são propriedades aceitáveis afinal, pois proposições a seu respeito poderiam ser justificadas, empiricamente verificadas, e até explicadas! Afinal, nesse caso poderíamos (virtualmente) fazer uma completa descrição das mudanças que ocorreram no seu sistema neural. É tentador portanto supor que evidência neurofisiológica robusta poderia resolver o problema de determinar se os qualia de alguém se invertem ou não, a contento do freguês.

Mas isso é um erro. Vejamos um caso de neurocirurgia alternativa. Existem ao menos duas maneiras através das quais o neurocirurgião poderia obter o efeito desejado:

1. Inverter um dos canais ‘iniciais’ de produção de qualia, i.e. no nervo óptico, fazendo com que os eventos de entrada sejam o oposto do seu valor original. Por hipótese, isso inverte os qualia.

2. Deixar todos esses caminhos de entrada intactos e simplesmente inverter alguns links de acesso à memória – o que quer seja que conecte os matizes de hoje com os de outrora. Por hipótese isso não inverte os qualia, apenas disposições ancoradas na memória de como reagir a eles.

Mais uma vez você acorda e exclama “Flying spaghetti monster! *Algo* se passou comigo, ou meus caminhos pré qualia foram invertidos e meus qualia também, ou minhas disposições relacionadas a eles foram modificadas. Eu me pergunto qual!”

De início poderia parecer que uma ferramenta como modificação neural fosse ser um auxílio aos defensores dos qualia, afinal estamos penetrando em propriedades internas da cabeça, onde nenhum homem jamais esteve, mas o exemplo de Dennett nos mostra o contrário, que de fato é possível atingir os mesmos efeitos da suposta inversão de qualia sem invertê-los absolutamente. Se existem qualia, então eles são ainda mais inacessíveis do que pensávamos: não apenas eles não podem ser intersubjetivamente descobertos e comparados, mas não podemos ter certeza sequer do nosso próprio caso, ao menos não através de introspecção.

Consideremos agora o caso de Chase e Sanborn, contratados como provadores de café, e cujo trabalho é manter o sabor do café constante ao longo dos anos. Um belo dia, após seis anos de trabalho, Chase confessa: “Eu não gosto mais desse trabalho, eu achava o café da Casa o melhor do mundo. Eu tinha orgulho da responsabilidade de preservar esse sabor. E nós fizemos bem o trabalho, o café tem o mesmo sabor hoje do que em meu primeiro dia, mas, sabe o que é, eu não gosto mais dele! Meus gostos mudaram. Eu me tornei um bebedor mais sofisticado. Eu não gosto mais *desse gosto* absolutamente.”

Sanborn responde: “Interessante você mencionar isso, algo semelhante aconteceu comigo. Eu achava o mesmo, o melhor sabor do mundo, e agora eu já não me importo mais com o café que estamos fazendo. Mas meus *gostos* não mudaram; meus... *gustadores* mudaram. Isso é, acho que algo aconteceu de estranho com minhas papilas gustativas ou alguma outra parte do meu sistema perceptivo. O café da Casa não tem mais o gosto que tinha quando entrei aqui, se ao menos ele tivesse, eu o adoraria, pois eu ainda acho que *aquele sabor* é o melhor sabor de café. Eu concordo que fizemos nosso trabalho bem. Todos nós concordamos que o gosto é o mesmo, e os outros provadores também, mas acho que eu não sirvo mais para esse trabalho.”

Chase e Sanborn são iguais no que diz respeito a antes gostar do café, e agora não mais se importar com ele. Mas eles alegam ser diferentes num outro aspecto, o café tem para um o gosto que sempre teve, e para o outro não. ¿Mas será que podemos tomar seus protestos ao pé da letra? ¿Se o relato de ambos depende da acuidade de suas memórias, há como checar essa acuidade?

Podemos esperar nesse ponto um argumento semelhante ao anterior, no qual

Dennett defende que cada um deles poderia estar equivocado a respeito de si próprio, e não haveria como sabermos. Além disso, podemos defender que num caso como esse simplesmente não há uma questão de fato, existe uma “invencibilidade epistêmica” de qualquer uma das teses, e ao diminuir uma asserção até que não haja mais nada para estar correto ou errado a respeito dela, estamos alocando-a no espaço daquilo que não nos deixa questão alguma de fato.

Revisões orwellianas e stalinescas

O intuition pump do neurocirurgião, mencionado acima, nos facilitará o caminho para entendermos a diferença entre revisões orwellianas e stalinescas, uma diferenciação importante que Dennett traça no livro *Consciousness explained* (1991), no qual demonstra que duas explicações alternativas para o “fenômeno fi” (ilusão de movimento aparente) são... menos alternativas do que possam parecer.

Em *1984*, de George Orwell, o ministério da verdade fabrica a história, reescrevendo-a, e nega qualquer conhecimento do passado. Essa distinção revela uma das distinções importantes para nós, a diferença entre aparência e realidade. A revisão orwelliana é uma maneira de enganar a posteridade. Outra é encenar julgamentos, apresentações organizadas de falso testemunho, e confissões mentirosas, com evidências simuladas. Chamemos esse modo de revisão stalinesco.

Quando somos enganados, é por sorte que sabemos de qual dessas duas maneiras fomos enganados. Qualquer caso de sucesso de reforma de nossas memórias não nos permitiria identificar o que ocorreu (todos os arquivos seriam efetivamente queimados, todos se esqueceriam dos fatos etc. não poderíamos diferenciar os casos), mas na maioria dos casos de fato ocorrentes podemos fazer essa diferenciação, no nível histórico e social ao menos. Poderíamos concluir que esse modelo se aplica até o limite interno, mas isso é uma ilusão.

Para ilustrar isso tomarei a discussão do fenômeno fi. No caso mais simples, dois pontos estão separados por 4 graus de ângulo visual. Eles acendem e apagam sucessivamente, mas aparentemente apenas um único ponto parece se mover para trás e para frente. Fi foi estudado numa versão a cores, com cada um sendo de uma cor. A questão é de que cor o ponto *parece* ser quando se move. ¿Haveria uma mudança gradual de cor quando *ele* “passasse” ao longo do caminho entre um ponto e outro? A resposta foi mais impressionante do que a imaginação que lhe antecedeu. Dois pontos

de cores diferentes eram acendidos por 150 mseg cada (com intervalo de 50 mseg), o primeiro parecia estar se movendo e subitamente no meio de sua passagem mudar de cor! Note que a *passagem* aqui referida simplesmente não existia na tela. A passagem ilusória foi o momento no qual os sujeitos percebiam a mudança de cor ocorrendo! O filósofo Nelson Goodman ponderou, diante desse experimento ¿como seria possível que nós preenchamos o espaço entre os dois, no tempo entre um e outro com a cor certa, *antes que a segunda cor apareça?* Abandonando hipóteses sobre pré-cognição e abobrinhas relacionadas, seria necessário que a modificação ocorra após a identificação do segundo estímulo. ¿Mas se o segundo estímulo já está “na experiência consciente” não seria tarde demais para modificar a percepção do momento intermediário entre um e outro?

A tarefa editorial do cérebro deve ser feita de maneira multi-tarefa. Qualquer informação que requer uma fonte deve esperar a informação daquela fonte. A discriminação da segunda cor (verde) tem de ocorrer antes da experiência de transição que aparenta ocorrer no tempo entre os dois flashes. A experiência é de vermelho, vermelho virando verde e verde. Então segue que sua consciência do sistema inteiro há de ser postergada até que o flash verde seja (¿inconscientemente?) percebido. Se essa intuição ainda lhe aparenta boa, seu modelo do funcionamento mental ainda está baseado no teatro cartesiano, o local hipotético unificador da consciência, onde o Self reina soberano.

Façamos uma heterofenomenologia das declarações de um sujeito que passou por esse experimento. Ele declara que “O ponto mudou de cor *aqui*”. No seu mundo heterofenomenológico, no seu mundo nocional, houve uma mudança no meio da trajetória. Consideremos a hipótese de um modelo stalinesco: Na sala de edição do cérebro, localizada antes da consciência, existe um atraso, assim como o atraso usado em programas ao vivo, que permite aos sensores algum controle sobre o conteúdo. Na sala de edição, chega o primeiro quadro A, do ponto vermelho, e posteriormente chega o quadro B do ponto verde. Haveria tempo para criar uma edição com dois quadros intermediários, formando a sequência A,C,D,B no caminho para a projeção, no teatro da consciência. O produto chega já em versão final à glândula pineal, ou o que quer que desempenhe o papel do Self.

Alternativamente poderíamos ter uma revisão orwelliana. Assim que os dois quadros passam pela consciência (sem nenhuma ilusão), um mecanismo revisor de histórias na estação de memórias do cérebro nota que a história atual desde pedaço não

faz nenhum sentido, então ele interpreta os eventos brutos, criando uma narrativa sobre a passagem entre os pontos, e instala a estória devidamente modificada nos arquivos da memória para referência futura. Como ele é rápido, trabalha antes mesmo de haver tempo para que você diga qual foi a sequência e qualquer fala já estará contaminada por essa estória construída para dar sentido à sequência. Você *dirá e acreditará* que elas tornaram-se conscientes não na sequência A, B, mas na sequência A,C,D,B, exatamente como ocorre com os sujeitos experimentais, mas isso seria na verdade uma alucinação.

Pode aparentar que de imediato possamos rejeitar o modelo stalinense, devido ao atraso entre os tempos iniciais de cada ponto colorido, 200 mseg. Já que por hipótese a experiência não pode ser composta pela sala de edição até que a informação sobre o verde chegue, a consciência do primeiro ponto vermelho deve esperar ao menos isso. Se a consciência recebesse de imediato a informação do vermelho, antes de receber o quadro B e fabricar o C e o D, o sujeito experienciaria uma lacuna no filme, tão longo quanto a pronúncia de uma sílaba, ou seja, a falta de cinco quadros de um filme, pois esse seria o tempo mínimo necessário para que a informação que gera a revisão chegue. Uma hipótese é que a consciência *sempre espere* ao menos 200mseg para passar o filme, mas isso é falso, pois se pedirmos aos sujeitos que apertem um botão assim que virem um ponto vermelho, eles apertam o botão (descontados os tempos de conexão) de modo que não há tempo para que haja essa lacuna.

Assim estaríamos convencidos de uma revisão posterior à consciência, uma revisão orwelliana. O sujeito aperta conscientemente o botão depois de ver o ponto vermelho, depois uma falsa memória com a sequência é inserida no lugar de tal modo que suas declarações digam que viu a mudança de cor naquele ponto. Sua memória real foi apagada e deu lugar a outra. Tal qual o caso dos neurocirurgiões, na realidade não teremos como diferenciar essa duas hipóteses, de revisão antes ou depois. Novamente a aparente mudança entre as cores não pode ser resolvida por nenhuma das teorias. Não apenas isso, mas o próprio sujeito não estaria capacitado a dizer qual dos dois ocorreu, não há uma evidência diferenciada, ele está tão pouco em posição de decidir a respeito das hipóteses quanto qualquer experimentador externo. O ponto de Dennett aqui é deixar claro que não há uma maneira de ordenar os eventos conforme eles ocorreram na consciência através dos relatos verbais. Há uma subdeterminação.

Esses casos são variações sobre um mesmo tema, e assim sendo podem ter uma moral comum. Uma boa moral é a de que, se quisermos estabelecer um único momento como o momento da consciência, ele terá de ser totalmente arbitrário: é sempre possível

traçar uma linha, mas ela nunca será uma linha com qualquer distinção funcional.

Além da linha vermelha invisível

Um crítico de Dennett bastante perspicaz é Joseph Levine, que traça a seguinte crítica (comunicação pessoal 2013). Uma teoria da consciência não pode lidar apenas com nossas declarações verbais sobre elas, as experiências mesmas são o explanandum básico ao qual devemos prestar contas. Levine sugere a seguinte sequência intermediária entre essas etapas:

- (a) “as próprias experiências”;
- (b) crenças sobre as experiências;
- (c) “julgamentos verbais” expressando essas crenças;
- (d) asserções de um tipo ou outro.

Para Levine, a informação primária é (a), mas para um heterofenomenólogo, (d) é a informação bruta e não interpretada. Antes de chegarmos à teoria temos de interpretar essa informação bruta, que nos leva através de (c), atos verbais, até (b), as crenças sobre experiências. Essas seriam nossas informações pré-teóricas, a informação interpretada para uma ciência da consciência. Mas Levine quer levar-nos até (a). Entretanto, se você tiver experiências conscientes (a) das quais não se lembra, e não acredita ter (b), então elas são tão inacessíveis a você quanto a um observador externo. ¿E se você acredita que tem crenças que são inexpressáveis em julgamentos verbais? Bem, aí temos uma nova informação bruta, e temos de acomodá-la a nossas teorias:

S julga ter crenças inefáveis sobre X.

O ônus da prova fica, é claro, com o sujeito, de demonstrar que de fato as crenças existem, e não apenas a crença na existência das crenças.

¿O que Dennett sugere que coloquemos no lugar das várias teorias refutadas, explícita ou tacitamente ao longo deste texto? Em primeiro lugar, uma teoria da consciência como fama no cérebro, sobre a qual já falei brevemente: a consciência seria não mais que o produto organizado de hierarquias de fama de diferentes sistemas mentais competindo por espaço, espaço de trabalho e espaço de memórias também. Um nome mais recente para a teoria da consciência de Dennett é *fantasy echo*, pois ele sugere em *Sweet dreams* (2005) que o fator crucial a respeito da consciência é sua transmissibilidade e sua capacidade de reverberar nas ações do indivíduo, seja na hora,

através de mecanismos por exemplo de controle e atenção mediados pela evolução, seja posteriormente, através de reverberações mediadas pelos processos de sedimentação de memória. Uma câmara de eco para as histórias e fantasias compõe o sujeito, conforme ele interage com o mundo ao seu redor – talvez uma visão mais poética do que a do reducionista médio. Há de se fazer saliente que a capacidade de reverberação não é uma propriedade secundária da consciência, ou um fator correlacionado, é o próprio fator determinante.

Outra teoria é a de que nós humanos temos uma capacidade de manter nossas câmaras reverberando muito melhor do que animais, como cães por exemplo. Isso depende de uma teoria de re-instanciação de memórias de novo e de novo na corrente da consciência. Por conta de ficarmos fazendo ensaios mentais do que ocorreu em certas ocasiões, constituímos uma relação de longo prazo de influências com outras partes de nós mesmos, outras fatias no nosso eu quadridimensional. As relações de fama que operam regendo a consciência em escala imediata na competição entre hierarquias informacionais têm análogos distribuídos pelo tempo, com memes antigos e novos disputando nossa atenção, preocupação etc. Segundo essa teoria, nossa mente funcionaria como uma máquina virtual que reinstancia sucessivamente diversos memes. Uma máquina memética, cujas disputas internas no tempo e no espaço regem a ausência de teatro e a multiplicidade de histórias que nos constituem. Isso não ocorre com um cachorro por exemplo, que, ainda que possa ter algum tipo de ativação simples de um sistema em função de outro, não terá reminiscências do sabor do osso de 1999 da Tia Diná. Nós podemos, no entanto, com um simples aroma reinstanciar uma completa situação ocorrida por vezes dezenas de anos antes. Nossas câmaras de eco reverberam nossas fantasias e reconstruções de maneiras muito mais profundas, sugere Dennett. Além disso, ele propõe que isso possa ser um fenômeno de origem cultural, que deu origem à consciência como a conhecemos, e essa arte de reensaíar estaria por trás do processo como um todo. A existência desse tipo de processo nos permite por exemplo aprender com apenas um exemplo, pois podemos, como criaturas gregorianas que somos, transformar esse um em muitos, ao re-revivê-los diversas vezes.

“Deixe-me resumir,” pede Dennett perto do final de seu *Sweet dreams* (2005, pp. 171-2):

Eu me aventurei com (1) a hipótese empírica de que nossa capacidade de reviver e repassar eventos repletos de conteúdo é o traço mais importante da consciência

– com efeito, é tão perto de um traço definidor da consciência quanto poderemos encontrar; e (2) a hipótese empírica de que essa capacidade de eco se deve em larga medida a hábitos de auto-estimulação que herdamos da cultura humana, de que a máquina joyceana⁴⁸ em nossos cérebros é uma máquina virtual feita de memes.

Essas propostas são independentes, e a refutação de uma não implica na da outra.

Isso nos dá o delineamento geral dos pensamentos de Dennett sobre a consciência, falando tanto sobre as ideias que ele refuta e combate, quanto sobre as que ele propõe. Antes de seguirmos à nossa primeira tentativa organizada de simular o pensamento dennettiano, vejamos duas respostas a críticas que emergiram em relação a sua teoria da consciência.

Crítica um) Existe uma tensão entre a teoria da consciência, que fala sobre um sistema disjunto, separado, com hierarquias mal definidas de fama etc., e o uso da perspectiva intencional, que consiste em *supor a unidade* de um agente para melhor compreender suas ações, crenças e objetivos.

Resposta um) Tudo depende de quão longe estamos observando. Como estratégia preditiva, e para uso cotidiano, a perspectiva intencional é uma maneira frutífera de pensar os agentes, tão frutífera aliás que está engendrada em nossas tendências genéticas de ver o mundo, em nossa cognição⁴⁹, e é através dela que podemos criar sentido de fato nas ações alheias. Para uma ciência da mente, por outro lado, é necessário decompor e entender as subpartes desse sistema, e a teoria da consciência está nesse nível de análise, o de uma psicologia cognitiva subpessoal⁵⁰.

⁴⁸ Em seus romances, James Joyce oferece uma visão da corrente da consciência como uma narrativa sequencial quebrantada que se assemelharia ao fluxo de pensamentos do personagem. Dennett entende que um dos desafios principais para uma teoria da consciência é conseguir ‘extrair’, de algum modo, uma corrente serial, como uma narrativa de Joyce, de um maquinário em paralelo como o cérebro. Daí sua sugestão de uma máquina virtual joyceana emulando uma corrente única num processamento paralelo.

⁴⁹ O livro *The stuff of thought* de Steven Pinker (2007b) discute alguns usos da linguagem na qual nossa visão de nós mesmos fica identificada a um ponto, e possíveis causas naturais para que a linguagem haja evoluído dessa forma. Quando falo de *meu braço*, ou *meu corpo*, por exemplo, estou indicando uma crença sobre mim mesmo como uma entidade pontual e não corpórea, um ponto que tem posse do corpo.

⁵⁰ No artigo “Three kinds of intentional psychology”, Dennett (1981b) sugere que o nível de análise que leva em conta coisas como crenças e objetivos não é o nível de análise de uma perspectiva causal. Crenças e objetivos seriam entidades virtuais; o nível de *wiring* (organização da fiação mental) no qual estariam engendrados os processos psicológicos reais, causalmente eficazes, seria um nível mais baixo. As crenças e objetivos seriam portanto um construto virtual num nível hierárquico superior. Note que Dennett está apresentando uma teoria verificacionista da subjetividade, que se utiliza da perspectiva intencional para definir crenças e objetivos virtuais, e uma outra teoria não verificacionista, mas

Crítica dois) ¿No fundo, no fundo, a posição de Dennett é a de um behaviorista, não?

Resposta dois) Definitivamente não a de um behaviorista skinneriano. A resposta de Dennett consoa com a de Wittgenstein no tópico. Se os estados mentais estão sendo vistos como um tipo de ficção, é uma ficção gramática, não uma ilusão. Costumamos falar de processos e estados, e deixar suas naturezas fora de questão, mas esse é justamente o ponto para o qual não podemos deslizar nesse caso. Nossa suposição tácita de que não há problema em deixar a natureza dos estados mentais fora de discussão até que tenhamos uma compreensão melhor do tema é justamente um comprometimento tácito com uma visão de que já sabemos o que é o estado. Isso porque temos uma concepção do que é entender melhor um processo. O fato é que não podemos cair nessa armadilha, e para isso temos de negar o processo não compreendido no meio ainda inexplorado. Fica aparente que se estão negando os processos mentais, mas não é isso que se quer. Os ‘objetos’ da experiência consciente não podem ser negados, nem pressupostos, mas sim, não mais e não menos, explicados.

cognitiva, das capacidades psicológicas que subjazem à esfera intencional. A perspectiva intencional paga o preço de não corresponder exatamente ao que está ocorrendo causalmente (mas apenas virtualmente) para poder comprar o benefício de ter uma descrição equivalente em diferentes sistemas (humanos, animais, alienígenas, computadores), por maiores que sejam suas diferenças de funcionamento interno. Esta perspectiva, portanto, é a mais evolutivamente condizente em termos de explicações adaptacionistas. A outra teoria, que se interessa mais com o nível causal, tem a tarefa de propor modelos para esse funcionamento que sejam condizentes com o nível virtual de análise, subestruturas que se organizam de tal forma a possibilitar o funcionamento que queremos explicar.

11. O algoritmo Dennett

Sempre quis tentar recapitular o que há de essencial em Dennett em um parágrafo, vejamos se minha falha é ao menos instrutiva:

O pensamento de Dennett se volta para uma unificação da perspectiva biológica genocêntrica e algorítmica com os conceitos intuitivos da psicologia popular. Através de, por um lado, uma defesa da consciência como uma entidade cognitiva e empiricamente acessível, e por outro, uma defesa de vermos o vocabulário mentalista como uma estratégia preditiva cujos termos referem a *abstractas*, oferece um modo de olhar a mente que engloba suas duas facetas mais aparentemente contraditórias: a multiplicidade de pensamentos, desejos, vontades, módulos e sub-inteligências que a compõe, e a unidade (aparente) do Self, capaz de poderes como abstração, resolução de problemas, criatividade, livre-arbítrio e conhecimento sobre o mundo. O problema que ele confronta é o problema de como constituir *uma mente de zilhões* de partes que não são mentes, e para isso é necessário não apenas distribuir a **inteligência** em sub-hierarquias até níveis em que não haja inteligência, auto-organizadas pela evolução, mas cabe também explicar como pode a *unicidade* surgir da *multiplicidade*. Sua sugestão é que a unicidade é aparente, e que aparenta ser assim porque essa é a melhor maneira de prever o futuro, quando se está analisando estruturas esculpidas pela evolução genética e memética, e, *além disso, se é uma dessas estruturas*. Finalmente, nossa capacidade de referência se restringe a mundos nocionais, a objetos nocionais, e nossa capacidade de sentido (e semântica) é herdada dos genes e memes, e não intrínseca.

Pouco a pouco, de aqui até o fim do capítulo, pretendo construir o algoritmo dennettiano, isto é, um conjunto simples de regras mais ou menos simples que formaliza parcialmente a forma de pensar de Dennett, e que pode ser rodado em outras máquinas, que implementariam o mesmo programa ao seguir essas regras.

Daqui em diante é particularmente importante manter em mente que o termo “Dennett” nesse livro vem se referindo à simulação possível de Daniel Dennett que se pode emular a partir do conhecimento de seus escritos, não a exatamente o que o homem diria se confrontado diante do mesmo contexto.

Começemos, arbitrariamente, com um assunto novo. ¿Dado um determinado

assunto novo, qual a primeira ação tomada por Dennett? Sugiro como primeiro corte o corte entre aquilo que exhibe e o que não exhibe design, então a primeira ação (representado por uma elipse, na Fig. 3) seria ver se aquele assunto exhibe ou não design.

Se sim, partimos a um escrutínio mais minucioso do assunto, sob a luz da biologia evolutiva. Isso se dá porque segundo a teoria do gene egoísta de Dawkins, e o darwinismo universal advogado por Dennett em *A perigosa ideia de Darwin*, todos os objetos do universo que exibem design o fazem por serem produto da evolução. Então a próxima ação seria decidir se o objeto desse novo assunto é ele próprio também produto da evolução biológica ou não.

Suponhamos mais uma vez que sim, que estamos tratando de algo que foi desenhado pela mãe natureza através de genes, probabilidade, aleatoriedade e seleção natural. Temos agora que circunscrever melhor de que tipo de coisa estamos tratando, já que ainda poderíamos estar falando das propriedades matemáticas dos espirais de caracóis, ou da tendência à obesidade de um certo grupo de crianças do sul da Eslovênia. A próxima pergunta que se faz relevante será: ¿Isso é um produto de evolução puramente morfológica⁵¹ ou comportamental? Com isso, poderemos discernir por exemplo entre a espessura dos pelos de um porco espinho e a Vontade de Poder que Nietzsche acreditava perpassar a condição humana, por exemplo – o que é o mínimo necessário para pensar um assunto. Nesse passo, temos então três possíveis respostas distintas, (1) apenas morfológica, (2) apenas comportamental, ou (3) ambas.

Estamos supondo que o assunto é novo, então por definição não é o caso que a literatura especializada (filosófica, artística, ou científica) já possua vasta informação a respeito. No entanto, o procedimento que se pode tomar a partir daí em qualquer dos três casos é o de analisar casos *análogos* na literatura. Lembremos a afirmação de Hofstadter de que o centro, o *core*, o fundamento da cognição humana é o uso de analogias. Diferentemente das elipses anteriores, nas quais existia mais de uma decisão possível, esse passo é apenas procedimental, e estará representado em nosso esquema (Fig. 3) como um retângulo, já que dele se segue apenas uma ação.

Após verificar a literatura em busca de casos análogos, provavelmente já teremos encontrado alguns casos mais interessantes que outros (recordem por exemplo a mosca *Sphex*, sempre presente em explicações de cognição avançada), então a próxima pergunta que caberia fazermos seguindo o algoritmo é: ¿encontramos até aqui algo

⁵¹ *Morfologia* é relativo à formação e à forma.

interessante? Esta pergunta presumivelmente admite resposta negativa ou positiva.

Estou descrevendo um processo linear dentro de um algoritmo de processamento paralelo para poder iniciar o contato com o algoritmo de maneira intuitiva, já que a cognição humana pensa histórias lineares muito melhor do que, por exemplo, redes neurais com processamento em paralelo.

Após esse parágrafo paralelo – logo auto-refutante – de autorização, posso supor que sim, havia algo interessante na literatura a partir do qual se possa aprofundar a compreensão do fenômeno em questão. Um traço importante de pensadores como Dennett, que inclusive escreveu sobre a criatividade, é não restringir-se à primeira solução para um problema que possa aparecer. Sugiro que na situação de estar munido tanto de um novo assunto quanto de uma maneira interessante de pensá-lo, Dennett produziria uma analogia, ou um intuition pump, a partir do qual possa trazer a questão sendo considerada a público, da maneira mais clara possível. Em particular, seria uma analogia que não permita desvios grandes de algum ponto que se possa querer defender, ou seja, que não sirva para gerar crenças falsas a respeito de aspectos particulares do objeto em questão.

Em seguida, sugiro que pensaria outra analogia, e fazendo todas as considerações necessárias, preferencialmente com diferenças em relação à primeira analogia, que sejam suficientes para permitir uma comparação entre suas vantagens e desvantagens.

O próximo passo evidentemente seria selecionar uma delas, pesando os prós e contras, (1) em termos de precisão das propriedades estruturalmente equivalentes entre o objeto e seu análogo em cada caso, (2) em termos da probabilidade de apreensão confusa ou equivocada da analogia, e (3) em termos (e isso é algo que aparece bastante principalmente nos escritos mais recentes) de quão engraçado ou atraente possa ser utilizar essa analogia.

Isso é suficiente para começarmos o esquema, retratando esses aspectos de uma progressão particular de algoritmos. Vejamos o que temos até aqui, na Fig. 3.

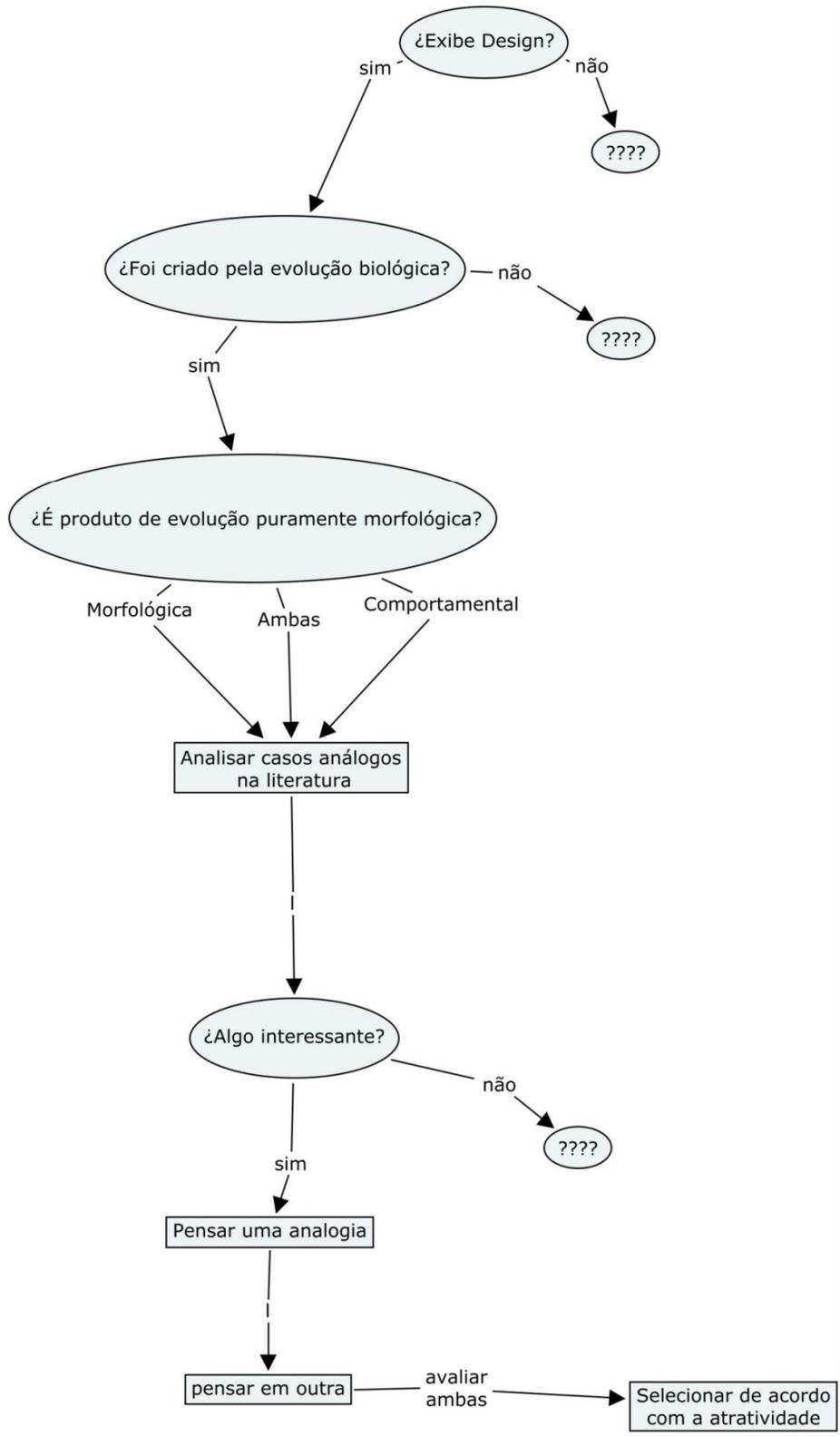


Figura 3: Parte do algoritmo de Dennett, partindo da pergunta se existe design.

Por enquanto estamos lidando com um esquema simples. No entanto, conforme o número de cursos possíveis for aumentando, e o comprimento de cada curso de ação também for ampliado, chegamos a um sistema com poder de processamento maior, e portanto mais complexo. Tomemos agora outro curso possível, partindo do mesmo ponto arbitrário de um assunto novo.

Suponhamos novamente que o objeto do assunto novo exhibe design, mas para logo de início criar uma nova rota, suponhamos que ele não seja produto da evolução biológica. Ora, se exhibe design e não é produto da evolução biológica, até onde sabemos, ele somente pode ser produto da evolução memética, já que não existe nenhum outro processo criador de design estável na natureza⁵². Esse é um procedimento de decisão tal qual o processo de decidir se é ou não biológico, mas agora a opção do “não” nos leva a um vértice sem saída, já que não pode haver nada que chegue nesse ponto. Digamos então que “sim”, estamos lidando com um objeto que evoluiu por razões meméticas. Uma maneira dennettiana de proceder seria procurar uma explicação naturalista e evolutiva que tenha paralelos interessantes. Esse processo pode ser estendido indeterminadamente até encontrar alguma explicação, ou seja, temos aqui nosso primeiro loop dentro do algoritmo. Humanos em geral não enfrentam problemas com loops, isso porque somos dotados de equipagem cognitiva que nos impede de executar a mesma ação recursivamente infinitamente, então não há motivo para preocupar-se com uma contagem de quantas vezes iremos procurar uma explicação paralela interessante. Programar um computador para isso seria bastante mais difícil. Finalmente, depois de algum esforço, encontramos algum paralelo interessante com outro processo de evolução natural, e agora iremos para um passo que já ocorreu no curso anterior, que é pensar numa analogia interessante que possa ir a público. Esse passo já está em nosso esquema, então o que ocorre daí em diante já está determinado (ou estaria, caso o esquema já estivesse completo).

Digamos então que já selecionamos entre nossas duas analogias, e agora cabe perguntar se essa nova ideia é merecedora de um artigo (por exemplo num periódico especializado, ou numa conferência) e caso não o seja se deve ir para um eventual livro. Em qualquer dos casos, numa resposta positiva, o próximo passo é logicamente escrever, e caso não seja o caso que o mérito esteja à altura do status de publicação, podemos

⁵² Emergência (por exemplo, a das estrelas) e eventos extremamente improváveis (a chance de $1/10^{172}$ que um carro se torne cada vez mais novo e limpo ao longo de 10 anos) não produzem o tipo de estabilidade que nos interessa nesse ponto, por isso estão sendo descartados como fonte de design.

voltar para o início de nosso processo, e mais uma vez iniciar o trabalho. A Fig. 4 mostra o que fizemos até aqui.

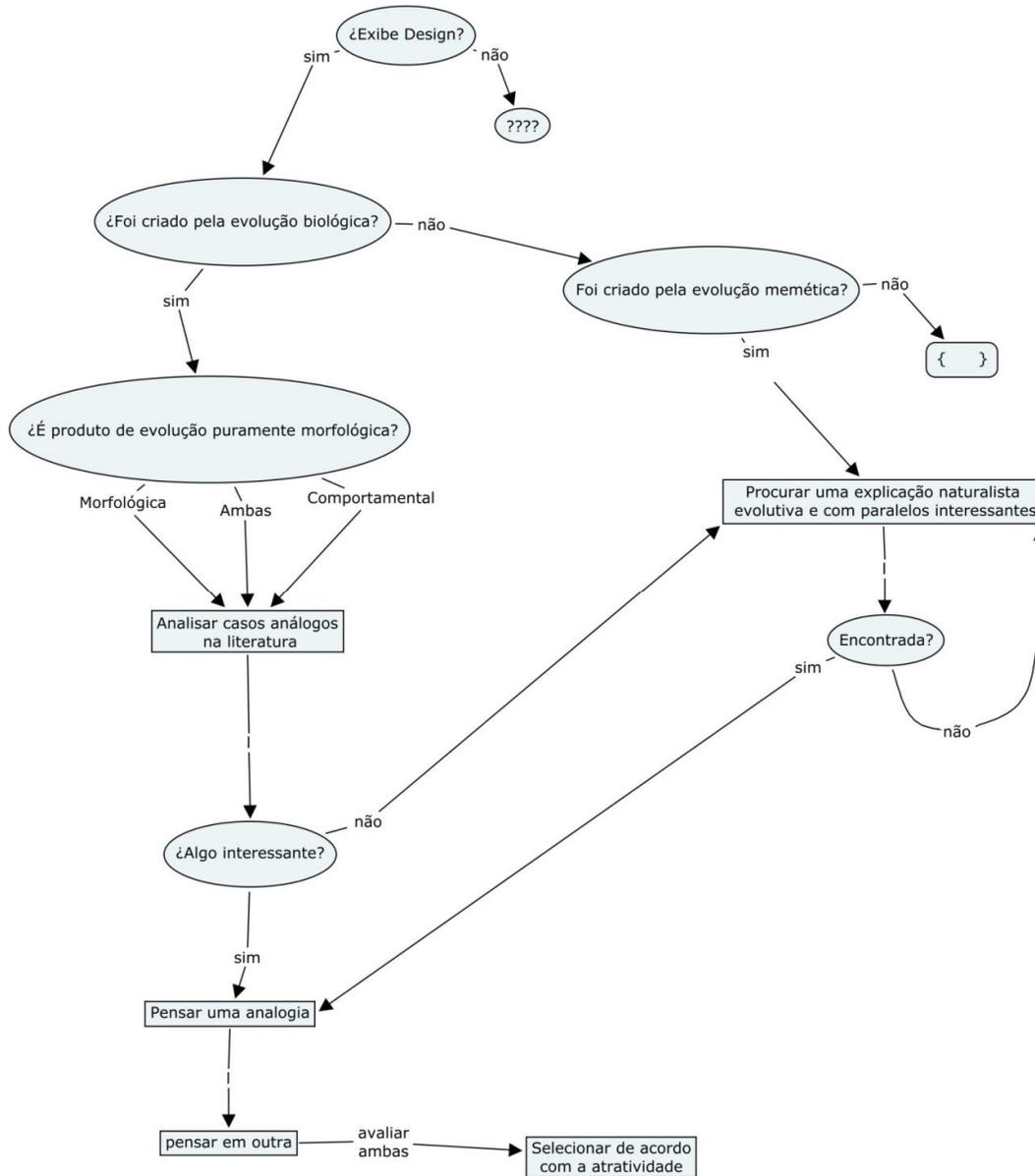


Figura 4: Extensão do algoritmo anterior, incluindo a evolução memética.

Nem tudo nesse mundo exhibe design, ou melhor, nem tudo exhibe design o suficiente para merecer essa designação, sem querer jogar com as palavras. Os biólogos, psicólogos, sociólogos e antropólogos costumam estudar principalmente aquilo que exhibe design. Já os filósofos, assim como os físicos, muitas vezes estudam também aquilo que ou não possui ou não aparenta nenhum design, ainda que possa exhibir outros traços característicos de interesse como por exemplo padrões ou leis regulares.

Digamos então que estamos lidando com algo que não exhibe design. Se algo que não exhibe design atraiu a atenção de Dennett, provavelmente isso se deve a uma dentre duas razões: ou é algo que pode ser usado, por exemplo nos procedimentos anteriores, como analogia para alguma explicação de algo que exhibe design, ou é algo que diz respeito à ontologia. Ontologia, rememoremos, é o que diz respeito ao que existe, ou ao conjunto das coisas existentes. Os filósofos têm muitas maneiras de falar sobre ontologia, e discordam demasiadamente a respeito de quais afinal são os habitantes de nosso mundo. O mobiliário do mundo é ora feito de partículas, ora de campos, ora de pessoas, ora de árvores, ora de universais e particulares. Em suma, a filosofia não progrediu muito nesse sentido. Dennett adota uma visão herdada de seu professor Quine a respeito da ontologia. Ele sacrifica a preocupação a respeito de um conjunto de coisas últimas que existiriam na realidade em favor de discussões mais específicas, conforme vimos ao falar sobre o que Dennett *não* estaria disposto a se enveredar. Ainda assim, existem dois temas dentro da ontologia os quais Dennett considera relevante abordar, que são a consciência e os padrões. Já falamos sobre consciência, então venhamos para os padrões.

Se um filósofo é convidado a pensar a respeito de ontologia, em geral ele tentará argumentar a favor de uma das seguintes posições: (1) realismo a respeito de X, (2) antirrealismo a respeito de X (por exemplo dizendo que X é uma ficção útil), ou (3) quase-realismo a respeito de X (dizer que não só X é uma ficção útil, como que ser uma ficção útil já é suficiente para constituir realidade, ou quase-realidade, a alguma coisa).

Não é o caso de Dennett, que em geral procura pensar essas questões na forma: ¿O que é como X? Por exemplo: ¿Seria X mais como uma bateadeira, ou mais como o tempo, ou mais como um elétron? ¿Seria X como uma voz? Ao invés de postular graus canônicos de realidade, e seccionar o inventário do mundo em diferentes classes, a ideia é dizer que o que se pode dizer a respeito da realidade de algo é, principalmente, e em primeiro lugar, que tipo de coisa é essa coisa. Por ora, devemos nos contentar com descrever o próximo passo como sendo analisar o objeto de acordo com o artigo “Real

patterns” (Dennett, 1991b). Esse tópico será mais vastamente abordado no capítulo 13 sobre padrões. Após fazer isso, sugiro procurar as razões pelas quais aquele padrão particular está recebendo uma atribuição de intencionalidade, isto é, porque se acha que ele é a respeito de alguma coisa, e finalmente procurar a resposta crucial, que a é a resposta a questão do ‘¿Cui bono?’ ¿Quem está se beneficiando? ¿Que espécie, grupo, cultura ou sistema ganha algo com a capacidade de detectar e interagir com certo padrão? ¿Para quem, em última instância, esse recorte ou pedaço do mundo pode ser visto, ou tem de ser visto, como um padrão?

Isso nos dá o algoritmo da Fig. 5:

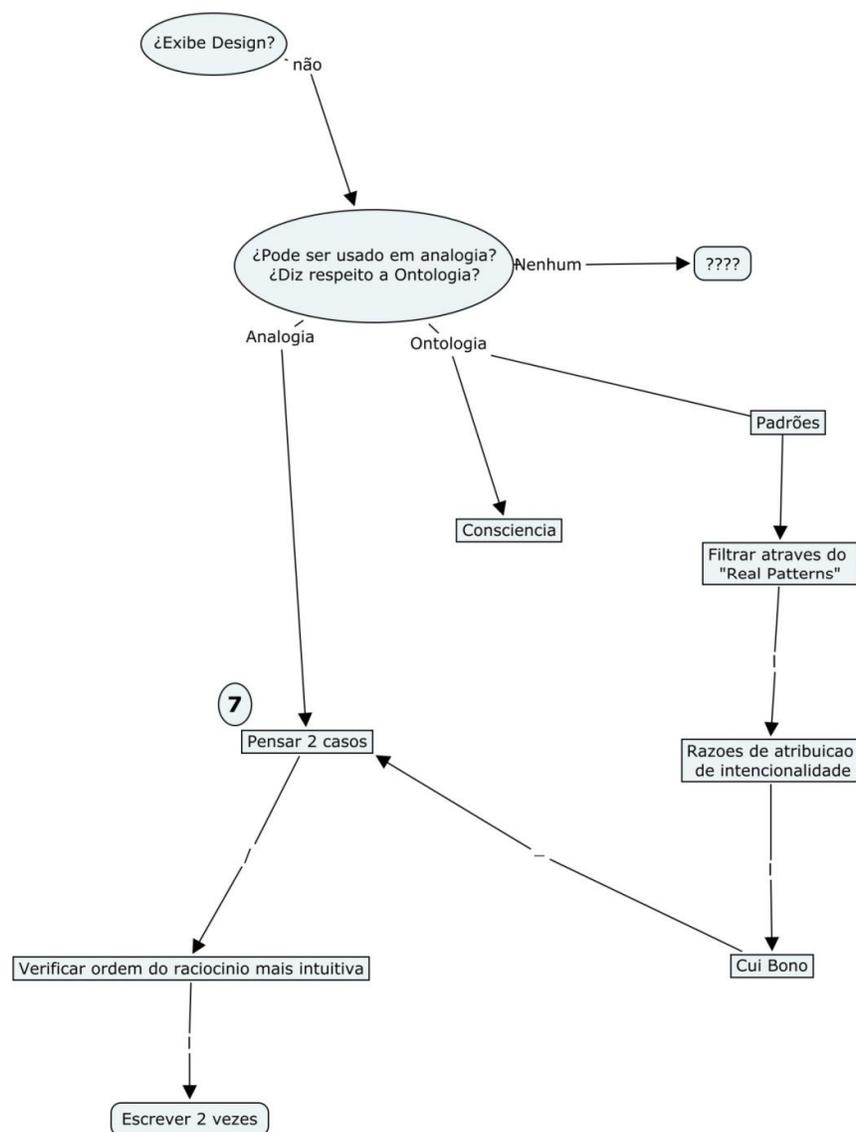


Figura 5: Algoritmo dennettiano no caso de um objeto que não exibe design.

Esses três pedaços mais lineares começam a nos mostrar como se porta o algoritmo que estou tentando desenvolver aqui, mas evidente que apenas quando juntarmos todas as peças, ao final, estaremos diante de uma rede que pode ser utilizada para simular Dennett a contento. Em breve chegaremos a esses passos mais avançados de nossa arquitetura.

Uma vez Richard Feynman, famoso físico, disse algo como “Não interessa quão bela sua teoria é, não faz a menor diferença quão inteligente você é, se os resultados discordam do experimento, ela está simplesmente errada”. Quero fazer uma adaptação dessa frase. Explico: até aqui, ao construir o algoritmo, parti do ponto inicial que é um assunto particular. No entanto, não é assim que ocorre na vida, não temos de início uma voz interior que nos diz o reto caminho sobre o que vai satisfazer nossas necessidades e curiosidades, temos apenas um fluxo constante de possibilidades novas e oportunidades que se abrem, muitas mais do que poderíamos tolerar ou entender. Então é errado começar uma explicação de como alguém pensa pressupondo um assunto dado, isso é algo que discorda da realidade, é simplesmente um equívoco. É muito mais condizente e razoável dizer que o que nos ocorre é tropeçarmos em assuntos particulares, e para entender como alguém pensa talvez seja até tão relevante ver o processo de filtragem ou de seleção, a partir dessa pleora de assuntos, quanto é relevante ver como essa pessoa pensa um assunto em particular. Assim sendo, no próximo caminho linear, partirei não mais de um assunto dado, mas sim do fato cotidiano que somos sistematicamente acometidos por impactos de um ou outro tema, que por qualquer razão entra em nossos interesses. O que acontece, de fato, é que tropeçamos num assunto.

Temos então aqui um assunto, não está esculpido nem tolhido, não necessariamente veio de um bom lugar, não há qualquer razão para sabê-lo importante, então ¿qual um bom algoritmo de seleção? Isto é, ¿qual um algoritmo de seleção que nos faria dedicar o mínimo de tempo para decidir a relevância de um assunto?

Sugiro os seguintes procedimentos de checagem. Primeiro, verificamos se ele é um assunto atual e conhecido, suponhamos que não, que é um assunto que parece relevante (para quaisquer que sejam nossos propósitos⁵³) mas que por outro lado não

⁵³ Lembrem que defendi mais acima que, para Dennett, fazer filosofia está associado a fazer a diferença, diferentemente de outros filósofos, com convicções diversas. A escolha pessoal de propósitos não estará incluída nesse algoritmo por duas razões: a primeira é que eu não faço ideia de como se poderia exhibir, ou entender, um processo de decisão que nos leva a ter os propósitos que temos, não porque eu não acredite que possa haver um, talvez até possa, mas porque se existir, seu nível de complexidade

está sendo abordado atualmente. Nesse caso, sugiro procurar informação acadêmica sobre um tema parecido. Uma vez que isso esteja feito, podemos elevar o olhar a respeito do tema, não pensando-o em referência a seu assunto, mas pensando nele como meme. Isto é, pensar em como esse tema surgiu, porque ele atingiu o grau de relevância memética que tem nesse momento, o que o levou a não ser abordado atualmente, quais as propriedades internas desse meme que o tornam um bom ou mau partido para ser preservado no futuro da evolução cultural etc. Isto é, dar um olhar memético, visualizando a inserção histórica desse tema, não como um historiador normalmente faria, mas como um biólogo ou um geneticista populacional o fariam. Simplesmente nos dispormos a pensar dessa maneira sobre a evolução de um assunto já pode nos trazer diversos insights a respeito de sua importância (em contraste por exemplo a seu grau de divulgação), a sua natureza, seu público alvo etc. Tudo isso pode nos trazer informação nova, que não costuma ser pensada por quaisquer outros que hajam tangenciado o tema, porque essa maneira de ver é tão particular e pouco disseminada. Como o próprio Feynman diria, às vezes conseguimos resolver um problema simplesmente porque temos uma caixa de ferramentas diferente de quem já tentou.

Uma vez que o assunto já esteja pensado à luz da memética, cabe pensá-lo também da perspectiva intencional. Finalmente pode-se avaliar o grau de influência do tema levando em conta esses passos anteriores. Se a influência parecer alta, então aí sim chegaremos ao passo no qual se iniciaram todas as descrições anteriores. Somente após esse processo de triagem ter sido executado é que passa a *valer a pena* estudar um assunto a fundo, ou executar o resto dos procedimentos descritos até agora. Como esse processo inicial parece mais aleatório, ele é muitas vezes menosprezado, ou mais particularmente desconsiderado, e creio que seja justamente esse o problema da vida intelectual de muitas pessoas. Como Dennett diz em uma palestra: “If it is not worth doing, it is not worth doing well”.⁵⁴ Esse conselho é especialmente importante no Brasil, onde é bastante conhecido o fato de que existe uma passividade intelectual

provavelmente não poderia ser resumido tranquilamente dentro de uma visualização simples. Quando executamos qualquer programa de computador, não estamos utilizando todas as suas funções, mas aquela que nos cabe melhor naquele momento. Da mesma maneira, podemos pensar que (1) estou aqui expondo aquelas partes que considero mais interessantes e compartilháveis do que entendo como um algoritmo de simulação de Dennett, e que (2) o simulador deixará mais salientes ou menos salientes em sua própria reconstrução interna do algoritmo aquelas partes que pareçam relevantes para ele, e portanto que se adequam ao seus propósitos. Se (2) for verdadeiro, então os propósitos de Dennett seriam irrelevantes para o simulador, e por consequência, não seriam interessantes de serem compartilhados, e nem escritos, de acordo com (1).

⁵⁴ “Se não vale a pena fazer, não vale a pena fazer direito.” O aforismo é originalmente do psicólogo Donal Hebb (Dennett, 2006b, p. 40).

grande com relação aos conteúdos que nos são apresentados (sejam vindos do exterior, sejam de nossos professores). O não se perguntar a respeito de haver outras opções melhores por aí pode minar os talentos de um gênio, transformando-o num brilhante estudante de pormenores irrelevantes.

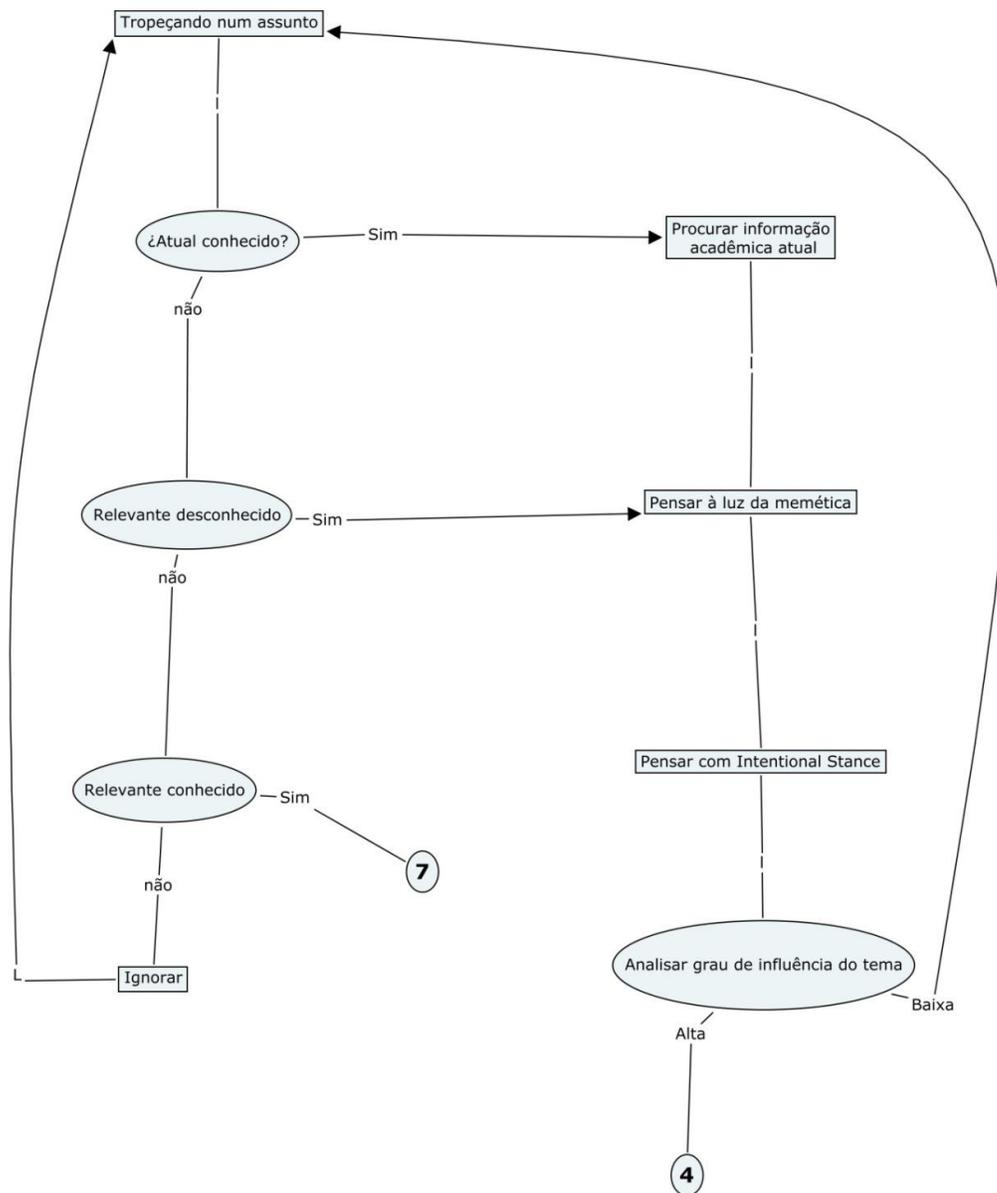


Figura 6: Heurística inicial ao tropeçar em um problema.

¿Qual a diferença entre pensar algo à luz da memética e pensar algo como produto da evolução memética? Pensar à luz da memética é tentar entender como o assunto está sendo tratado na mídia, é entender como a *teoria* está sendo replicada, como chegou até você, quais são as propriedades atrativas da teoria. Uma teoria polêmica ou moralmente controversa geralmente tem maior força de replicação que suas rivais, *independentemente de evidências a seu favor*. Pensar sobre a evolução memética de algo, por outro lado, é pensar como o *objeto* evoluiu memeticamente. A diferença centra-se em estar tentando entender o objeto, digamos, a religião islâmica, ou a teoria, digamos a teoria que diz que a religião islâmica evoluiu a partir de proto-religiões menos eficientes no Oriente Médio antigo. Saber como as teorias evoluem serve para evitarmos alguns bias⁵⁵ ou vieses muito fortes a respeito de sua veracidade, como o bias de acreditar mais em algo que é exposto com maior frequência, chamado ‘efeito da mera exposição’.

Também sobre um aspecto meta-teórico é a caixa “Pensar com Intentional Stance” (Fig. 6). Pode-se avaliar antes de adentrar um tópico quais intenções, desejos e crenças estão envolvidos na transmissão desse conhecimento que chegou até nós. Fazemos isso naturalmente, é claro, por exemplo suspeitando de reportagens que falem bem a respeito de um novo medicamento.

Após esta introdução a alguns dos fluxos particulares que o algoritmo pode tomar, com um olhar mais nítido sobre ele, posso mostrá-lo inteiro e apenas comentar alguns aspectos que ainda não haviam aparecido e seus porquês. A forma quase final do algoritmo está na Fig. 7, e a forma final foi exibida anteriormente, no capítulo 4, Fig. 2.

Alguns pedaços novos merecem comentários. O retângulo que diz “Verificar ordem do raciocínio mais intuitiva” é um deles: ao pensar um intuition pump, uma analogia, ou qualquer forma de dizer algo em termos de outro algo, é fundamental saber qual a sequência adequada para introduzir o novo tema, afinal, conforme o discurso vai se compondo, algumas partes tornam-se mais ou menos salientes, o que por sua vez pode tornar a analogia mais ou menos eficiente em transmitir o que quer que se queira passar com ela. Compare por exemplo as seguintes exposições da mesma ideia. Cada ideia será representada por uma sequência de números de um a nove.

⁵⁵ *Bias* ou *vieses cognitivos* são erros sistemáticos que nosso sistema cognitivo comete porque somos *hardwired* (predispostos por nossa formação estrutural, de maneira inata) de uma certa maneira que não permite um raciocínio ideal naquelas condições. Para uma lista: http://en.wikipedia.org/wiki/List_of_cognitive_biases

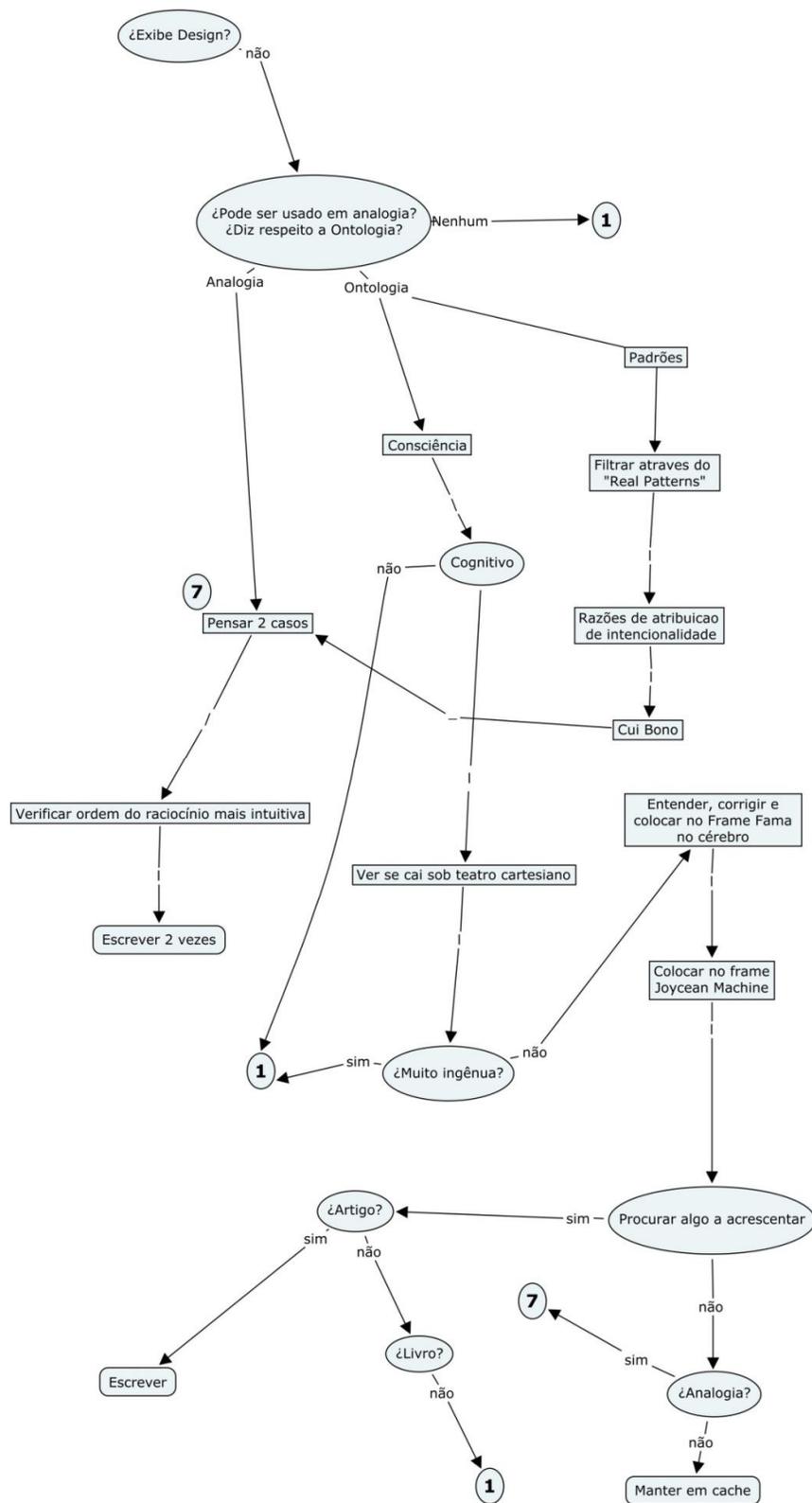


Figura 7: Forma quase final do algoritmo dennettiano. Para a final, ver Fig. 2, p. 63.

1. O curso da evolução é produto da seleção natural.
2. Quem é selecionado na seleção natural são os genes.
3. Os indivíduos muitas vezes são descartados em prol dos genes, a natureza age contra o indivíduo (por exemplo acelerando o envelhecimento dele) em favor de algum gene.
4. A natureza portanto é muitas vezes prejudicial aos indivíduos.
5. Existem propriedades humanas das quais não gostamos, como o fato de que pessoas morrem jovens, podem ter doença de Alzheimer, a possibilidade de contaminação por malária, o câncer.
6. Temos boas razões para combater aquilo que nos gera sofrimento, e por isso utilizamos, por exemplo, a medicina.
7. Temos a obrigação moral de auxiliar aqueles que estão em desespero ou em estado miserável.
8. Essa obrigação moral não cessa com relação à distância (África Subsaariana) ou diferença entre pessoas (negros, gays, crianças, velhos, carecas etc.)
9. Temos obrigação moral de modificar a condição humana, independentemente de se isso é contra ou a favor da natureza.

Compare essa sequência acima com a seguinte cena: Fórum social mundial, sessão sobre vegetarianismo, centenas de ávidos ouvintes procurando soluções para um mundo melhor e mais harmonioso, entra um palestrante e inicia:

9. Temos obrigação moral de modificar a condição humana, independentemente de se isso é contra ou a favor da natureza.
8. É necessário modificar a condição humana principalmente de gente como os velhos, os carecas e os africanos subnutridos.
7. Essas pessoas as vezes vivem num estado miserável, do qual é necessário tirá-las.
6. Interferir na natureza para eliminar doenças, problemas, utilizar a medicina para interferir na vida das pessoas e retirá-las de seus estados de sofrimento.
5. ...
4. ...
- ...

Não há como não se sentir tocado pelo argumento do primeiro, e entender que, de fato, para fazermos um mundo melhor, modificar a condição humana é essencial. Por outro lado, nenhum dos ouvintes vegetarianos no Fórum Social Mundial tem qualquer

chance de não morrer de ódio do palestrante que inicia sua palestra com tamanho descaso pela natureza. A ordem dos fatores altera o produto a tal ponto que diferentes ordens simplesmente estão falando de diferentes produtos.

Todo o fluxo a partir da decisão de que um tema é relativo à consciência ainda não foi comentado nesse capítulo, mas um recurso ao capítulo sobre consciência provavelmente daria conta da maioria das preocupações que possam surgir. Ainda assim, algumas valem a exposição, ¿Porque ignorar aquilo que não é cognitivo a respeito da consciência? Essencialmente porque se não vale a pena fazer algo, não vale a pena fazer direito, e existe muita gente muito inteligente fazendo ótimos trabalhos nessa área, o que seria um vórtice enorme para o tempo de uma maneira improdutiva, ou seja, é algo no que se deve dar o drible. ¿Como colocar algo no quadro da máquina joyceana? Uma maneira interessante é justamente tentar entender a teoria, ou explicação, ou objeto que se está tentando analisar, utilizando-a sobre uma história contada como fluxo de consciência. Seja uma história de Joyce mesmo, ou simplesmente deixar-se passar por um fluxo mental por algum tempo, e analisar quão condizente parece essa sequência de eventos com o que quer se esteja propondo nessa particular análise da consciência, ou nessa teoria, experimento etc. Em suma, verificar se a teoria condiz ou não com a experimentação, através de uma experimentação literária, ou de qualquer tipo de fluxo de consciência.

Concluimos aqui a formação de um algoritmo dennettiano. Algo que se nota é que, apesar de bastante grande, o algoritmo é bastante simples, isso porque quando excluimos aquilo que é típico da espécie humana (na verdade dos leitores de filosofia), e os pormenores particulares dos sub-temas e micro-interesses específicos, não sobra mais tanta informação a respeito de como alguém pensa. Isso vale para qualquer pessoa. Evidente que poderíamos ter qualquer grau de espessura para fazer nossa análise. Um algoritmo de alguém pode variar do infinitamente fino, uma cópia micromolecular do cérebro⁵⁶, ao infinitamente grosso, como um modelo skinneriano aplicado a um humano, em termos de descrições simples de estímulo e resposta. O que nos interessa como apreciadores de um sujeito em geral são os traços intermediários, aqueles que diferenciam Einstein e Dennett do idiota do vilarejo. Estes traços constituem uma diferença de inteligência que pode nos parecer astronômica, mas que é pequena quando

⁵⁶ Filósofos funcionalistas diriam que uma cópia isofuncional do cérebro é melhor do que uma cópia micromolecular, na medida em que uma cópia que tenha as mesmas funções é um modelo cognitivo adequado mesmo que o corpo seja colocado em outra posição no espaço e no tempo. E estariam certos em dizer isso. Só usei 'micromolecular' porque passa a ideia de melhor cópia possível de algo.

comparando inteligência com coelhos, esquilos e computadores. A razão pela qual as diferenças nos parecem tão grandes é que, como um primata social que somos, temos detectores precisos dessas diferenças intra-específicas de inteligência, já que no passado nossos ancestrais com esses detectores fizeram melhores alianças, conseguiram mais comida, mais sexo e mais filhos. Assim como conseguimos diferenciar faces humanas mas temos dificuldade de fazer o mesmo com orangotangos, temos facilidade de discernir maneiras de cognição e pensar humanas e suas diferenças, e dificuldade de fazer o mesmo com golfinhos.

¿Isso significa que estou argumentando que nesse algoritmo estão todos os segredos de Dennett, todo o elixir do conhecimento que transborda suas sinapses e correntes corticais? Evidentemente que não, nem no algoritmo, nem em todo o texto, nem em todo o material compilado escrito por ele. Mas ao menos temos aqui uma ideia muito mais clara e bem dividida do pensamento dennettiano do que tínhamos no começo do texto. Os próximos capítulos continuarão guiando um zoom maior em diversas partes desse algoritmo, de modo que cada re-olhada direcionada a ele poderá se tornar mais rica de informação semântica, conforme a leitura progredir. No próximo capítulo entraremos mais a fundo na intencionalidade, falando de sistemas intencionais, da origem do vocabulário em termos de intenções, e da necessidade do plano de análise intencional.

12. Intencionalidade, psicologia popular e a necessidade do plano de análise intencional

Douglas Adams (2004) descreve o espaço da seguinte maneira:

O Espaço é grande.
Realmente grande.
Assim, você simplesmente não vai acreditar quão vastamente, enormemente, desmioladamente grande ele é.
Você pode achar que é um longo caminho até a farmácia, mas isso são ervilhas no espaço.

Em todo este vasto espaço, mais de 99,999% não possuem uma propriedade que nos interessa muito. Com efeito, essa propriedade da matéria está envolvida em quase tudo o que nos interessa, e a história do mundo simplesmente não seria a mesma sem ela. Ela só existe, até onde sabemos, no centro de uma região de cento e cinquenta anos-luz, ao redor de nosso pequeno planeta azul. Sabemos que essa esfera está aumentando quase na velocidade da luz, e está sendo transmitida por ondas eletromagnéticas. Não é a vida, mas dependeu da vida para surgir. Quanto mais nos afastamos da terra, menor sua quantidade. Essa propriedade é a *intencionalidade*.

‘Intencionalidade’ é uma dessas palavras que, infelizmente, têm um significado diferente no vocabulário especializado da filosofia, e no vocabulário do dia a dia. “Mas ninguém fala ‘intencionalidade’ no dia a dia”, objeta o astuto leitor. De fato, mas falamos o tempo todo de intenções, de qual era a intenção dele ao fazer isso ou aquilo. Para piorar, intencionalidade se refere a dois aspectos muito próximos em alguns casos, e bastante distintos em outros.

Intencionalidade₁: A propriedade de algumas coisas de *ser a respeito de outras*. Exemplo: numa placa de zoológico está escrito “Esta tartaruga chegou no zoológico em 1999”; as palavras ‘esta tartaruga’ são a respeito da tartaruga. Existe uma estrutura física material, feita de tinta numa placa de metal, e esta estrutura é, num sentido importante, a respeito da tartaruga que está dentro da jaula.

Além de placas, muitas outras coisas possuem intencionalidade, como mapas, livros, aquelas curiosas sequências de sons guturais e estalidos da boca e língua que chamamos de “palavras”, e os códigos magnéticos 010101 que gravam o nome do seu melhor amigo no seu celular. De maior interesse para a filosofia, foi sugerido no início

do século XX que a propriedade essencial dos estados mentais seria a intencionalidade₁. Se eu *creio*, eu creio em monstros, se *desejo*, desejo um monociclo, se tenho *medo*, tenho medo de aranhas. Ou seja, propunha-se que os estados mentais se distinguem dos físicos por serem a respeito de algo.

Intencionalidade₂: Propriedade daquilo que tem intenções. “Ele andava desanimado, caído, quase em estado vegetativo, mas parece ter recuperado a intencionalidade₂ e agora está cheio de projetos.” “No início, era apenas uma célula, depois foi se desenvolvendo, ganhou um sistema nervoso, e agora é uma criança cheia de intencionalidade”.

A razão pela qual esses dois sentidos são bastante próximos em alguns casos é que todo ato que é intencional₂ é também intencional₁. Se eu tenho a intenção de caçar uma baleia, eu tenho um estado mental do tipo “intenção” que é a respeito de um tipo de ato, a caça, e uma ordem de animais, as baleias. Ou seja, o meu estado mental, que é justamente o que determina que eu seja um sujeito com intencionalidade₂, é ele mesmo uma coisa do mundo que possui intencionalidade₁.

Sempre que uma entidade (coisa) tem intenções, boas ou más, alguma parte dela ganha de quebra a intencionalidade, porque não há intenções sem objeto. O objeto não precisa ser físico, alguém que tem intenções de matar tem um estado mental que é a respeito de um ato, mas esse ato não necessita ter um objeto. Caso haja um objeto, a intencionalidade₁ também não exige que o objeto seja real. Se uma placa diz “Sherlock Holmes esteve aqui”, ela é a respeito de Sherlock Holmes.

Para os filósofos da mente, a principal questão referente à intencionalidade₁ é tentar entender e explicar como é possível que algo seja a respeito de outra coisa. Uma pedra ou um vulcão não são a respeito de nada, no entanto, a Pedra da Rosetta está carregada de intencionalidade em seus escritos trilíngues. É muito estranho que seja possível que uns certos pedacinhos de matéria sejam a respeito de outros, ou que uma estrutura seja a respeito de um conceito abstrato, como uma soma ou multiplicação. No entanto, isso acontece.

Existem algumas teorias principais a respeito da intencionalidade. Uma teoria sobre a intencionalidade é aquela que pretende explicar fatos do tipo: “A se refere a B”. Essas teorias são chamadas de “teorias da referência” quando A é uma palavra de uma linguagem. Ou seja, uma teoria da referência é uma sub-área de teorias da intencionalidade. Existe uma diferença grande entre uma palavra como “tartaruga” e um *token* (ocorrência) da palavra “tartaruga”. A palavra “tartaruga” é uma entidade abstrata,

pertencente a uma organização formal altamente complexa que é o português. Isso é diferente de uma placa na qual está escrito tartaruga: a palavra “tartaruga” está ocorrendo de fato na placa, e naquele contexto ela se refere àquela tartaruga particular da jaula que chamaremos de Pescoçuda. Aquele token se refere à Pescoçuda. O *type* “tartaruga”, ou seja, a palavra “tartaruga”, se refere a muitas coisas em diferentes contextos, seja a Pescoçuda, seja a Donatelo e Michelângelo, as tartarugas ninjas, seja à espécie como um todo.

Uma teoria da referência está interessada em como os *types* se relacionam com seus objetos nas linguagens humanas. Uma teoria da intencionalidade é mais abrangente, ela se interessa por toda e qualquer ocorrência de algo que remete a outro algo. Pode-se argumentar que parte do meu código de DNA tem intencionalidade, até dupla intencionalidade, na medida em que é a respeito de, num certo nível, a produção de algumas proteínas, e em outro nível, em características físicas minhas, como o formato do meu nariz, ou algumas formações cerebrais que me predispõem a gostar de ciência.

Existe um sentido muito fraco no qual podemos dizer que a radiação astronômica de fundo, que vem dos confins do espaço, é “a respeito do Big Bang”. Isso porque é através de exames dessa radiação que chegamos a conclusões sobre o Big Bang e o começo do tempo. Existe um sentido muito forte no qual podemos dizer que o grande ‘M’ amarelo na frente de uma filial do McDonald’s se refere ao McDonald’s. Em algum ponto entre esses, está o limiar que interessa aos filósofos da mente e em particular a Dennett.

O grau de força de conexão que há de haver entre algo que refere e algo que é referido, para que estejamos falando de intencionalidade, é arbitrário, e é importante notar que existe variação em quão bem e precisamente algo refere. Os lugares mais distantes do espaço em que encontramos intencionalidades do tipo interessante a nós são as ondas eletromagnéticas transmitidas pelas primeiras estações de rádio.

Existem várias teorias a respeito de intencionalidade, como já mencionei. Dois tipos são os básicos, os que discutem como a intencionalidade funciona, e os que discutem a origem da intencionalidade. Por vezes as duas se misturam. Uma teoria que proponha que “a origem da intencionalidade é uma relação de referência que ocorre espontaneamente entre a mente humana e, digamos, uma mesa” é uma teoria de origem e referência ao mesmo tempo. Examinemos algumas teorias mais críveis.

Teoria Causal da Intencionalidade.⁵⁷ A partir do “batismo”, um objeto (por exemplo, uma pessoa) ganha um nome, que é atribuído ostensivamente (apontando, por exemplo). Desse momento em diante, existe uma sequência causal que conecta aquele nome àquele ato de batismo. Toda vez que me refiro a Dennett, seria possível (teoricamente) traçar uma sequência causal passando pelo meu cérebro, por eu ter ouvido Dennett, por Dennett escrevendo seus livros e chegaria, indo mais e mais para trás no dia em que seus pais disseram, “vamos chamá-lo de Daniel, Daniel Dennett”. Dessa maneira se explica a conexão desses sons com aquela pessoa.

Teoria Cluster da Intencionalidade. A intencionalidade de um objeto é determinada por uma rede conceitual. Uma ocorrência da palavra ‘religião’ num livro em geral se refere a uma série de conceitos distintos que podem determinar uma religião, sem que nenhum deles seja essencial. Elenco alguns desses traços, para ilustração: religiões têm Deuses, religiões envolvem comportamento ritualizado, religiões envolvem fé, religiões envolvem livros sagrados, religiões envolvem encontros sociais, religiões envolvem orações, religiões envolvem princípios morais, religiões envolvem dogma.

Nenhuma dessas propriedades é essencial a uma religião, isto é, é possível ser uma religião mesmo que não tenhamos uma dessas propriedades em particular. No entanto, a maioria das religiões possuem algumas, se não a maioria dessas propriedades. Segundo a teoria cluster da intencionalidade, o que determina a referência de uma ocorrência de ‘religião’ num texto é o agrupamento de todas aquelas coisas que têm a maioria dessas propriedades (ou de um grupo de propriedades mais adequado escolhido como determinante).

Teoria Mágica da Intencionalidade. Como logo se vê pelo nome, essa teoria não é tanto uma teoria como o nome utilizado por filósofos para acusar uns aos outros de não estarem sendo explicativos com suas teorias sobre intencionalidade. Segundo a teoria mágica da intencionalidade, existe uma flecha invisível mágica que conecta o referente e aquilo que a ele se refere, e existe uma dessas flechas para cada ocorrência

⁵⁷ A teoria causal e a teoria *cluster* (de agrupamentos) são geralmente vistas como teorias de referência, não de intencionalidade, mas para nossos propósitos, elas podem ser pensadas como teorias de intencionalidade. Alguns proponentes dessas teorias são: causal (Saul Kripke), *cluster* (John Searle), mágica (senso comum), sintática (Dan Dennett), evolutiva (Hilary Putnam), semântica originária (Jerry Fodor).

de um referente. Ou seja, ela simplesmente postula aquilo que pretende explicar.

“The method of ‘postulating’ what we want has many advantages; they are the same as the advantages of theft over honest toil. Let us leave them to others and proceed with our honest toil.”⁵⁸

Vale notar que, como a grande maioria das pessoas nunca pensou no problema da intencionalidade, e simplesmente assumiu que os termos referem, existe num certo nível uma pressuposição tácita da teoria mágica em nossa omissão.

Teoria Sintática da Intencionalidade. O mundo está cheio de sistemas físicos, disso não há dúvida. Há dúvida se de fato é possível que esses sistemas físicos representem ou sejam a respeito de outros sistemas físicos. Que fazer? Podemos supor que o cérebro é uma máquina sintática, de processamento puramente mecânico, que aproxima-se bem o suficiente dos objetos intencionais para que nossos pensamentos tenham a eficácia causal esperada, ao menos na maioria das vezes. O mesmo vale para a linguagem, podemos ver as palavras da linguagem como sistemas físicos que, mecanicamente, fazem um trabalho tão parecido, mas tão parecido com o trabalho de significação, de referir-se a um objeto, que vale a pena observá-las *como se* elas estivessem de fato significando algo. Em computação, símbolos são partes do sistema que têm efetividade causal, que interferem no funcionamento do sistema de uma maneira específica. Normalmente, pensaríamos esses interferidores de uma maneira específica como algo que tem *significado*, isto é, como *signos*. A diferença entre *símbolos* e *signos* é que no caso dos símbolos, as propriedades que lhes dão a função que têm são propriedades físicas. Isso significa que sua constituição material é relevante para seu significado. A teoria sintática da intencionalidade dirá que toda a intencionalidade que há por aí é uma aproximação (muitas vezes incrivelmente boa) de uma verdadeira semântica, uma verdadeira significação. Ainda assim, no fundo, no fundo, estamos falando apenas de sistemas sintáticos, físicos, mecânicos que se comportam de tal maneira que parece que estão magicamente conectados com os objetos a respeito dos quais são. Isso se dá porque esses sistemas possuem símbolos, e todo o processo de referência é o processo através do qual alguns símbolos interpretam outros símbolos de certa maneira, como se fossem referentes a algo externo ao sistema.

⁵⁸ “O método de ‘postular’ o que queremos tem várias vantagens; são as mesmas vantagens do roubo sobre o trabalho honesto. Deixemos essas vantagens aos outros e continuemos com nosso trabalho honesto” (Russell, [1919] 2008, p. 71).

Aqui não é o ponto em que eu esperaria que você teria entendido a teoria sintática da intencionalidade. Ao fim da lista farei uma descrição bastante mais completa e minuciosa dessa teoria bastante complexa, para que possamos seguir adiante sem o desagradável sentimento de que faltou alguma coisa fundamental aqui.

Teoria Evolutiva da Intencionalidade. Somos seres vivos. Seres vivos complexos, como nós, o milho, e as baleias são resultado de recombinação e mutação gênica ao longo de bilhões de anos. A seleção natural opera no nível do gene, isto é, o gene é “egoísta” e os genes mais adaptados vão dominando o cenário genético ao longo do tempo. Nós e as baleias temos sistemas nervosos organizados por nossos genes para que consigamos lidar com o mundo. Em particular, temos de sobreviver, comer e nos reproduzir. Tudo isso requer comportar-se no mundo de maneira adequada. Para comportar-se de maneira adequada, é necessário, ao menos num certo nível, construir representações não arbitrárias e não absurdas do mundo. Ou seja, temos crenças aproximadamente verdadeiras ou verdadeiras porque se não tivéssemos, não deixaríamos descendentes, e o mesmo argumento vale para nossos pais. Segundo a teoria evolutiva da intencionalidade, a razão pela qual existe uma adequação entre aquilo que significa, e aquilo que é significado, seja em nossas mentes ou nas construções de nossa cultura, é que se não houvesse um certo grau de conexão, de adequação, simplesmente não estaríamos aqui para contar a história. Digamos que outra espécie surja, e que todos os seus termos se refiram a coisas completamente diferentes, e não haja nenhum isomorfismo estrutural entre suas crenças, desejos e linguagem, e a forma como o mundo se comporta. Eles seriam extintos em uma geração, ou melhor, em alguns minutos. Simplesmente não é viável se jogar em um planeta com um arcabouço de teorias falsas, de sistemas cognitivos organizados em desacordo com o planeta; daí que, onde haja vida complexa, haverá intencionalidade.

Teoria da intencionalidade originária humana. Uma maneira diferente de olhar a questão da intencionalidade é negar que exista intencionalidade em coisas materiais, como sentenças de linguagem natural, ondas eletromagnéticas etc. Os autores que seguem essa linha de raciocínio (Fodor, Searle) dirão que as sentenças de uma linguagem e seus *tokens* (ocorrências) não possuem significado a não ser quando ele é conferido por pessoas, que interpretam essas sentenças e que falam essas sentenças para expressar seus pensamentos. Asserções tomariam de empréstimo sua intencionalidade

derivada da intencionalidade *original* dos humanos com mentes que utilizam-na para seus propósitos. Se, como argumenta Fodor, existir uma linguagem do pensamento, com propriedades sintáticas e semânticas, talvez a intencionalidade originária seja uma propriedade dos símbolos mentais. A intencionalidade seria, em primeira e última medida, uma propriedade das pessoas e seus pensamentos.

Essas diversas teorias a respeito da intencionalidade não são necessariamente excludentes entre si. Algumas combinações são. Dennett defende uma teoria da intencionalidade que combina a abordagem evolutiva com a abordagem sintática.

O aspecto sintático está presente no pensamento dennettiano na medida em que não existe realidade ontológica⁵⁹ em estados intencionais. O que lhes confere intencionalidade é uma série de nuances sintáticas, nuances funcionais que os fazem comportar-se de uma certa maneira. Essa certa maneira de se comportar desses estados físicos faz com que eles se pareçam correlatos de certas outras entidades em algum lugar, e nessa medida esses estados sintáticos finos⁶⁰ são representações daqueles objetos com os quais seu funcionamento está correlacionado. Pode-se pensar, para começar a construir uma intuição a respeito, que estamos falando de algum tipo de isomorfismo funcional, isto é, funcionamento equivalente de dois sistemas nos seus aspectos relacionais. Ou ainda, operações paralelas nos dois sistemas chegam ao mesmo resultado quando traduzidas de uma certa maneira. Para uma definição de semelhança de relações, ver Russell ([1919] 2008, pp. 52-4).

Dennett então espousa uma visão da intencionalidade na qual uma série de símbolos sintáticos possuem relações quase-isoestruturais com objetos da natureza, e a razão pela qual possuem essas relações depende de suas propriedades físicas. Isso é, o tamanho, densidade, velocidade, temperatura, condutividade elétrica etc. de um pedaço de matéria é relevante para que ele tenha, dentro do sistema no qual está inscrito, o papel que ele tem. O papel que ele tem, por sua vez, é quase-isoestrutural ao papel que aquilo ao que dizemos que ele se refere tem na natureza.

Um exemplo: “Não pise na grama” escrito numa placa nos gramados de Harvard

⁵⁹ Uma definição simples e útil de “realidade”, compatível com o uso nessa sentença, é a de David Deutsch (1997, p. 87), físico famoso por sua defesa do multiverso: “Se algo chuta de volta, é real”. Ou seja, algo é real na medida em que tem efeitos que podem ser sentidos ou detectados de alguma maneira. Lembrando, “ontológico” é o que é relativo a coisas, ou o que compõe o mobiliário do mundo. Dennett não utilizaria o termo aqui, preferindo, como expliquei previamente, comparar o grau de realidade dos estados intencionais com o de uma voz, ou com a saúde, por exemplo.

⁶⁰ *Narrow syntactical states*, em inglês.

é feito de tinta numa placa de metal. Essa tinta está inscrita em um sistema complexo envolvendo o reflexo da luz do sol, os olhos e cérebros dos alunos, e as correntes corticais que se ativam ao ler a placa, e que, em média, correlacionam com pessoas não pisando na grama. Além de gerar o comportamento de não pisar na grama, se insere num contexto maior em que alunos discutem extenuantemente sobre se se deve ou não pisar na grama, se a grama é ou não uma entidade real, se é ou não possível pisar e não pisar na grama ao mesmo tempo etc. Todo esse sistema intrincado possui uma série de relações internas, e aquela placa tem relações que são isoestruturais a algumas das relações envolvendo grama, pessoas e o ato de pisar. Nessa medida, essa placa se refere a não pisar na grama.

¿Porque um computador que joga xadrez consegue jogar bem? O computador não pode ver o tabuleiro, ele apenas “vê” (acessa informação de) uma série de comandos entre estruturas matemáticas que se conectam de um ou outro modo entre si. A razão pela qual o computador consegue jogar xadrez é que seus programadores escolheram objetos matemáticos e computacionais que são isoestruturais a um jogo de xadrez e colocaram-nos como input no computador. Então as relações que existem no tabuleiro (as relevantes) estão todas lá. Não estará presente a tonalidade do bege do bispo das brancas, mas sim a informação matemática dos movimentos possíveis do bispo, etc. Um cientista de computação mais ortodoxo poderia até afirmar que um tabuleiro de xadrez é apenas uma bugiganga dispensável que ajuda as intuições humanas a fazer cálculos, dispondo informação matemática com cara de guerra.

Imaginar a abordagem sintática à intencionalidade é imaginar que as ocorrências de palavras, placas, jogos de tabuleiro, estados como desejar uma maçã, ter medo de um lobo, acreditar em OVNI, querer conquistar o mundo, são todos símbolos que operam em um determinado sistema de forma quase isoestrutural, ou quase funcionalmente equivalente, à forma que os objetos aos quais se referem se comportam no sistema no qual esses objetos estão inseridos.

A abordagem evolutiva se funde à abordagem sintática na filosofia de Dennett. Isso acontece no seu livro *Kind of minds* (1996a), sobre o qual já falamos no capítulo 6, “Tipos de mentes”. Poderíamos fazer uma recapitulação breve para encaixar as peças das duas teorias e mostrar uma forma mais sofisticada de teoria da intencionalidade, mas o próximo capítulo será uma recapitulação do caminho visto até aqui, e portanto continuaremos com a discussão de o que é e porque é útil a psicologia popular (*folk psychology*).

Psicologia popular, pensando pessoas

A psicologia popular é uma teoria tácita envolvida no uso popular de termos mentais. “Marvin Minsky *acredita* que combinou com Dennett de se encontrar no topo do domo do MIT, mas *sabe* que chegará atrasado porque *esqueceu* a escada, mesmo que tivesse toda a *intenção* de chegar no horário. Além disso, *teme* a fúria dennettiana, e *sente-se culpado*.”

Existe, segundo os defensores da existência da psicologia popular (quase todos os filósofos), uma teoria implícita nessas afirmações, essa é uma teoria a respeito de que tipo de coisas são acreditar, saber, esquecer, ter intenção de, temer, e sentir-se culpado. São o tipo de coisa que funciona de maneira análoga ao uso dessas palavras na linguagem. Ou seja, acreditar é um estado correlacionado com a funcionalidade da palavra ‘acreditar’ em linguagem popular. ¿Isso significa que as pessoas de fato têm uma teoria complexa de estados mentais, mas simplesmente esquecem de comentá-la com você quando vocês sobem juntos no elevador? Não, mas significa que em nível inconsciente, tanto intra-pessoal, quanto inter-pessoal, existe uma teoria latente sobre o funcionamento do mental, que possui alguns termos típicos e usos típicos na linguagem cotidiana. Da mesma maneira temos uma teoria popular da física, segundo a qual um corpo deixado em livre movimento para depois de um tempo, giroscópios e sifões não são possíveis, etc. A existência da teoria não precisa significar: 1) Que alguém tenha estruturado ela e escrito a respeito 2) Que ela esteja clara e bem elaborada 3) Que ela seja verdadeira, e nem 4) Que todas as versões em cada ser humano sejam compatíveis. É suficiente que ela seja falsa, confusa, variada e nunca tenha sido escrita ou verbalizada, contanto que existam invariâncias fortes na grande maioria dos usos dos termos que constituem seu vocabulário.

Isso ocorre com o vocabulário mentalista dos pedestres do dia a dia, e portanto se qualifica como uma teoria popular de algo, mais especificamente, dos estados mentais.

¿Porque os filósofos se interessariam por uma teoria falsa, confusa, variada e não verbalizada? Por duas razões distintas. Primeiro, porque a existência da teoria ilustra como nossos processos de pensamento são capazes de organizar o mundo e compartimentar os mais abstratos objetos. Segundo, porque é possível que haja algo de interessante a se recuperar daí, é o rascunho inicial do qual todas as teorias da mente partiram, e é aquilo ao que temos acesso mais imediato num nível intuitivo,

diferentemente por exemplo de estruturas neurais. No início do século vinte, não tínhamos quase nenhuma compreensão de como o cérebro funcionava, e portanto, pensadores a respeito da mente como Freud, por mais que fossem favoráveis ao progresso da neurociência, tinham de se valer de conceitos populares a respeito dos estados mentais, e a partir desses conceitos formular teorias mais elaboradas a respeito do funcionamento mental.

Hoje em dia, nossa precisão de análise funcional é de cubos de aresta 2mm a 4mm, e sabe-se que o tamanho ideal seria 1,5 mm (Hyde, Biswal & Jesmanowicz, 2000), ou seja, estamos perto da funcionalidade máxima de testes funcionais de ressonância magnética. Mesmo com a existência de tecnologia desse tipo, ainda existem 400 mil neurônios em cada cubo, e por um bom tempo estamos ainda fadados a trabalhar nesse nível intermediário com a informação sobre o que se passa no cérebro quando estamos pensando.

No seu artigo “Three kinds of intentional psychology” (1981b), Dennett sugere que nós utilizamos a psicologia popular para fazer previsões. Uma das pressuposições tácitas da psicologia popular é que somos racionais. Isto é, que se acreditamos que ter um cachorro e gostar dele faz com que devamos levá-lo para passear, então uma pessoa que tem um cachorro e gosta dele *achará* que deve levá-lo para passear. Alternativamente, uma pessoa que não tem cachorro, mas conhece esta primeira, se perguntada, confirmará que é bem provável que seu amigo *ache* que deva levar o cachorro para passear. Para saber disso, é irrelevante saber se o sujeito leva ou não o cachorro para passear, é suficiente compreender seu sistema de crenças, e pressupor que ele é um ser racional com algumas crenças (eu tenho um cachorro, e gosto dele) e desejos (levá-lo para passear).

Dennett (1981b, p. 40) sugere que *crenças* poderiam ser definidas grosseiramente como “estados portadores de informação de pessoas, que surgem de percepções, e que, juntamente com *desejos* apropriadamente relacionados, levam à *ação* inteligente.”

Para que nossas teorias naturais de psicologia popular funcionem, não é necessário nenhum conhecimento do que se passa dentro da cabeça. Com efeito, gregos antigos, que em grande proporção acreditavam ser o coração o centro da cognição e das emoções, poderiam ter uma psicologia popular tão sofisticada quanto a nossa. A única razão pela qual provavelmente temos uma teoria mais sofisticada hoje é que vivemos em sociedades maiores, com vocabulário mais extenso, e amplo acesso à informação

social relevante para construção dessas normas tácitas, principalmente advinda de *reality shows*, novelas, e séries de tevê.

A estratégia intencional

Mesmo estando completamente alienadas do funcionamento interno do cérebro, a maneira como entendemos crenças, desejos e intenções, segundo Dennett, é explicativa. O que acontece é que, embora não sejamos racionais o tempo todo, e de fato não sejamos racionais, somos “bem racionaizinhos”, somos um bom tanto racionais. Não estamos adaptados a compreender a taxa de juros exponencial, ou as flutuações na bolsa de valores, mas somos bastante bons em combinar um encontro entre quatro pessoas numa sorveteria a mais de cem metros de distância do local combinado, com três dias de antecedência, e de fato estar presentes no local e hora marcada. Não só isso, mas podemos fazer uma tarefa que invejaria qualquer físico venusiano que passasse por aqui, podemos *predizer* com precisão relativa que quatro primatas inteligentes com polegares opositores de tal e tal jeito estarão, daqui a três dias, naquele espaço particular. Se isso fosse um problema computacional no nível físico, seria intratável, e no entanto nossos computadores internos são plenamente capazes de resolver o problema. Deve haver um atalho.

Já falamos sobre esse atalho, a estratégia intencional, e existem outros atalhos dos quais nos utilizamos, além desse. O único psicólogo (até hoje) a receber o prêmio Nobel, Daniel Kahneman, estudou principalmente alguns tipos de atalhos que nos dão alguma precisão intuitiva, mas não são tão precisos quanto um critério estrito de racionalidade nos faria pensar (ver Gilovich, Griffin & Kahneman, 2002). Ao estimar números, somos influenciados por um número anterior que nos foi apresentado. Ao estimar médias (por exemplo de uso de cocaína por atores de Hollywood), utilizamos um recurso de memória no qual procuramos um ator prototípico e verificamos se ele usa cocaína. Há uma série de estratégias como essas, que são atalhos rápidos para uma quase racionalidade funcional. Ou, como diria Dennett, *suficientista*: não somos acertadores, somos *suficientizadores*. A evolução não é *preciosista*.

Ainda no artigo, Dennett sugere três campos de desenvolvimento para a pesquisa em psicologia. De início, a tarefa da psicologia seria sistematizar essa teoria popular, para ver como ela funciona em seu melhor estado, o que ela tem de melhor a oferecer.

Posteriormente, estaríamos munidos de conceitos como *crenças*, *intenções* e

desejos. Como vimos, Dennett considera esse tipo de estado psicológico como *abstracta*, estados para facilitação de cálculo, que ajudam o poder de previsão, e funcionam como um centro de gravidade, tendo o mesmo grau de realidade que um. *Abstracta* é metade de uma dicotomia, seu antônimo é *illata*. Os *illata* são entidades que têm poder causal efetivo, e que portanto têm um grau maior de realidade que os *abstracta*. Enquanto o centro de gravidade da terra é melhor visto como um *abstracta*, a força eletromagnética que impede que a terra colapse sobre si mesma é melhor vista como um *illata*. Os disparos neuronais, modificações químicas nas sinapses, e transmissões de energia elétrica pelo sistema nervoso são melhor vistos como *illata*, já que possuem essa efetividade causal bastante clara. Um dos críticos dessa diferenciação feita por Dennett o pressionou a explicar mais claramente a diferença entre os *illata* e os *abstracta*, e Dennett respondeu que a principal diferença é que é possível adquirir nova informação empírica sobre um *illata*, mas não sobre um *abstracta*. Vejamos. É possível que descubramos, com o tempo, alguma nova propriedade da força eletromagnética, com efeito, descobrimos há não muitas décadas que ela pode ser unificada com uma das forças atômicas, a chamada força fraca. O novo nome dessa força tornou-se então eletrofraca, e nosso entendimento da realidade mudou, porque nosso entendimento do eletromagnetismo mudou. Da mesma maneira, poderíamos obter novas informações sobre o que ativa os neurônios; uma das maneiras pela qual isso poderia acontecer é se desenvolvêssemos a tecnologia para uma análise funcional de 1,5mm, e não de 2mm como temos agora, e encontrasse novos padrões de ativação.

Se a sugestão de Dennett estiver correta, então o mesmo não poderia ser feito com centros de gravidade. O centro de gravidade da Terra é o ponto no espaço a partir do qual suporemos que a força da gravidade é exercida. Isso se quisermos fazer uma aproximação da força gravitacional terrestre sobre um objeto. O ponto em si é apenas uma abstração geométrica, e a força é exercida pela matéria da Terra inteira. Imaginemos então que se faça alguma nova descoberta sobre a força de gravidade, e que ela seja maior do que se esperava. Isso não modificaria o centro de gravidade, ele continuaria sendo, por sua natureza, uma suposição útil para calcular a gravidade, mas ele não passaria a exercer alguma nova força, ele não se moveria, não derreteria seus arredores etc. Sua natureza continuaria a mesma. É isso que o torna um *abstracta*. Quando Dennett se coloca como um tipo de realista a respeito dos estados mentais, está se referindo a um realismo-abstracta sobre eles. Dennett crê que *crenças, desejos, medos, intenções* são maneiras de aproximarmos o que está de fato acontecendo, no fundo no

fundo, no cérebro e seus caminhos tortuosos. São aproximações. Se formos procurar apenas os *illata*, encontraremos neurônios, sinapses, colunas corticais, encontraremos até símbolos, que tem quase-funções especiais, e que ainda assim exercem seus efeitos por sua interação física, e não por seu significado. Símbolos são a ponte entre a sintaxe e a semântica, entre a física e o vocabulário mentalista. Símbolos são o que permitem a transformação de carne em conceitos, de cérebro em ideias. Eles são justamente o processo de construção de significado através de isoestruturas, através de *representações* do mundo.

Ação e representação

¿O que é representar o mundo? É reagir a ele de maneira sensível a algumas propriedades, e não a outras. Representamos o mundo na medida em que nossa interação com o mundo é regulada por ele também, não apenas por nós.

Estávamos nos munindo com *crenças, desejos e intenções*, estados psicológicos, que importamos da psicologia popular e refinamos numa teoria preditiva, através do uso da perspectiva intencional. ¿Como continuar o projeto da psicologia? Uma vez armados com esses novos termos, agora técnicos, formaríamos uma Teoria dos Sistemas Intencionais, uma teoria de competências. Nessa teoria, trataríamos as propriedades como *acreditar, desejar* etc. como propriedades do sistema como um todo, e trataríamos o sistema como uma caixa preta. A teoria lidaria com a “produção” de novos pensamentos, crenças, desejos etc. a partir de antigos. Isso através de entendimento das relações entre crenças, desejos, propriedades do ambiente, e as ações do sistema. Como já temos as regras de atribuição, poderíamos atualizar constantemente o estado do sistema, e isso geraria uma impressão de que há uma descrição naturalista dos processos internos do sistema. Essa ilusão seria análoga a de um estudante de física que confundisse um gráfico de soma de forças com cordas, roldanas e rodas, num desenho de um exercício. Essa seria a ilusão daqueles que acreditam na Teoria da Intencionalidade Originária Humana. Mais genericamente, esse seria o erro de todos aqueles que atribuem realidade, no sentido de *illata*, aos estados mentais.

A segunda parte da nova psicologia estaria, essa sim, preocupada com os estados naturais internos que de fato regulam o sistema, que operam causalmente nele, que regem seu comportamento, e suas mudanças internas. A outra teoria, Psicologia Cognitiva Sub-pessoal, teria esse papel. Essa seria uma teoria de performance, mais

como uma engenharia reversa: o cérebro funciona como uma máquina semântica, e a tarefa seria descobrir quais inputs significam o quê.

O cérebro é um operador de significados, serve para que entendamos os significados, e ajamos de acordo, mas:

como a fisiologia ou o simples senso comum nos mostram, ele é apenas um *motor sintático*; tudo que ele pode fazer é discriminar seus inputs por suas propriedades estruturais, temporais e físicas, e deixar suas atividades inteiramente mecânicas serem governadas por esses traços “sintáticos” de seus inputs. Isso é tudo que cérebros *podem fazer*. Agora, como consegue o cérebro extrair semântica da sintaxe? Como poderia *qualquer* entidade (como poderia um gênio, um anjo, ou Deus) extrair a semântica de um sistema de nada mais que sua sintaxe? Ele não poderia. A sintaxe de um sistema não determina sua semântica. Por qual alquimia, então, o cérebro consegue extrair resultados semanticamente confiáveis de operações sintaticamente dirigidas? Ele não pode fazer uma tarefa impossível, mas poderia ser desenhado para e aproximar da tarefa impossível, para copiar [*mimic*] o comportamento de um objeto impossível (o motor semântico), capitalizando sobre correspondências fortuitas próximas (suficientemente próximas) entre regularidade estruturais – do ambiente e de seus próprios estados internos e operações – e tipos semânticos. (Dennett, 1981b, pp. 53-4)

Infinitudes e generatividades

Digamos que você queira saber se Marvin Minsky já matou alguém, uma maneira de fazer isso sintaticamente é pegar toda sua correspondência, e fazer um computador executar o seguinte algoritmo: caso encontre a palavra “matei” ou “assassinei” ou “estranglei”..., até “morreu”, emitir um output “SIM” em letras garrafais numa folha impressa. Esse não é um bom mecanismo, por uma série de razões, mas podemos pensar nele como uma variedade extremamente simples do que estamos procurando aqui. Poderíamos sofisticar o mecanismo, utilizando um buscador, incluindo também os vídeos em que ele aparece e um software que escreve as falas etc. No entanto, nunca poderíamos fazer a mesma coisa para todos os aspectos do ambiente no qual estamos interessados. A estratégia da evolução não pode ter sido essa para nos fazer capazes de capturar a semântica. Há de haver uma estratégia gerativa, que seja capaz de capitalizar recursos finitos e discretos, e a partir disso gerar infinitamente mais possibilidades de disposições, reações e formas de ser no ambiente. Só há um caso do qual temos notícia no qual uma estrutura exibiu essas propriedades, é o caso da linguagem. A linguagem tem o poder de expressão praticamente ilimitado a partir de um

número de palavras e regras finito. Algo similar tem de estar ocorrendo em nossos estados mentais internos. Isso nos dá uma ideia genial, e uma confusão de igual proporção. A ideia genial é a de que pode existir algo análogo a uma gramática mental, que nos permite exibir a variedade de comportamentos, maneiras de viver e crenças que exibimos, sem cair em contradição. Pense por exemplo no fato de que você acredita piamente que 132 é maior que 122. Se existisse um só neurônio responsável por essa crença, e você não fosse um total ignorante matemático, isso consumiria todo o cérebro. Afinal, você também acredita que 122 é maior que 121, e que 20.... A ideia ruim que é trazida pela analogia com a linguagem é a de que possa existir uma linguagem do pensamento, com depósitos cheios de palavras, frases e sentenças. Não haveria espaço para tantas sentenças, então não pode ser o caso que no nível de implementação, no nível físico, e no nível simbólico, o tipo de entidades que encontraremos será semanticamente carregado. A infinitude dos significados possíveis, e o fato de que temos, literalmente, infinitas crenças, desejos e intenções, nos garante que quaisquer que sejam as explicações que, ao fim e ao cabo, encerrarão o projeto da psicologia, não existem sentenças armazenadas em nossa cabeça.

A função da Psicologia Cognitiva Subpessoal será então reproduzir, inclusive com as falhas e desvios de racionalidade, os *illata* do cérebro, e entender seu funcionamento estrutural num nível mais fino do que o nível dos significados. Ela irá além da assunção de racionalidade, para o nível de implementação dos mecanismos que geram a inteligência, as emoções, e até o uso de linguagem. Essa tarefa é aquela a qual tem se proposto a Neurociência Cognitiva. Pode ser o caso, ao longo do tempo, que encontraremos de fato correspondentes adequados aos significados exteriores nas engrenagens internas do sistema. Podemos esperar ansiosamente por isso, mas não devemos segurar o fôlego.

A tarefa redutiva final, depois de implementarmos essas duas teorias, seria justamente a de eliminar a Teoria dos Sistemas Intencionais, trocando seus termos por termos fisiológicos, através da Psicologia Cognitiva Subpessoal. *¿Hein?* Não, isso seria um reducionismo cobiçoso da pior espécie, seria justamente o projeto que está falhando.

O que devemos fazer, de fato, é o contrário disso, é mostrar como a um sistema descrito em termos fisiológicos é capaz de se dar uma interpretação como uma realização de um sistema intencional. Agora sim, isso soa (espero que também para você) mais Dennett.

This page intentionally left self-referential.

13. Padrões, consciência, significado, e a realidade dos *abstracta*

Deve estar evidente nesse ponto que a diferença entre *abstracta* e *illata* é de fundamental importância no trabalho de Dennett. Todas as discussões ontológicas para ele estarão ou comparando o grau de realidade de uma entidade ao grau de realidade de outra, ou classificando-as como *illata* ou *abstracta*. Dennett defende que os significados de sentenças e de frases são apenas *abstracta*, e faz o mesmo a respeito do self, ao tratá-lo como um centro de gravidade narrativo, e também com a consciência, ao tratá-la como a história joyceana melhor capaz de recontar os variáveis processos internos que ocorrem em nossas mentes, os inumeráveis rascunhos que se interpolam e se atropelam na construção de nossos comportamentos, dos mais simples aos mais sofisticados. Além de todos esses aspectos que são considerados *abstracta*, Dennett também discute uma outra parte da natureza, colocando questões ontológicas sobre ela, os padrões. No algoritmo passamos rapidamente pelos padrões, sob promessa de que mais tarde eles voltariam à tona (para que voltar atrás no processo algorítmico ganhe mais e mais sentido). Nosso ângulo de ataque, mais uma vez, é a pergunta sobre qual é o grau de realidade de padrões. Duas perguntas serão respondidas conforme o artigo que Dennett entende como seu segundo artigo mais inteligente, “Real patterns” (1991b): ¿O que são padrões? ¿Como saber que um padrão é real?

Por razões de simplicidade, adotei anteriormente o critério de David Deutsch de realidade, dizendo que algo é real se chuta de volta. Podemos agora nos desmamar desse conceito simples, e compreender melhor, de uma perspectiva diferente, o que garante que um padrão seja real. Vejamos alguns comentários sobre a analogia de Dennett para os estados mentais:

O problema com esses supostos paralelismos [...] é que eles são todos, falando estritamente, *falsos*, apesar de não haver dúvida de que são simplificações úteis para vários propósitos. É falso por exemplo que a atração gravitacional entre a Terra e a Lua envolva duas massas pontuais; mas é uma primeira aproximação boa o suficiente para muitos cálculos. Mas não é isso que Dennett realmente quer dizer sobre estados intencionais. Pois ele insiste que adotar a perspectiva intencional e interpretar um agente como agindo sob certas crenças e desejos é discernir um padrão em suas atividades que está genuinamente lá (um padrão que é perdido se adotarmos ao invés disso uma perspectiva científica): Dennett certamente não sustenta que o papel de atribuições intencionais é meramente o de nos dar uma aproximação útil de uma verdade que pode ser mais acuradamente

expressa em termos não intencionais. (Smith, 1988, p. 22).

Vamos comparar agora este trecho com a afirmação igualmente confiante de realismo, vinda de Dretske:

Eu sou um realista a respeito de centros de gravidade [...] A terra obviamente exerce uma atração gravitacional em *todas* as partes da lua – não apenas o centro de gravidade. A força *resultante*, a soma dos vetores, age através de um ponto, mas isso é algo bastante diferente. Deve-se ter clareza sobre o que centros de gravidade sejam *antes* de se decidir sobre ser literal a respeito deles, *antes* de se decidir sobre ser ou não um realista sobre centros de gravidade. (Dretske, 1988, pp. 511-2.)

Há dois caminhos na decisão sobre ser ou não ser (um realista) sobre alguma coisa, o metafísico e o científico. O metafísico está cheio de pontos interessantes, como o baricentro do triângulo de vértices Leitor, Dennett e Diego. Esse é um ponto movimentado e dinâmico, apesar de irrelevante, mas metafisicamente seu status é o mesmo do centro de gravidade da terra. O caminho científico opera em dimensões mais úteis. Um ponto é real para a ciência se é o tipo de coisa interessante, que pode ser usado, que deve ser aprendido, e é em geral útil. Dennett defende justamente que as crenças sejam reais nesse sentido. Mas uma questão permanece, ¿se esses padrões de crença são, afinal, padrões, *do que* são padrões?

¿É real? Pode apostar!

¿Quando é um padrão real? Essa versão genérica da questão talvez ilumine-nos com novos e diferentes insights a respeito de como pensar a sub-questão “¿Quando é um padrão mental real?”

Para pensar essa questão Dennett nos sugere considerar a Fig. 7. Todas as figuras foram geradas a partir de um padrão (o que intuitivamente vemos na Fig. 7), mas em cada um deles foi introduzida uma quantidade de “ruído” [*noise*] ou interferência aleatória. Essa quantidade está definida nas seguintes proporções: A: 25%; B: 10%; C: 25%; D: 1%; E: 33%; F: 50%.

Não é possível saber que F foi gerado pelo mesmo sistema que gerou os demais números, apenas com mais ruído. Isso porque 50% de ruído, quando só se pode variar entre dois estados, é o mesmo que total aleatoriedade. Teremos que confiar que Dennett tenha utilizado o mesmo programa de geração de pixels a partir do padrão para criar o desenho. O padrão de F é indiscernível de um padrão aleatório.

Mas eu proponho que o ar auto-contraditório de ‘padrão indiscernível’ deva ser levada a sério. Podemos até dar um sentido estendido ou metafórico para a ideia de padrões indiscerníveis (ou figuras invisíveis, ou sinfonias silenciosas), mas no cerne um padrão é ‘por definição’ um candidato para *reconhecimento* de padrão. (Dennett, 1991b, p. 32.)

Um padrão é definido pela possibilidade de ser representável em menos informação do que a que nos é apresentada de início. Os padrões de A a E poderiam ser representados a partir do código de barras inicial, com uma descrição pixel-a-pixel apenas da interferência, o mesmo não ocorre com F, seu algoritmo de geração não permite tal manobra, já que 50% de interferência, ou modificação, equivale ao mesmo que se o padrão não estivesse presente de início.

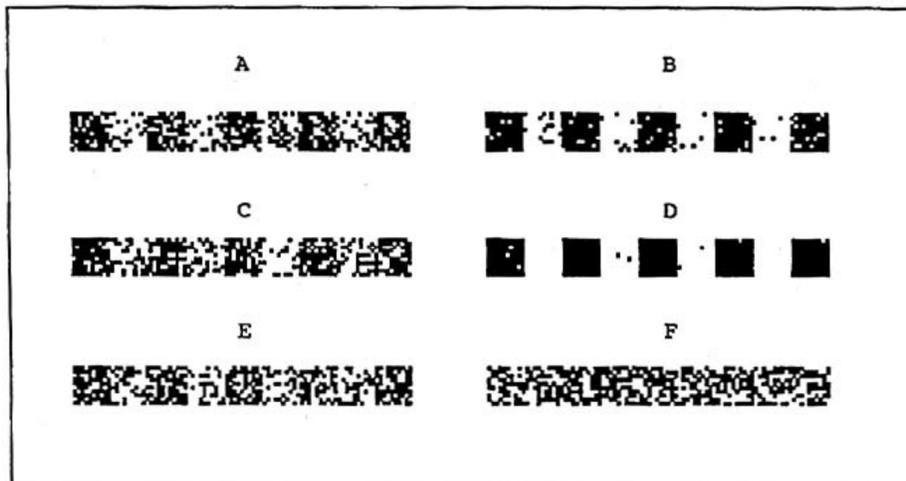


Figura 7: Padrões com ruído.

Poderíamos criar descrições diferentes da mais óbvia a olhos humanos, mas que ainda assim reduzissem o tamanho da descrição em relação a uma descrição extensiva. Por exemplo, se tomarmos o padrão de código de barras, e deslocarmos esse padrão apenas um pixel para o lado, especificando claramente cada pixel que desvia desse novo padrão, ainda estaríamos lidando com uma descrição menos extenuante que a descrição pixel a pixel no caso de D. Isso implica que há mais de um padrão que pode ser encontrado em D, mais de uma maneira de descrever D que reduziria a quantidade de informação necessária para descrever D. Não é apenas em D que isso ocorre, isso só é mais facilmente visualizável a nossos olhos em D. Mas nos outros casos (salvo F)

também há descrições mais curtas e simples que são suficientes para disponibilizar toda a informação que acessamos, sem a necessidade de uma informação pixel a pixel. Qualquer uma dessas descrições encurtadas é um padrão real na informação.

Agora digamos que, ao invés de deslocar todas as colunas uma linha para a direita, mantendo o padrão bastante visível a nossos olhos humanos, nós tivéssemos escolhido algo como “a cada duas linhas mova todos os pixels 3 pixels para a direita, e as linhas intermediárias mover 2 pixels para a esquerda”. Nesse caso não teríamos mais a capacidade de discernir quase nenhum padrão no desenho. Mas nesse caso outras criaturas, que vivessem num universo um pouco diferente do nosso, com outros órgãos sensoriais, e outro sistema cognitivo poderiam perceber o padrão. O padrão estaria na informação o tempo todo. A única coisa é que nós não perceberíamos.

¿Então como fazer para saber se um padrão está de fato em algum lugar? Não estamos providos de todos os sentidos possíveis, de todo o aparato biológico complexo para fazer, por exemplo, detecção de calor com a língua (cobras), de campos elétricos com a pele (peixes) ou ecolocação (morcegos e umas trinta pessoas). No entanto, podemos e devemos crer nesses padrões, isto porque existe alguém na natureza capaz de ver por trás das cortinas, e entender que roldanas estão sustentando aquilo que nos parece voar, mesmo que *não possamos* ver as roldanas (não com nossos sentidos brutos ao menos). Óculos, microscópios, detectores de terremotos etc. são maneiras de ampliar o escopo de nossos sentidos e ver padrões que não podíamos ver antes.

A questão que fica é: ¿e quando não temos a aparelhagem de detecção necessária, como podemos avaliar se um padrão está ou não presente? Digamos por exemplo que duas espécies diferentes consigam fazer uma detecção de um padrão no ambiente que nos é totalmente inacessível. O que para nós parecem apenas folhas completamente iguais é por elas distinguidas cuidadosamente. As que elas selecionam não são venenosas, e as que evitam, em geral são venenosas. Para nós, a aparência é igual, e esse tipo de veneno não nos afeta, então até onde podemos identificar, elas estão escolhendo aleatoriamente. Um cientista no entanto descobriu que não podia dar qualquer folha para aquela espécie de, digamos, inseto. Boa parte das folhas matavam os animais. Mais tarde, descobriu-se que um aracnídeo que vive do outro lado do rio também se alimenta dessa mesma folha, mas sua taxa de acerto de plantas não venenosas é um pouco pior, fazendo com que muitos morram após comer parte das folhas envenenadas. Ambas essas espécies encontraram padrões.

Digamos agora que apresentemos uma sequência de folhas para ambas as

espécies e descobrimos que todos os espécimes de inseto sempre vão em direção às mesmas folhas. Digamos a 1, 6, 7 e 9, dentre as 10 primeiras. Os aracnídeos escolhem a 1, 3, 7 e 9... Ao obrigá-los a comer todas as folhas, a 2 e a 8 matam ambas as espécies. Ambas então aprenderam a desviar, ambas encontraram um padrão na natureza que, se seguido cuidadosamente poderia preveni-las da morte por envenenamento. ¿Qual delas encontrou o padrão certo? Pode não haver uma boa resposta para essa questão. Digamos que uma delas encontrou um padrão P com algum $z\%$ de interferência, a outra encontrou o padrão R, com $y\%$ de interferência. O que ocorre é que ambas estão apostando em seus padrões para determinar o que fazer a respeito da próxima folha de uma série. Ou seja, ambos estão *antecipando* o valor de sobrevivência da folha. Se ambas as espécies estão, em geral, sobrevivendo, então ambos os padrões são reais. Isso vale tanto se a taxa de sobrevivência for exatamente a mesma, quanto se, como foi o caso descrito, os insetos sobrevivam mais frequentemente que os aracnídeos. O resumo de toda a ideia é pensar isso da perspectiva dos genes. Chamemos de *rico* um gene muito frequente. Então, se alguma suposição de padrão está deixando um gene *rico*, o padrão existe na natureza. Se ele está deixando uma pessoa rica, ele existe na natureza. Se *alguém* fica rico, o que quer que seja, ele é real.

Na figura, A e C têm o mesmo tanto de interferência em relação ao padrão fixo de um código de barras. Ambos são a manifestação do padrão “Código de barras com 25% de aleatoriedade”. O que muda é qual foi a manifestação aleatória que interferiu com o padrão. Há uma desvantagem no uso da palavra ‘interferência’ para traduzir ‘noise’ do inglês. ‘Noise’ significa ruído, ou barulho, e aqueles mais afeitos a computação devem estar acostumados a utilizar ‘ruído’ como descrição da informação que não é parte de um padrão que chega a um sistema. A vantagem de ‘ruído’ sobre ‘interferência’ é que ‘interferência’ passa a ideia de que, de início, existe um padrão limpo, perfeito, cristalino. Durante o processo de transmissão de informação esse padrão é cortado por algo de fora que age causalmente sobre aquela estrutura, modificando suas propriedades, e tornando-a gradativamente mais aleatória e menos padronizada, até o ponto em que ela chega no receptor. Mas não é disso que se fala aqui. A informação bruta que está chegando até nós já é a informação que parece, a nossos olhos primatas, estar *alterada*, mas *não está alterada*, ou ao menos não sabemos disso ainda. Tudo o que temos disponível ao reconhecer um novo padrão é um *provisão* hipotético de que aquele padrão está lá, e uma estimativa de quanto ruído ele possui. Não há nada na natureza que impeça que o padrão já venha em sua forma *aparentemente* alterada, que a natureza já

seja assim. De nossa perspectiva epistêmica, ao detectar um novo padrão (digamos a radiação de fundo do universo), só podemos ver que existe um *algo* ali. Isso não significa que houve um *superalgo* cristalino em algum lugar muito longe, e que desde então aquilo veio se degenerando até chegar até nós.

Ser um padrão real então é ser de tal natureza que, se apostássemos nele, ele renderia lucro..., algo mais eficiente do que um mapa pixel a pixel, bit a bit. Às vezes, por considerações pragmáticas, de nossos sentidos, de nossos computadores, ou até mesmo da mãe natureza, em toda a sua ignorância e cegueira, somos levados a fazer aproximações idealizadas dos padrões, fingindo que de fato aquele padrão cristalino e ideal está lá no fundo, pois essas heurísticas tornam algum processo mais eficiente. Mesmo que pudéssemos visualizar objetos em longa distância com mais precisão, é possível que não valha a pena o dispêndio biológico, como vale para uma águia por exemplo. Os padrões da natureza estão disponíveis por vezes para o mais minucioso exame, mas isso não significa que nosso maquinário esteja à altura de detectá-los, e muito menos que seria eficiente fazê-lo. O caso dos insetos *versus* aracnídeos é interessante na medida em que vemos dois padrões diferentes ocorrendo na mesma série de informação, e devemos confiar na realidade de ambos, *mesmo antes* de nos fazermos a pergunta sobre a possibilidade de um padrão mais profundo, que, se explicado, explicaria os dois.

Nosso aparato sensorial nos dota com algumas maneiras de detectar padrões, que nos permitem caminhar pela rua, desviar de bolas de beisebol, e jogar xadrez. Essas compressões de informação que estamos fazendo seriam completamente intratáveis de uma perspectiva física, minuciosa. Simplesmente não podemos construir um computador que calcule a trajetória de uma bola de beisebol apenas a partir dos fótons, sua trajetória até a retina etc. A intratabilidade é o que nos leva a usar a física newtoniana, a perspectiva intencional, a tirar médias, a usar a perspectiva do design... Somos, ao menos em parte, dispositivos sólidos de compactação, facilitação de cálculo e redução de complexidade. Ao menos em parte. E sempre que é possível fazer isso, sempre que há algo ali que pode ser reduzido, facilitado, compactado, e que ainda assim gere resultados positivos, pode estar certo, existe um *padrão real* aí.

Apostando nos *abstracta*

Falando sobre padrões, consciência e significado, Dennett deixa bastante claro seu interesse e gosto pelos *abstracta*. É patente que esse tipo de entidade para ele

permeia o mundo humano em diversos aspectos, e que controla nossas ações de diversas maneiras, muitas vezes com mais precisão e acurácia do que o controle exercido pelos neurônios, elétrons e demais entidades micro-cósmicas. Julga todas essas entidades como sendo desse tipo mais abstrato, que exercem sua influência causal justamente por serem interpretados como tais, e não porque haja algo ali anterior e imanente que faz um papel equivalente ao de uma onda eletromagnética, só que em um nível superior. A organização do mundo físico é diferente, em escala microcós mica e em escala macrocós mica ou intermediária, e isso, apesar de um problema para alguns fisicalistas, alguns filósofos que dizem que tudo é físico e tem de funcionar tal qual as equações da mecânica quântica prevejam, não é um problema para os *abstractistas*. Dennett não precisa se afastar do monismo, abraçar o cartesianismo, ou abandonar a crença na existência de significado, e pode, ao mesmo tempo, acalentar-se na esperança de muitos reducionistas de que o mundo é um sistema físico, nada mais.

Compatibilizar o monismo com as coisas de tamanho médio e com as coisas menos concretas é um dos grandes trabalhos da filosofia pós-cartesiana. Diversas soluções foram propostas, por grandes nomes como Russell, Quine, Davidson, Dennett, Sellars, Wittgenstein, Chalmers, Searle, Fodor, Block, Levine, entre outros... Algumas dessas soluções são mais próximas dessa que viemos examinando até aqui, a de Davidson em particular, outras mais distantes, como a de Chalmers, a de Russell na década de 1910... Como vimos, não é tudo que Dennett diz serem *abstracta*, algumas coisas, como os qualia, ele simplesmente nega que existam. A jornalista e cientista Susan Blackmore (2006, p. 6) organizou um livro de entrevistas com diversos pensadores a respeito da consciência, no qual Dennett é retratado como capitão do time A, e Chalmers como capitão do time B. A principal diferença entre ambos os times é aceitar ou não a realidade ontológica (ôntica) dos qualia.

Tem qualia que o Dennett não compra, para todas as outras, existem os *abstracta*.

Essa moral não é tão distante dos pontos de vista a respeito das grandes questões filosófico-semânticas que Dennett se dispôs a enfrentar ao longo da carreira. Se Dennett fosse ficar famoso por apenas dois aspectos de sua obra, seria sua incessante defesa da perspectiva intencional, e sua estratégia de salvar sistematicamente o aparentemente controverso utilizando-se das noções *illata* e *abstracta*.

14. IA, ciência cognitiva, etologia e neurociências: ¿um cientista ou um filósofo?

Richard Dawkins pontua, em uma apresentação, que ele (um biólogo) sempre se deleita em apresentações de cientistas, no entanto, uma pergunta para ele sempre fica em pauta: “What is the point of philosophers?”⁶¹ Ele diz que sempre que essa pergunta lhe atormenta, lembra-se da figura de Dennett, que toma como seu irmão mais velho intelectual, e se assegura de que, afinal, existe algum propósito na existência dos filósofos. Muitos outros cientistas compartilham essa opinião, a opinião de que a filosofia em geral de nada serve. Em parte essa opinião tão presente entre cientistas se deve a uma visão equivocada da função do conhecimento, que muitas vezes leva os cientistas a considerar como conhecimento apenas a capacidade de prever o futuro, fazer previsões testáveis, e demais cânones que são ensinados nos primeiros anos de graduação da maioria dos cursos de ciência. Independente de suas razões, vale notar que mesmo entre aqueles cientistas que consideram os filósofos como a escória do mundo, os que têm contato com o trabalho de Dennett acabam, de duas uma: ou abrem uma exceção e dizem que existe um filósofo que de fato está trabalhando, ou simplesmente desclassificam Dennett como um filósofo e o chamam de cientista. Mesmo alguns colegas filósofos dizem que o que ele faz hoje (na década de 1990) não é mais filosofia mas sim ciência.

É difícil não simpatizar com aqueles que querem classificar Dennett como um cientista. Suas listas de referência ao final dos livros são majoritariamente científicas, ele escreveu um livro sobre biologia defendendo pontos de vista genocêntricos, participou com um grupo do MIT na criação de um robô que aprende com o ambiente. Tem diversas publicações em revistas de ciência, inclusive com títulos como “Out of the armchair and into the field”⁶². Mesmo em parte de seus artigos mais antigos há constantes referências a temas de ciência.

Não é apenas Dennett que flutua entre essas categorias: a taxonomia da ciência tem se expandido bastante, com o surgimento de ciência cognitiva, neurociência cognitiva, economia comportamental, genética comportamental, genética de populações, psicologia de populações, biogerontologia, biofísica e mais uma profusão de outras áreas interdisciplinares da ciência que vêm surgindo nas últimas décadas para acomodar

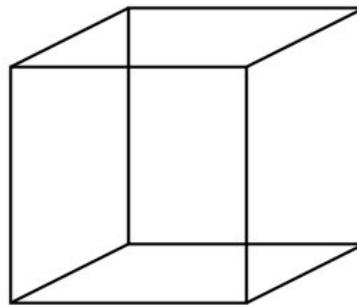
⁶¹ Qual é o propósito dos filósofos?

⁶² “Saindo da poltrona de filósofo para ir a campo”, publicado em Dennett (1998).

nosso conhecimento exponencialmente maior... Procurar uma solução que encaixe um ou outro pesquisador numa ou outra ciência não teria mais valor do que jogar uma moeda. Assim como não existe um ponto claro entre o Homo sapiens e o Homo rhodesiensis que os divida, e há casos intermediários, assim como não há um ponto claro no qual está garoando ou chovendo, da mesma maneira não há uma questão de fato na pergunta: “¿Mas ele é um filósofo ou um cientista?” Podemos salientar as *questões de fato* que *podem ser respondidas*. Por exemplo, Dennett participou efetivamente de poucos experimentos, famosamente de uma pesquisa em etologia na savana africana, mas visitou diversos laboratórios de neurociência e robótica. Em seu escritório, na Universidade de Tufts, há exemplares de muitos anos da revista *Animal Cognition*. Numa conferência em que estive, um cientista cognitivo resumiu muito bem: Dennett é um grande conhecedor da ciência, mas não é propriamente um experimentador, ele não faz filosofia experimental, mas filosofia esclarecida pela ciência.

¿Quais são as ciências das quais Dennett mais se serve? Interessantemente, a parte das ciências que mais lhe interessou acabou por ser uma parte mais 'filosófica' da ciência. Tomemos por exemplo a reconcepção feita por Richard Dawkins da teoria genocêntrica da evolução. Segundo Dawkins, em seu famoso livro *O gene egoísta* (1976), o gene é a unidade correta para se pensar a evolução, em oposição ao nível do indivíduo (em suma, às vezes vale a pena se matar por dois irmãos ou oito primos). Esse livro, no entanto, foi tomado como um livro majoritariamente filosófico até pelo próprio Dawkins. Para ele, o papel do livro na biologia seria o mesmo papel cumprido por nosso cérebro ao olhar para um cubo de Necker e variar o lado que aparentemente está mais próximo de nós.

Figura 8: Cubo de Necker.



Dawkins estava propondo uma mudança na forma de ver muito mais do que uma nova teoria a ser testada em laboratório, até porque a teoria de seleção genética, considerando graus de parentesco, já existia desde um pouco antes, tendo sido formulada por outro biólogo, Bill Hamilton. No caso do genocentrismo, Dawkins já

havia feito o trabalho “sujo” para Dennett, por assim dizer, que apenas se encarregou de refinar a sua teoria e derivar algumas de suas consequências mais filosóficas em *A perigosa ideia de Darwin*. No caso dos memes, a situação não foi tão simples. Embora o mesmo Dawkins tenha proposto a existência de memes, no mesmo livro, Dennett foi um dos principais e mais resistentes defensores da memética. A perspectiva memética toma poder central em sua obra ao longo dos anos 1990, com artigos e capítulos de livros cada vez mais repletos do meme “meme”. A memética foi tomada como ciência apenas por uma parte dos pesquisadores em ciências sociais, e mesmo a comunidade genocêntrica da biologia ainda tem reservas a respeito do poder explicativo e da veracidade dessa teoria. Se a memética não for verdadeira, a história se encarregará de erradicá-la como ciência. Se ela for verdadeira, a história dos memes não necessariamente verá seus dias de glória, afinal, nunca foi parte da teoria memética que o meme “meme” é um bom meme. Em seu livro sobre a religião, *Quebrando o encanto*, Dennett (2006a) se dispõe a expor uma teoria de *aspecto* científico, desafiando aqueles que dizem que não se poderia fazer uma ciência da religião, ou estudá-la seriamente.

Genes e memes mostram que a ciência de Dennett realmente opera na interface entre o claramente científico e o abstrato. Ele mesmo no entanto tem grande comprometimento com criar hipóteses testáveis, e seus livros estão carregados de previsões e testes de suas teorias, a partir de *A consciência explicada*, de 1991.

As teorias científicas mais controversas que Dennett aceita são o genocentrismo e o memecentrismo, que juntos formam o ultra-darwinismo. Além dessas duas, há diversas teorias que originalmente pertencem ao campo da filosofia que ele deseja transformar em ciência, e portanto que têm a *versão dennettiana* como uma hipótese testável. Aqui entram suas teorias sobre a natureza da consciência, ao pensá-la como reverberação, *fantasy echo*; também entram suas hipóteses sobre como o tempo é experienciado, e mesmo seu computacionalismo brando a respeito dos estados mentais.

No início do texto, sugeri que o papel que Dennett tenta cumprir é o de um engenheiro de obras que está tentando com um túnel conectar o físico a alguns objetos teimosos, que insistem em não se adequar facilmente a ele. O reducionismo de Dennett é o de um engenheiro, tentando bloquear infiltrações no caminho entre o físico e o mental, entre o físico e o intencional, entre o físico e o design. Como um bom engenheiro, Dennett serviu-se de ciência, e bastante ciência, para atingir parte de seus propósitos. Agora, no próximo capítulo, faremos o papel de um certificador, averiguando o caminho traçado até aqui, e garantindo que as peças se encaixem umas nas outras.

15. Uma brevíssima história de quase tudo

No início não era o verbo. Na verdade mesmo não houve um início, mas isso se lê em livros de Stephen Hawking e está fora de nosso escopo. Eventualmente havia um monte de energia se expandindo, resfriando, e boa parte dela transformou-se em matéria. Parte dessa matéria por sua vez foi se aglomerando, e formando estrelas, mais tarde também planetas, com rios, atmosferas, raios, vulcões entre outros fenômenos planetários interessantes a diretores de cinema. Cerca de 3 bilhões e meio de anos atrás surgiram os primeiros replicadores. Esses replicadores não tinham nada de diferente da matéria inorgânica, salvo o fato de que estavam num meio ambiente no qual algumas de suas propriedades (geométricas, químicas, físicas) tendiam a criar cópias dos replicadores. Por algum tempo isso não fez diferença alguma. Até que um belo dia, ou melhor, um dia tempestivo e nebuloso, numa atmosfera cheia de amônia, um péssimo dia: dois desses replicadores se diferenciaram, por acaso. Mais algum tempo passou com as duas “espécies” de replicadores, até que, em outro péssimo dia, dois deles concorreram para o uso de um recurso, de tal maneira que apenas um deles pôde acessá-lo, e o outro morreu. Tinha início a seleção natural. Esses replicadores foram se modificando aleatoriamente e sendo selecionados até que surgiram formas interessantes de vida, como as bactérias. Bastante tempo depois surgiram células maiores, as eucariontes, e até pluricelulares. Durante essa época, estamos vendo o surgimento das criaturas darwinianas, aquelas que aprendem com a morte. Um processo ineficiente e quase ridículo. Embora permaneçam a grande maioria, essas criaturas começam a conviver com criaturas skinnerianas, conseguindo se afastar de um ou outro problema, e se aproximar de uma ou outra solução por meio de instintos, mecanismos de comportamento pré-programados geneticamente. Comportam-se aleatoriamente de início (dentro de um espaço de projeto) e possuem processos de reforçamento que os levam a manter a estratégia de maior sucesso. O tempo continua passando, muito, muito tempo se passa.

Surgem alguns mecanismos mais complexos de controle de comportamento, num mundo ainda dominado por uma massa de criaturas darwinianas e skinnerianas, com sensores mais sofisticados, capazes de detectar fótons, e portanto de obter informação sobre eventos bastante mais distantes que suas antecessoras experimentais. Além disso, eles são capazes de computar, dentre seus movimentos possíveis, aqueles

que têm maior chance de dar certo *antes* de tentar.

Aqui já temos graus significativos de intencionalidade operando na natureza, afinal, ¿como *prever* e *calcular* a relação do ambiente se não tendo, em algum nível, uma representação interna desse ambiente, um isomorfismo entre estruturas do processo de previsão e cálculo, e estruturas da natureza que está sendo apreendida? ¿E como fazer essas isoestruturas? Bem, isso é uma tarefa para a mãe natureza cega, surda, desprovida de intenções etc. E a maneira que ela encontrou foi aproximando padrões reais na natureza a padrões internos do sistema de comportamento de um organismo. ¿Isso significa que o interior físico de um organismo tem de ter regras lógicas representadas fisicamente? Não, apenas precisa reagir de acordo com algumas regras (por exemplo da lógica e da aritmética), o que não significa possuir internamente uma representação *explícita* desses mecanismos quase-regulatórios. Estas criaturas popperianas só precisam acertar o suficiente em suas previsões, não tendo a necessidade matemática de *escrever* em si mesmas o que estão fazendo. É aqui, nessa grande explosão gerada pelas criaturas popperianas, que começam a surgir formas de inteligência mais sofisticadas, que ultrapassam a metodicidade da mosca *Sphex*, para atingir a flexibilidade comportamental de um camundongo. O mundo se tornara um lugar muito mais interessante. Com organismos portadores de grandes genomas, tornava-se possível o surgimento de sociedades, já que não engolir seus primos e irmãos passa a ser uma estratégia evolutivamente boa quando compartilhamos muitos genes com eles. Desde o surgimento dos pluricelulares podemos falar em sociedades, sociedades de células idênticas, lutando juntas pelo ideal comum (do qual nenhuma delas faz a menor ideia, evidentemente), o de colocar aquela célula de linha germinativa num bom ninho cheio de recursos e energia para que ela repita a mesma operação. Mas falamos agora em sociedades mais complexas e intrincadas, sociedades de animais, que são sociedades de células, que são sociedades de replicadores. Estamos falando de sociedades de terceiro nível. Organizações hierárquicas fascinantes começam a surgir, e estratégias curiosas de reprodução também. O sexo, também conhecido como a arte de sacrificar metade dos seus genes em troca de diversidade contra doenças, surge também nesse período intermediário. Muita coisa ocorre enquanto surgem os pequenos Poppers da natureza. Triunfal, a evolução continua a toque de caixa, gerando sistemas imunológicos complexos, comportamentos de caça, comportamentos estereotipados, mutualismo, inquilinismo... O amor – *ah, l'amour!* Esse período é o da cavalcada das Valquírias, em que os mares, terra e ares entram em ressonância com os padrões

informativos que chamamos de vida e explodem num milagre informativo que deixaria a internet com o rabo entre as pernas encolhida num canto. A única diferença é que essa cavalgada durou dois bilhões de anos, o que evidencia a diferença de velocidade entre a evolução genética e a evolução memética. Mas estamos nos antecipando, nesse momento, ainda falamos de um mundo sem memes, sem palavras, sem linguagem.

Os guindastes da intencionalidade

Pouco a pouco o mundo vai construindo uma série de intencionalidades, primeiro, de maneira vaga e imprecisa, com tantas vilosidades e erros nos processos semi-intencionais que é só olhando retrospectivamente que conseguiríamos perceber que é dali que surgiria o que hoje entendemos como intencionalidade. Após as semi-quase-hemi-intencionalidades, surgem aquelas que valeria chamar de fato de intencionalidades. Estas são de primeira ordem, isto é, A é a respeito de B. Depois até intencionalidades de segunda ordem começam a surgir, A é a respeito de B que é a respeito de C. Isso ocorre por exemplo no pensamento avançado social de algumas espécies de primatas, que são capazes de disfarçar um comportamento para um companheiro e mostrar para outro. Ora, ele só poderia fazer isso se estivesse com algo na cabeça, a respeito de algo na cabeça do outro, que por sua vez é a respeito do comportamento que ele está fazendo. Essa brincadeira de construir intencionalidade sobre intencionalidade poderia continuar indeterminadamente, mas não o faz. Com efeito, nós mesmos não conseguimos ultrapassar em muito 5 ou 6 níveis de intencionalidade, Dennett (1987, p. 243) brinca: “eu suspeito que você se pergunta se eu percebo como é difícil, para você, ter certeza de que você compreende se eu pretendo dizer que você reconhece que eu posso acreditar que você quer que eu explique que a maioria de nós só é capaz de lidar com cerca de cinco ou seis ordens [de intencionalidade], na melhor das hipóteses.” Muita intencionalidade é ou intratável computacionalmente, ou simplesmente o tipo de coisa que não compensou para nenhum animal detectar e processar, durante a interessante cavalgada da seleção natural das cognições.

Mas ainda estávamos em sistemas simples de dois níveis de intencionalidade, e éramos ainda criaturas popperianas. Um afunilamento violento em um grupo de hominídeos começa a tomar força, e surgem grupos diferentes, espécies diferentes que

têm uma nova propriedade, a cultura, e a curiosidade que vem com ela. Não falo aqui da cultura no sentido de cantos endêmicos a uma população, já que essas são compartilhadas pelas baleias. Não falo de uso de ferramentas apenas por alguns grupos, já que isso ocorre em chimpanzés, orangotangos, macacos capuchinhos e, quem diria, até em moluscos! Falo aqui de uma cultura mais claramente insignada nos afazeres e nas mentes de um povo, de uma cultura que possui a curiosidade necessária para florescer e a flexibilidade necessária para acessar cada vez mais e mais campos do conhecimento. Uma cultura que opera com proto-linguagens e linguagens. Ainda não há telefones celulares, pois aquilo a que me refiro nós compartilhamos com os neandertais, com os homens de Flores, e com os escultores das pedras de Levallois, potencialmente nossos ancestrais. Evidente que, sendo a cultura mais um dos produtos da seleção natural, somente uma pessoa que ainda está presa à ideia de que tudo tem de ser tudo ou nada se preocuparia em traçar uma linha exata entre o que é ou não essa cultura. Nós, que já ultrapassamos essa fase do conhecimento, podemos simplesmente prosseguir com essa definição intermediária de cultura, aquela que compartilhamos com os neandertais e com os soviéticos, mas não com os chimpanzés, baleias e polvos. Nesse momento surgem as criaturas gregorianas, criaturas que têm capacidade para melhorar sua própria adaptabilidade e velocidade de solução de problemas quando apresentados a designs anteriores criados por outros indivíduos para fazer algo. Nenhum de nós inventou o machado, mas vê-lo auxilia em muito o tempo que demoraríamos para inventar um espontaneamente. Elas podem aumentar sua velocidade de encontro com vizinhanças boas, no espaço de design dos comportamentos possíveis. Copiar tecnologia de informação não apenas deixa-as mais inteligentes, mas acelera o passo no qual ficam mais inteligentes.

As gregorianas trazem com elas o germe de todo o edifício da cultura humana, seja ela a civilização ocidental ou as sociedades indígenas, os feudos medievais ou os ciganos, os japoneses ou os coreanos. Mas mais interessante para nós que as diferenças de proliferações culturais entre esses grupos é algo que surge em todos eles, que não varia muito entre eles, que é o surgimento de um *artefato* que só poderia vir a ser num mundo em que a colaboração e a intencionalidade fossem senhores da vida social, da convivência em conjunto. Um mundo organizado em torno da cultura e fortemente grupal, o artefato que surge nesse mundo é a *linguagem*. Até há pouco, a linguagem possuía a mais forte forma de intencionalidade da qual tínhamos notícia, Wittgenstein, por exemplo, dedicou alguns anos de sua carreira a mostrar ao mundo filosófico todo o

isomorfismo e simbolismo carregado nas palavras, frases e sentenças que compõe nossa vida cotidiana. O jogo da vida, para o velho Wittgenstein, é predominantemente composto de jogos de linguagem, e estes por sua vez são determinados por nossos comportamentos isoestruturais ao que dizemos, e nossas falas isoestruturais a como nos comportamos. Digo que até há pouco a linguagem era a mais forte forma de intencionalidade porque agora ela tem um concorrente, e o júri ainda está de pé para ver qual dos isomorfismos é mais adequado, ou mais adequador entre dois sistemas (um externo e um interno). O concorrente são os *estados mentais*.

Quando há concorrência, no mercado, podem ocorrer diversas coisas. Uma empresa pode fagocitar a outra, e esse é o projeto de Fodor, que sugere que na realidade há uma *linguagem* mental, e portanto os estados mentais são apenas um subgrupo da já mais ou menos conhecida estrutura das nossas linguagens. Duas companhias podem partir para nichos diferentes, caso no qual teríamos de ter toda uma nova ciência dos estados mentais. Ou algum nível intermediário de sobreposição poderia acontecer entre as duas, o que em nosso caso significaria que em parte podemos usar analogias linguísticas para entender os estados mentais, e vice-versa. Mas haverá também distinções importantes que levarão às diferenças que serão manifestas nas psicologias e linguísticas que se seguirem. O cenário mais provável é esse último, no qual podemos exaptar⁶³ algumas, mas não todas as analogias entre um campo e outro.

O mundo está populado. Bactérias, sapos, estados intencionais, estados mentais, propriedades linguísticas, culturas, festas, cerimônias. Todas essas entidades, depois de alguns bilhões de anos surgiram e tomaram algum espaço no nicho físico, biológico, ou memético da natureza que nos cerca. É aqui que surgem, *ipso facto*, as pessoas. Pessoa no sentido lato se refere de maneira vaga a um corpo passeando pelo tempo, suas ideias, seus afazeres, e seus estados relacionais, como estar casado, por exemplo. Dennett é radical dissidente em sua teoria de ordenação do surgimento das pessoas, da linguagem e da consciência. Enquanto quase todos colocam a consciência, depois as pessoas e por fim a linguagem, ele inverterá completamente a ordem. Pessoas seriam quase essencialmente compostas de palavras, de um discurso (gestual e falado) a partir do qual podemos gerar um centro de gravidade narrativa. Sem linguagem, fica mais difícil organizar uma série de acontecimentos sob o nome “Minsky” e dizer que aquele sujeito

⁶³ *Exaptação* é uma adaptação biológica que não evoluiu inicialmente dirigida por pressões seletivas relacionadas à sua função atual, mas a outras funções, até que eventualmente chegou a um estado em que veio a ser utilizada para a nova função.

constitui um *self*. Não é apenas o fato de que o discurso é várias vezes a respeito de si (a palavra mais usada na maioria das línguas é a palavra para *eu*) e *portanto* organiza a centralidade da pessoa. Mais do que isso, o raciocínio auxiliado pelo uso de palavras faz muito melhor a função interna de organizar um sujeito; ele cria através da própria gramaticalidade do sujeito algo ao que aquele sujeito se anexa, algo que vai se tornando ele ao longo do tempo. Estamos o tempo todo “Making sense of ourselves” (Dennett, 1981c) (fazendo sentido de nós mesmos).

Uma das formas de demonstrar isso empiricamente são os estudos que mostram *choice supportive bias*, ou um bias a favor das escolhas feitas. O teste funciona da seguinte forma: A diversas pessoas normais médias (alunos de graduação de Harvard!) são dispostas fotos para que escolham a que mais gostam (Gilbert, 2012). Algumas semanas depois as pessoas são trazidas de volta, já tendo esquecido o que haviam escolhido, em geral. De volta à sala de escolha, os experimentadores mentem para elas, dizendo que escolheram uma outra foto; a partir disso, elas passam a dar de uma série de motivos para sua escolha (que na verdade não foi sua escolha original), e concordam que aquela é mais bonita dentre as fotos. Ou seja, o simples fato de achar que se escolheu aquela foto é suficiente para que queiramos dar suporte a essa decisão. Isso se deve entre vários fatores a querermos parecer racionais e inteligentes, já que essas são características socialmente desejadas. Afinal, se você sabe que algo é bom, você sabe *desde a primeira vez*, não precisa mudar de ideia.

Então temos a ascensão da linguagem como uma condição importante para o surgimento daquilo que chamamos de pessoas, e seus *selfs*. Também com a linguagem, e a capacidade de um raciocínio linear, repetível e reensaiável, surge a possibilidade de reverberação memética, a possibilidade de eco na câmara de pensamento humana, a possibilidade daquilo que Dennett entende como consciência. As palavras e a linguagem permitem não apenas, como destacou Chomsky, gerar infinitas ideias a partir de palavras finitas e regras finitas, mas também repetir sentenças e frases, ou frases de mesmo sentido, até que incorporem aquilo a nossa maneira de ver a nós mesmos. Poucas coisas são mais fortes como constituição da nossa consciência e do que entendemos por nós mesmos do que algo que já vimos repetir em nós (explícita ou implicitamente) diversas vezes.

Nosso mundo está cada vez mais populoso, agora, além de bactérias, ferro, níquel, ele possui extintores, números, humanos, marxistas, memes. Talvez apenas mais um habitante pré-2010 deva ainda adentrar nossa já rechonchuda ontologia: os

computadores. Computadores não são demasiado diferentes de outras entidades que exibem intencionalidade, principalmente quando os pensamos da perspectiva intencional. Da perspectiva física, porém, a implementação da “inteligência” que um computador possui é demasiadamente diferente da dos organismos que conhecemos. De início, um computador é em geral um imóvel, isolado do mundo para melhor funcionamento de suas partes internas, e armazenando informação em forma eletromagnética distinta de todas as formas utilizadas por lontras, zebras e nematelmintos. Dentro de um computador a compartimentalização de funções é muito mais determinada do que num cérebro animal, o processamento do computador é majoritariamente linear, enquanto o nosso cérebro é paralelo (se Dennett estiver certo, imitando um processamento linear, mas ainda assim paralelo). A velocidade relativa de computação de diferentes funções é muito diferente para um computador e um esquilo, ou um computador e um humano. Os computadores também têm uma origem distinta dos demais mecanismos inteligentes que habitam nosso planeta, uma das razões causais para suas diferenças.

Falamos anteriormente de mundos nocionais como os mundos que estão representados no inventário mental de um ser vivo, ou de um computador. Mundos representacionais, que mapeiam o território daquilo que o organismo processa como o mundo a seu redor. Que a ontologia do mundo não esteja sendo cravada em suas juntas na descrição que fiz até aqui é um truísmo, mas enquanto sabemos que o mundo não é apenas composto desses simples objetos até aqui descritos, podemos afirmar que conhecemos agora muito melhor um mundo nocional. Nossa brevíssima história de quase tudo revela a ontologia, a epistemologia, e, mais importante, o mundo nocional de Daniel Dennett.

15. A perspectiva dennettiana

Nossa jornada por uma simulação intelectual do funcionamento mental de Daniel Dennett começa a se encerrar por aqui. Neste capítulo examinaremos de início como a perspectiva dennettiana facilita a compreensão de algumas partes do início do texto, agora que estamos mais familiarizados com ela.

No início do texto, ao falar sobre reducionismo, mencionei a frase atribuída (falsamente, ou interpretativamente) a Einstein: “Devemos tornar as coisas o mais simples possível, mas não mais simples que isso.”

¿Por que escolhi dizer que é possível dizermos ‘falsamente’ ou ‘interpretativamente’ no mesmo caso? Porque da perspectiva dennettiana o que define Einstein é o centro gravitacional narrativo daquilo que circunda Einstein, e essa frase faz parte daquilo que Einstein poderia muito bem ter dito. A criação de falsos memes de citação como esse é muitas vezes justamente um indicativo de que *aquela pessoa*, aquele *self*, poderia muito bem ter dito aquilo. Então assim como atribuímos a Einstein um traço de personalidade (que não está na microfísica de Einstein, mas serve para prevermos o que ele fará amanhã), podemos atribuir essa frase a Einstein para prever o que ele escreverá amanhã. Algumas das frases que Einstein disse podem até ser ditas *menos* einstenianas do que essa, que ele nunca disse.

Logo em seguida comentei que Dennett posiciona-se vigorosamente contra um tipo de reducionismo que ele chama de reducionismo cobiçoso (*greedy reductionism*). E introduzi um abtruso parágrafo:

Essa forma exagerada de reducionismo, presente por exemplo no pensamento do psicólogo behaviorista Skinner, seria basicamente um desejo de reduzir absolutamente um fenômeno de larga escala a um fenômeno de pequena escala. Mas isso sem procurar compreender e reconhecer a existência de uma reestruturação numa hierarquia intermediária que permite que o fenômeno se comporte de maneira diferente numa escala maior do que seria esperado que ocorresse se estivéssemos apenas prevendo o comportamento microscópico de suas partes, não porque as macro partes se comportem mal, mas porque nós não conseguimos calcular seu comportamento. A razão pela qual essas reestruturações se dão, segundo Dennett, é a seleção natural, na qual a existência de replicadores permite uma guerra de reproducibilidade mais acelerada, e esta por sua vez gera novas estruturas materiais que se reproduzem mais rápido que suas antecessoras, ganhando espaço possível de design. (Caleiro, 2014, pp. 22-3.)

Skinner é um reducionista cobiçoso porque ele não se importa com a intratabilidade do cálculo de todos os movimentos de Marvin Minsky da perspectiva física. Para ele, era só declarar que “em princípio” o reducionismo é possível e a questão estaria resolvida. Os níveis intermediários, onde se encontram os estados biológicos, os órgãos, os estados mentais e as colunas corticais, não interessam ao psicólogo Skinner. Para ele, podemos dar o salto computacional diretamente da perspectiva física para o todo. Já sabemos que, para Dennett, ao menos dois outros degraus não de existir na nossa análise, a perspectiva de design e a perspectiva intencional. Ambas reduzem substancialmente a intratabilidade, e, admitamos, não temos ainda total certeza de que aquele “em princípio” desejado por Skinner de fato ocorra. A razão pela qual as outras perspectivas são necessárias no lugar de uma análise puramente reducionista tem a ver com a seleção natural. Uma aproximação newtoniana prevê bem o comportamento de uma rocha, de magma, ou uma cachoeira, a aproximação começa a se perder ao tentarmos computar um coqueiro, ou, ainda pior, uma sala na qual ocorre uma conversa entre Marvin Minsky e Karl Popper. Skinner colocaria os pombos e Popper junto com as rochas, e se eu e você estamos rodando versões parecidas da simulação de Dennett, não podemos permitir isso.

Convido o simulador a reler as definições de conceitos listados no capítulo 1 (pp. 14-5), para reencaixar suas versões ressignificadas com todo o edifício dennettiano que viemos até aqui construindo:

Pessoa: Centro de gravidade narrativo, um *abstracta* que organiza nossas representações mentais a respeito de outros e de nós mesmos. Centro do conjunto de histórias atribuídas a um indivíduo.

Abstracta: Qualquer coisa que tem o mesmo grau de realidade que um centro de gravidade, que existe, mas não tem eficácia causal, que *está lá*, mas não é causalmente responsável por nada. Ainda assim podem figurar nas explicações causais como parte de um modelo que as descreva.

Máquinas de Turing: *Abstração matemática* que a partir de regras muito simples consegue executar tarefas complexas, desde que programada para tal e desde que haja um sistema de codificação e decodificação.

Computador: *Objeto físico* que a partir de regras muito simples executa operações muito parecidas com uma Máquina de Turing, e as disponibiliza em interfaces cada dia mais naturais.

Mente: “Si abbiamo un’anima. Ma è fata di tanti piccoli robot.”
Tradução: Sim, temos uma alma, mas ela é feita de muitos pequenos robôs.

Evolução: Explicação para todo o *design* que persiste no mundo, *only game in town* (i.e., única possibilidade). Perigosa ideia de Darwin, pois corrói todas as estruturas e ameaça convicções milenares; principal explicação para os botos, o botox, os botões e os batons, todos produzidos pelo mesmo algoritmo.

Algoritmo: Conjunto de operações extremamente simples, descritíveis

numa única linguagem, que qualquer idiota conseguiria executar. Sequência ordenada de operações simples.

Sistemas intencionais: Sistemas para os quais a melhor estratégia preditiva é supor que eles tenham *crenças, intenções e desejos*, e uma racionalidade parcial limitada a seus mundos nocionais.

Sintaxe: Conjunto de regras formais de operação dentro de um sistema *ou* conjunto de operações ocorrendo em sistemas que não têm processamento de “significado”, por exemplo sistemas físicos.

Memes: Unidades replicadoras a partir das quais a evolução cultural deve ser pensada. [...].

Um *strange loop* afinal

Este texto, como qualquer outro, é um aglomerado de memes, que tendem a se replicar melhor quando juntos do que quando separados. A repetição é uma forma de reforçar as correntes corticais responsáveis pelo pensamento num meme, e a repetição num lugar diferente é uma forma de reforçar conexões conceituais mais vastas e intrincadas entre os diferentes memes que coabitam e disputam espaço em nossas cabeças. Neste momento, enquanto você processa informação visual e semântica (¿semântica?) dessas palavras, alguns milhares de resquícios de conceitos que passaram por sua mente estão reverberando, auxiliando (e atrapalhando) o processamento dessas novas ideias, conectando-se com elas, aumentando seu poder de previsão da próxima letra, da próxima xxxxxxxx, ou até mesmo xx xxxxxxxx xxxxx⁶⁴. Essa reverberação de todos esses memes internos, que agora já estão parcialmente desfigurados, mas ainda operantes e determinantes em seu comportamento, isso é *você*. Ser você é estar processando informação sob essas circunstâncias e desse jeito, é isso que diferencia sua consciência da de outros, seu self do de outros. Um conjunto de símbolos opera dentro de você, organizando um indivíduo que é assim, desse jeitinho tão seu. Você está consciente de tudo aquilo que está reverberando no seu funcionamento mental e interferindo em como você entende que é entendido, como se comporta e como interpreta o comportamento de outros. Isso tudo, é claro, se o aglomerado de memes dentro de minha mente que eu acho que descrevem bem o aglomerado de memes na mente de Dennett estiver certo. Aliás, porque não colocar mais níveis de intencionalidade, e dizer logo de uma vez que você só é isso se o aglomerado de memes que a sua mente está usando para interpretar o aglomerado que é esse livro for bem correlacionado com o aglomerado que é o meu sub-aglomerado que representa a visão

⁶⁴ Aqui caberiam diversas formas de dizer o mesmo meme: da próxima ‘sentença’, ‘frase’, ‘meme’, ‘ideia’. Se você simulou o que há de invariante entre as quatro sentenças, isso já demonstra o ponto.

de Dennett, sendo que a própria visão de Dennett é um aglomerado a respeito do mundo do sistema cognitivo e sensorial de um primata americano. Mais uma vez, não devemos elevar demais o número de níveis de intencionalidade numa sentença, sempre que pudermos evitar.

Volto-me agora por fim a uma questão parcialmente negligenciada, que Dennett (1997, pp. 17-18) comenta em seu “Fé na verdade”, que citei no início do texto (p. 12):

É um fato que as pessoas não querem muitas vezes saber a verdade. E é um fato mais inquietante que as pessoas não queiram muitas vezes que os *outros* saibam a verdade. [...] eles podem muito bem não querer saber a verdade, e podem ter razão em não querer saber a verdade, mas o fato é que acreditam na verdade; isso é claro. Eles sabem que a verdade está aí, para ser evitada ou abraçada, e sabem que a verdade é importante. É *por isso* que eles podem muito bem não querer saber a verdade. Porque a verdade pode magoar. magoar.

Algumas verdades magoam muito mais do que outras, Dennett parece se envolver fortemente com questões pesadas nesse sentido. A biologia contemporânea magoa muita gente, todos aqueles que tinham a esperança de que o design viesse de cima, que de algum jeito existissem *skyhooks*, ganchos celestiais, ou, mais tradicionalmente, a mão divina. A perigosa ideia de Darwin corrói fortemente essa perspectiva, e, ao explicar toda a maravilha da natureza, modifica a visão de mundo de muitos para algo pior do que se ela fosse falsa. No entanto, para a maioria, a visão de mundo trazida por Darwin e consagrada por Dawkins e Dennett é muito mais interessante, fantástica e bela do que seriam as antigas cosmogonias. Que a natureza haja esculpido, a partir de tão simples regras, tamanha complexidade, é algo maravilhoso.

Para Dennett (1995, p. 520), conseguimos na biologia, através da engenharia reversa, a explicação de porque a natureza parece como parece, e é como é:

O que é o trabalho do *design*? É aquele casamento maravilhoso do acaso e da necessidade, acontecendo em um trilhão de lugares de uma vez, em um trilhão de níveis diferentes. E que milagre causou isso? Nenhum. Simplesmente calhou de acontecer, na totalidade do tempo. Você até poderia dizer, de certa maneira, que a Árvore da Vida criou a si mesma. Não num *vuuush* milagroso e instantâneo, mas lenta e lentamente, ao longo de bilhões de anos.

É essa Árvore da Vida um Deus que se poderia adorar? Para o qual se poderia rezar? Temer? Provavelmente não. [...] A Árvore da Vida não é perfeita nem infinita no espaço ou no tempo, mas é atual, e se não for o que Anselmo chamou “Aquilo do qual não se pode pensar nada maior”, ela é seguramente um ser maior do que qualquer coisa que qualquer um de nós jamais vai conceber em detalhes que mereçam seus detalhes. É algo sagrado? Sim, eu digo, com

Nietzsche. Eu não poderia rezar para isso, mas eu posso erguer-me em afirmação a sua magnificência. Este mundo é sagrado.

Nietzsche foi um dos primeiros a compreender de uma maneira mais naturalista o todo da natureza. Em sua época, não havia companhia para seus pensamentos, sua visão de mundo estava solitária, tamanhamente solitária, que ele criou a ideia de espíritos-livres, que servem para ele como companhias intelectuais, amigos, com os quais podemos sentar, tagarelar, e rir. Os tempos de Nietzsche eram demasiadamente áridos intelectualmente, e sua mente estava situada no topo, onde o ar é rarefeito. Graças a Nietzsche, Russell, Darwin, Mendel, Sagan, Dawkins, Dennett, e centenas de outros, não vivemos mais num tempo onde compreender os mistérios do universo é algo que leva à solidão intelectual, e a uma angústia intempestiva diante da surdez do cosmos às nossas ansiedades. O cosmos pode não responder a nossas angústias, mas parte dele pode compartilhá-las conosco, conversar sobre elas, e até resolver algumas. Nós seres humanos, parte do cosmos, como seres sociais, somos capazes de compartilhar nossas angústias, de interconectar nosso mundo mental com o de outra pessoa, de empatizar, e de, ao gerar uma fundação comum, tornarmo-nos parte um do outro, em certo nível. Essa possibilidade de compartilhar e fundir, não só em pensamento, mas em emoção, no choque diante da maravilha de existir, é algo de fascinante, e o poder expressivo da linguagem tem dificuldade de fazer jus a uma cena como essa (leia como se fosse o narrador):

Depois de tantas eras, dias e noites no escuro, sob camadas e camadas de lençóis e névoas, posso agora me levantar, e olhar além. Em pouco mais de vinte anos, refiz a jornada mental de toda a minha espécie, compreendi milhares de vezes mais do que um ser humano jamais poderia compreender sozinho. A herança escrita me permitiu entender o mundo que me circunda. Posso ver com uma clareza que nunca se imaginou atingir o que são essas mãos, esses pés, dedos. Dedos!

Convido meus amigos para sentarmos no parque, tomando nossas bebidas favoritas, num fim de tarde ensolarado. Ali, juntos, observamos ainda outra vez a queda da bola de fogo. Se a visão é mesmerizante por ela mesma, é o que dizer do fato de podermos nos olhar nos olhos e sentir que compartilhamos o conhecimento de que lá, distante, existe de fato uma bola gigantesca, queimando e se deixando entender por nós? Que possamos olhar para o horizonte, numa encosta, abraçados, e compreender tão bem de onde viemos, porque estamos juntos, porque estamos ali, como chegamos, o que

somos, e o que estamos vendo, o que é *ver*. Que possamos, além de entender o mundo ao nosso redor, e a nós mesmos, sentir as emoções esculpidas dentro de nós, que possamos ao mesmo tempo experienciar o amor ao próximo e a totalidade. Que sem medo possamos de um só golpe amar e amar ao amor, e entender o amor e amar o entendimento. Por estes prêmios devemos agradecimento a esses intelectuais.

Que pela primeira vez podemos conviver moralmente ao mesmo tempo com nossas emoções morais e com nossa visão da realidade, porque não fomos mais criados dentro do dogma. Que possamos saber porque um abraço é prazeroso, o que é entender o pensamento de um amigo, que possamos ouvir uma ideia e compreender como ela chegou até ali, tudo isso são coisas que compelem a suspirar em respeito à força que é a natureza. A verdade pode magoar, mas a verdade também pode inspirar. A Natureza, a verdadeira Natureza, tal como ela é, não tal como fingíamos que era, merece reverência.

Bibliografia

- Abrantes, P.C. & Almeida, F.P.L. (2011). Evolução humana: A teoria da dupla herança. In: Abrantes, P.C. (ed.). *Filosofia da biologia*. Porto Alegre: Artmed, p. 261-95.
- Adams, D. (2004). *Guia do mochileiro das galáxias*. Rio de Janeiro: Sextante. Original em inglês: 1979.
- Adorno, T. & Horkheimer, M. (1985). *Dialética do esclarecimento*. Rio de Janeiro: Jorge Zahar.
- Ainslie, G. (2001). *Breakdown of will*. Cambridge: Cambridge University Press.
- Arntzenius, F. (2008). No regrets, or: Edith Piaf revamps decision theory. *Erkenntnis* 68, 277-97.
- Aunger, R. (ed.). (2000). *Darwinizing culture*. Oxford: Oxford University Press.
- Bennett, M.R. (2007). *Neuroscience and philosophy: Brain, mind, and language*. New York: Columbia University Press.
- Blackmore, S.J. (2000). *The meme machine*. Oxford: Oxford University Press.
- (2003). Consciousness in meme machines. *Journal of Consciousness Studies* 10, 4-5.
- (2006). *Conversations on consciousness: what the best minds think about the brain, free will, and what it means to be human*. Oxford: Oxford University Press.
- Block, N. (1980). Troubles with functionalism. *Readings in the philosophy of psychology*, vol. 1. Cambridge: Harvard University Press, p. 268-305.
- (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18, 227.
- (2002). The harder problem of consciousness. *Journal of Philosophy* 99, 391-425.
- Block, N. & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review* 108, 1-46.
- Boroditsky, L. (2009). How does language shape the way we think. In: Brockman, M.

- (ed.). *What's next: Dispatches on the future of science*. New York: Vintage, p. 116-29.
- Bostrom, N. (2002a). *Anthropic bias: Observation selection effects in science and philosophy*. New York: Routledge.
- (2002b). Existential risks. *Journal of Evolution and Technology* 9.
- (2003a). Are we living in a computer simulation? *Philosophical Quarterly* 53, 243-55.
- (2003b). Astronomical waste: The opportunity cost of delayed technological development. *Utilitas* 15, 308-14.
- (2004). The future of human evolution. In: Tandy, C. (ed.). *Death and anti-death: Two hundred years after Kant, fifty years after Turing*. Palo Alto: Ria University Press, p. 339-71.
- (2005). The fable of the dragon tyrant. *Journal of Medical Ethics* 31, 273-77.
- (2009). The future of humanity. *Geopolitics, History, and International Relations* 2, 41-78.
- (2011a). Infinite ethics. *Analysis and Metaphysics* 10, 9-59.
- (2011b). Information hazards: A typology of potential harms from knowledge. *Review of Contemporary Philosophy* 10, 44-79.
- (2013). Existential risk prevention as global priority. *Global Policy* 4, 15-31.
- Bostrom, N. & Ord, T. (2006). The reversal test: Eliminating status quo bias in applied ethics. *Ethics* 116, 656-79.
- Bostrom, N. & Sandberg, A. (2009). The wisdom of nature: an evolutionary heuristic for human enhancement. In: Savulescu, J. & N. Bostrom, N. (eds.). *Human enhancement*. Oxford: Oxford University Press, p. 375-416.
- Boyer, P.; Robbins, P. & Jack, A.I. (2005). Varieties of self-systems worth having. *Consciousness and Cognition* 14, 647-60.
- Brandt, L. & Brandt, P. A. (2005). Making sense of a blend: A cognitive-semiotic approach to metaphor. *Annual Review of Cognitive Linguistics* 3, 216-49.
- Brockman, J. (ed.). (1995). *Third culture: Beyond the scientific revolution*. New York: Simon and Schuster.

- Bromhall, C. (2004). *The eternal child: How evolution has made children of us all*. London: Ebury Press.
- Buss, D. M. (1999). *Evolutionary psychology: The new science of the mind*. Boston: Allyn & Bacon.
- (ed.) (2005). *The handbook of evolutionary psychology*. New York: Wiley.
- Caleiro, D. (2008). Why truthful memes have a bigger survival chance. Manuscrito não publicado.
- (2009). Every conscious machine brings us closer to death. Online: <http://www.tau.ac.il/humanities/philos/dascal/DA.pdf>. Apresentado no *VIII Encontro Brasileiro Internacional de Ciência Cognitiva*. Campinas.
- (2010). What isn't it like to be knowledge. Manuscrito não publicado.
- (2012). Fishing the content and epistemology of phenomenal belief. Manuscrito não publicado.
- (2014). *Simulando Dennett: Ferramentas e construções de um naturalista*. Dissertação de mestrado, Faculdade de Filosofia, Letras e Ciências Sociais, Universidade de São Paulo.
- Cantor, M., & Whitehead, H. (2013). The interplay between social networks and culture: theoretically and among whales and dolphins. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368, 1618.
- Carr, B. (Ed.). (2007). *Universe or multiverse?*. Cambridge University Press.
- Chalmers, D. J. (1996). Does a rock implement every finite-state automaton?. *Synthese* 108, 309-33.
- (2002a). Consciousness and its place in nature. *Philosophical Studies* 44, 197-200.
- (ed.) (2002b). *Philosophy of mind: Classical and contemporary readings*. Oxford University Press.
- (2003). The content and epistemology of phenomenal belief. In: Smith, Q. & Jolic, A. (eds.). *Consciousness: New philosophical perspectives*. Oxford: Oxford University Press, p. 220-72.
-

- Chalmers, D. J., & Jackson, F. (2001). Conceptual analysis and reductive explanation. *Philosophical Review* 110, 315-60.
- Ćirković, M.M. & Rees, M.J. (2008). *Global catastrophic risks*. Oxford: Oxford University Press.
- Ćirković, M.M.; Sandberg, A. & Bostrom, N. (2010). Anthropoc shadow: Observation selection effects and human extinction risks. *Risk Analysis* 30, 1495-1506.
- Clastres, P. (2003). *A sociedade contra o estado*. São Paulo: Cosac & Naify.
- Cole, D. (2013). The chinese room argument. In: Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, disponível online.
- Davidson, D. (1978). What metaphors mean. *Critical inquiry* 5, 31-47.
- (1984). *Inquiries into truth and interpretation*. Oxford: Oxford University Press.
- Darwin, C. (2002). *A origem das espécies*. Trad. E. Amado. Belo Horizonte: Itatiaia. Original em inglês: 1859.
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press. Em português: *O gene egoísta*. Trad. G.H.M. Florsheim. Belo Horizonte: Itatiaia, 1979.
- (1982). *The extended phenotype: The long reach of the gene*. Oxford: Oxford University Press.
- (1986). *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. New York: Norton. Em português: *O relojoeiro cego*. Trad. L.T. Motta. São Paulo: Cia. das Letras, 2001.
- (2005). *The ancestor's tale: a pilgrimage to the dawn of evolution*. London: Phoenix.
- Dennett, D.C. (1971). Intentional systems. *Journal of Philosophy* 68, 87-106. Republicado em Dennett (1978d / 2006). Sistemas intencionais, p. 33-55.
- (1975). Why the law of effect will not go away. *Journal for the Theory of Social Behaviour*, 5, 169-88. Republicado em Dennett (1978d / 2006). Por que a lei do efeito não irá embora, p. 119-40.
- (1976a). Are dreams experiences? *Philosophical Review* 85, 151-71. Republicado em Dennett (1978d / 2006). Os sonhos são experiências?, p. 187-209.

- (1976b). Conditions of personhood. In: Rorty, A.O. (ed.). *The identities of persons*. Berkeley: University of California Press, p. 175-96. Republicado em: Goodman, M.F. (ed.). *What is a Person?* Clifton (NJ): Humana, 1988, p. 145-67. Republicado em Dennett (1978d / 2006).
- (1978a). Artificial intelligence as philosophy and as psychology. In: Ringle, M. (ed.). *Philosophical perspectives on artificial intelligence*. New York: Humanties Press, p. 57-78. Republicado em Dennett (1978d / 2006). Inteligência artificial como filosofia e como psicologia, p. 163-83.
- (1978b). Skinner skinned. In Dennett (1978d), p. 53-70. Em português: Skinner “esfolado”, op. cit., p. 95-117.
- (1978c). Two approaches to mental images. In Dennett (1978d), p. 53-70. Em português: Duas abordagens às imagens mentais, op. cit., p. 241-59.
- (1978d). *Brainstorms: Philosophical essays on mind and psychology*. Cambridge: MIT Press. Em Português: *Brainstorms: Ensaios filosóficos sobre a mente e a psicologia*. Trad. L.H. de Araújo Dutra. São Paulo: Unesp, 2006.
- (1978e). The abilities of men and machines. In Dennett (1978d), p. 256-66. Texto original de 1970. Em português: As habilidades de homens e máquinas, op.cit.
- (1981a). True believers: the intentional strategy and why it works. In: Heath, A.F. (ed.), *Scientific explanation*. Oxford: Oxford University Press, p. 53-75. Republicado em Dennett (1987), p. 13-42.
- (1981b). Three kinds of intentional psychology. In: Healey, R. (ed.), *Reduction, time and reality*. Cambridge: Cambridge University Press, p. 37-61. Republicado em Dennett (1987), p. 43-68.
- (1981c). Making sense of ourselves. *Philosophical Topics* 12, 63-81. Republicado em Dennett (1987), p. 83-101.
- (1982a). How to study human consciousness empirically or nothing comes to mind. *Synthese* 53, 159-80.
- (1982b). Beyond belief. In: Woodfield, A. (ed.). *Thought and object*. Oxford: Clarendon. Republicado em Dennett (1987), p. 117-202.
- (1983a). Styles of mental representation. *Proceedings of the Aristotelian Society* 83 (1982/83), p. 213-26. Republicado em Dennett (1987), p. 213-25.
- (1983b). Intentional systems in cognitive ethology: the “Panglossian paradigm”

- defended. *Behavioral and Brain Sciences* 6, 343-55. Republicado em Dennett (1987), p. 237-68.
- (1986). The logical geography of computational approaches: A view from the East Pole. In: Harnish, R. & Brand, M. (ed.). *The representation of knowledge and belief*. Tucson: University of Arizona Press, p. 59-79.
- (1987). *The intentional stance*. Cambridge: MIT Press.
- (1988). Quining qualia. In: Marcel, A. & Bisiach, E. (eds.). *Consciousness in Modern Science*. Oxford: Oxford University Press, p. 42-77.
- (1990a). Memes and the exploitation of imagination. *The Journal of Aesthetics and Art Criticism* 48, 127-35.
- (1990b). The interpretation of texts, people and other artifacts. *Philosophy and Phenomenological Research* 50, 177-94.
- (1991a). *Consciousness explained*. New York: Penguin.
- (1991b). Real patterns. *Journal of Philosophy* 88, 27-51.
- (1992). The self as a center of narrative gravity. In: Kessel, F.; Cole, P. & Johnson, D. (eds.). *Self and consciousness: Multiple perspectives*. Hillsdale (NJ): Erlbaum, p. 103-14.
- (1994). Real consciousness. In: Revonsuo, A. & Kamppinen, M. (ed.). *Consciousness in philosophy and cognitive neuroscience*. Hillsdale (NJ): L. Erlbaum, p. 55-63.
- (1995a). *Darwin's dangerous idea: evolution and the meanings of life*. New York: Simon & Schuster. Em português: *A perigosa ideia de Darwin: a evolução e os significados da vida*. Trad. T.M. Rodrigues. Rio de Janeiro: Rocco, 1998.
- (1995b). Animal consciousness: what matters and why. *Social Research* 62, 691-710.
- (1996a). *Kinds of minds: Toward an understanding of consciousness*. New York: Basic Books. Em português: *Tipos de mentes*. Tradução A. Tort. Rio de Janeiro, Rocco, 1997.
- (1996b). Consciousness: More like fame than television. Online: <http://pp.kpnet.fi/seirioa/cdenn/concfame.htm>. Publicado em alemão: Bewusstsein hat mehr mit Ruhm als mit Fernsehen zu tun. In: Maar, C.; Pöppel, E. & Christaller, T. (eds.). *Die Technik auf dem Weg zur Seele*. Rowohlt, p. 61-90.

- (1996c). Facing backwards on the problem of consciousness. *Journal of Consciousness Studies* 3, 4-6.
- (1997). Fé na verdade. Trad. D. Murcho. *Disputatio* (Lisboa) 3, 3-21. Online: <http://www.disputatio.com/wp-content/uploads/1997/11/003-1.pdf>. Original em inglês: Faith in the truth. *Free Inquiry* 21(1) (2000) 40-2. Online: <http://ase.tufts.edu/cogstud/papers/faithint.htm>.
- (1998). *Brainchildren: Essays on designing minds, 1984-1996*. Cambridge: MIT Press.
- (2001). In Darwin's wake, where am I?. *Proceedings and Addresses of the American Philosophical Association* 75, 11-30.
- (2002a). *Content and consciousness*. Routledge.
- (2002b). The new replicators. In: Pagel, M. (ed.). *The encyclopedia of evolution*, vol.1. Oxford: Oxford University Press, p. E83-E92.
- (2003a). Who's on first? Heterophenomenology explained. *Journal of Consciousness Studies*, 10, 9-10.
- (2003b). *Freedom evolves*. New York: Penguin.
- (2004). Could there be a Darwinian account of human creativity. In: Moya, A. & Font, E. (eds.). *Evolution: from molecules to ecosystems*. Oxford: Oxford University Press, p. 272-9.
- (2005). *Sweet dreams: Philosophical obstacles to a science of consciousness*. MIT press.
- (2006). *Breaking the spell: Religion as a natural phenomenon*. New York: Penguin.
- (2007a). What RoboMary Knows. *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, 15-31.
- (2007b). Philosophy as naive anthropology: Comment on Bennett and Hacker. *Neuroscience and philosophy*, 73-95.
- (2008a). Making sense of ourselves. *Philosophical Topics*, 12(1), 63-81.
- (2008b). Autobiography: Part 1. *Philosophy Now* 68.
- (2013). *Intuition Pumps and Other Tools for Thinking*. WW Norton & Company.

- Deutsch, D. (1997). *The fabric of reality*. New York: Penguin. Em português: *A essência da realidade*. Trad. B.R. Fernandes. São Paulo: Makron, 2000.
- Diamond, J. M. (1992). *The rise and fall of the third chimpanzee*. New York: Vintage.
- (1998). *Why is sex fun? The evolution of human sexuality*. New York: Basic Books.
- (2006). *Collapse: How societies choose to fail or succeed*. New York: Penguin.
- Drescher, E. (2006). *Good and real: Demystifying paradoxes from physics to ethic*. Cambridge: MIT Press.
- Dretske, F. (1988). The stance stance. *Behavioral and Brain Sciences 11*: 511-2.
- Egan, G. (2010). *Permutation city*. London: Gollancz.
- Ellison, P.T. & Gray, P. B. (2009). *Endocrinology of social relationships*. Cambridge: Harvard University Press.
- Fabiano, J.L. (2013). Humble primary intensions: Fixing two-dimensional semantics. *Analysis and Metaphysics 12*, 105-15.
- Feynman, R. (1985). *QED: The strange theory of light and matter*. Princeton: Princeton University Press.
- Gazzaniga, M.S. (ed.). (2004). *The cognitive neurosciences III*. Cambridge: MIT Press.
- Gilbert, D. (2006). *Stumbling on happiness*. New York: Knopf.
- (2012). *The surprising science of happiness*. Disponível em: http://www.ted.com/talks/dan_gilbert_asks_why_are_we_happy.html
- Gilovich, T.; Griffin, D. & Kahneman, D. (eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.
- Gould, S.J. (2001). *Lance de dados: A ideia de evolução de Platão a Darwin*. Rio de Janeiro: Record.
- Gregory, T.R. (2005). Genome size evolution in animals. In: Gregory, T.R. (ed.). *The evolution of the genome*. San Diego: Elsevier, p. 4-87.
- Hall, J.S. (2007). *Beyond AI: Creating the conscience of the machine*. Amherst (NY): Prometheus Books.
- Harris, S. (2011). *The moral landscape: How science can determine human values*. New

York: Free Press.

Hawking, S. (1988). *Uma breve história do tempo: do Big Bang aos buracos negros*. Rio de Janeiro: Rocco.

Hawkins, J. & Blakeslee, S. (2005). *On intelligence*. New York: St. Martin's Griffin.

Henrich, J.; Boyd, R. & Richerson, P.J. (2008). Five misunderstandings about cultural evolution. *Human Nature* 19, 119-37.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An eternal golden braid*. New York: Vintage.

——— (1996). *Metamagical themas: Questing for the essence of mind and pattern*. New York: Basic Books.

——— (2007). *I am a strange loop*. New York: Basic Books

Hofstadter, D.R., & Dennett, D C. (ed.) (2006). *The mind's I: Fantasies and reflections on self & soul*.

Hölldobler, B. & Wilson, E.O. (2009). *The superorganism: the beauty, elegance, and strangeness of insect societies*. New York: Norton.

Horgan, T. & Tienson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In: Chalmers, D. (ed.). *Philosophy of mind: Classical and contemporary readings*. Oxford: Oxford University Press, p. 520-33.

Horton, J.C. & Sincich, L.C. (2004). A new foundation for the visual cortical hierarchy. In: Gazzaniga, M.S. (ed.). *The cognitive neurosciences III*. Cambridge: MIT Press, p. 233-44.

Hyde, J.S.; Biswal, B.B. & Jesmanowicz, A. (2000). Optimum voxel size in fMRI. *Proceedings of the International Society of Magnetic Resonance in Medicine* 8, 239.

Jablonka, E. & Lamb, M. J. (2005). *Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. Cambridge: MIT Press.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 127-136.

Jeffreys, W.H. & Berger, J.O. (1992). Ockham's razor and Bayesian analysis. *American Scientist*, 80(1), 64-72.

- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan. Em português: *Rápido e devagar: duas formas de pensar*. Trad. C.A. Leite. Rio de Janeiro: Objetiva, 2012.
- Kant, I. (1980). *Crítica da razão pura*. São Paulo: Abril Cultural.
- Kaplan, D. (1989). Demonstratives. In: Almog, J.; Perry, J. & Wettstein, H. (eds.). *Themes from Kaplan*. New York: Oxford University Press, p. 481-563.
- Knobe, J. (2007). Experimental philosophy. *Philosophy Compass* 2(1), 81-92.
- Knobe, J. & Nichols, S. (eds.) (2008). *Experimental philosophy*. Oxford: Oxford University Press.
- Koch, C. (2004). *The quest for consciousness*. Englewood (CO): Roberts & Company.
- Kripke, S.A. (1980). *Naming and necessity*. Cambridge (MA): Harvard University Press
- Kurzweil, R. (2006). *The singularity is near: When humans transcend biology*. New York: Penguin Books.
- Lakoff, G. (1999). *Philosophy in the flesh*. New York: Basic Books.
- Lewis, D.K. (1973). General semantics. In: Davidson, D. & Harman, G. (eds.). *Semantics of natural language*. Dordrecht: Reidel, p. 169-218.
- (1974). Radical interpretation. *Synthese* 27: 331-44
- (1979a). Attitude de dicto and de se. *Philosophical Review* 88, 513-43.
- (1979b). Scorekeeping in a language game. *Journal of Philosophical Logic* 8, 339-59.
- (1980a). A subjectivist's guide to objective chance. In: Harper, W.L. et al. (eds.). *Ifs*. Dordrecht: Reidel, p. 267-97.
- (1980b). Probabilities of conditionals and conditional probabilities. In: Harper, W.L. et al. (eds.). *Ifs*. Dordrecht: Reidel, p. 129-147.
- (1983). *Philosophical papers*. Vol. I. New York: Oxford University Press.
- (1986a). Probabilities of conditionals and conditional probabilities II. *Philosophical Review* 95, 581-9.
- (1987). *Philosophical Papers*. Vol. II. New York: Oxford University Press.

- (1989). Dispositional theories of value. *Proceedings of the Aristotelian Society* (suppl.) 63, 113-37.
- (2000). Causation as influence. *The Journal of Philosophy* 97, 182-97.
- (2013). *Counterfactuals*. New York: John Wiley & Sons.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* 8, 529-66.
- Lieberman, M.D. (2009). What makes big ideas sticky? In: Brockman, M. (ed.). *What's next*. New York, Vintage, p. 90-103.
- Lyubomirsky, S. (2008). *The how of happiness*. New York: Penguin Press.
- McElreath, R. & Boyd, R. (2007). *Mathematical models of social evolution: A guide for the perplexed*. Chicago: University of Chicago Press.
- Minsky, M. (2007). *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. New York: Simon and Schuster.
- Mirolli, M. & Dennett, D. (2002). A naturalistic perspective on intentionality. Interview with Daniel Dennett. *Mind & Society* 3(2), 1-12.
- Mithen, S. (1996). *The prehistory of the mind: a search for the origins of art, religion and science*. London: Thames & Hudson. Em português: *Pré-história da mente: uma busca das origens da arte, da religião e da ciências*. São Paulo: Unesp, 2002.
- Morris, D. (2004). *The nature of happiness*. London: Little Books.
- Nagel, E. & Newman, J.R. (2001). *Gödel's proof*. New York: New York University Press
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review* 83, 435-50.
- Omohundro, S. M. (2008). The basic AI drives. *Frontiers in Artificial Intelligence and applications* 171, 483-92.
- Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.
- (2007). Is personal identity what matters? Texto premiado pelo *Ammonius Foundation*. Online: <http://marcsandersfoundation.com/assets/pdfs/ammoniusfinal.pdf>.

- Pearl, J. (2000). *Causality: models, reasoning and inference*. Cambridge (MA): MIT Press.
- Pereira, F., Mitchell, T. & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 45, S199-S209.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- (2003). *The blank slate: The modern denial of human nature*. New York: Penguin.
- (2007a). *The language instinct: How the mind creates language*. New York: Harper Perennial Modern Classics
- (2007b). *The stuff of thought: Language as a window into human nature*. New York: Penguin.
- (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences* 107 (supplement 2), 8993-99.
- (2011). *The better angels of our nature: Why violence has declined*. New York: Penguin.
- Popper, K.R. & Eccles, J.C. (1977). *The self and its brain*. Berlin: Springer.
- Putnam, H. (1975). The meaning of 'meaning'. In: Keith Gunderson, K. (ed.). *Language, Mind and Knowledge*. Minnesota Studies in the Philosophy of Science, vol. 7. Minneapolis: University of Minnesota Press, p. 131-93.
- (1981). *Reason, truth and history*. Cambridge: Cambridge University Press.
- (1988). *Representation and reality*. Cambridge (MA): MIT Press.
- Quine, W.V.O. (1948). On what there is. *Review of Metaphysics* 2, 21-38.
- (1976). *The ways of paradox, and other essays*. Cambridge: Harvard University Press.
- (1980). *From a logical point of view*. Cambridge (MA): Harvard University Press.
- (2000). Two dogmas of empiricism. In: Stainton, R.J. (ed.). *Perspectives in the Philosophy of Language: A Concise Anthology*. Peterborough (Canada): Broadview, p. 189-210.

- Rawls, J. (1999). *A theory of justice*. Cambridge: Harvard University Press.
- Rees, M. (2005) Earth in its final century? TED conference speech. Disponible em: http://www.ted.com/index.php/talks/martin_rees_asks_is_this_our_final_century.html.
- Richardson, P.J. & Boyd, R. (2005). *Not by genes alone*. Chicago: University of Chicago Press.
- Romo, R.; Lafuente, V. & Hernández, A.. (2004). Somatosensory discrimination: neural coding and decision-making mechanisms. In: Gazzaniga, M.S. (ed.). *The cognitive neurosciences III*. Cambridge: MIT Press, p. 197-214.
- Rorty, R. (1980). *Philosophy and the mirror of nature*. Princeton: Princeton University Press.
- Russell, B. (1905). On denoting. *Mind* 14(56), 479-93.
- (1910). Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society II*, 108-128.
- (1923). Vagueness. *Australasian Journal of Psychology and Philosophy* 1(2), 84-92.
- (1926). *Icarus: or, the Future of Science*. Lonon: K. Paul, Trench, Trubner.
- (1927). *The analysis of matter*. London: Kegan Paul.
- (1949). *History of western philosophy*. New York: Simon & Schuster.
- (1957). *Why I am not a Christian, and other essays on religion and related subjects*. New York: Simon & Schuster.
- (1998). *Mortals and others: Bertrand Russell's American essays, 1931-1935*. London: Routledge.
- (2008). *Introduction to mathematical philosophy*. Nottingham: Spokesman Books. Original: 1919.
- (2009). *ABC of Relativity*. New York: Taylor & Francis.
- (2010). *The problems of philosophy*. Boston: Mobile Reference.
- Russell, B. & Whitehead, A. N. (1968). *Principia mathematica*. Cambridge: Cambridge University Press.

- Russell, S.J.; Norvig, P.; Canny, J. F.; Malik, J. M. & Edwards, D. D. (1995). *Artificial intelligence: a modern approach*. Vol. 2. Englewood Cliffs (NJ): Prentice Hall.
- Sandberg, A. & Bostrom, N. (2008). Whole brain emulation: A roadmap. Future of Humanity Institute, Oxford University. Online: <http://www.fhi.ox.ac.uk/Reports/2008-3.pdf>.
- Seabright, P. (2010). *The company of strangers: A natural history of economic life*. Princeton: Princeton University Press.
- Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences* 3, 417–57.
- Shoemaker, S. (1975). Functionalism and qualia. *Philosophical studies* 27, 291-315.
- Smith, P. (1988). Wit and chutzpah. *Times Higher Education Supplement*, 7 de agosto, p. 22.
- Soltis, J.; Boyd, R. & Richerson, P. J. (1995). Can group-functional behaviors evolve by cultural group selection? An empirical test. *Current Anthropology* 36, 473-94.
- Stalnaker, R. (2002). What is it like to be a zombie. In: Gendler, T.S. & Hawthorne, J. (eds.). *Conceivability and possibility*. New York: Oxford University Press, p. 385-400.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT Press.
- Tomasello, M. & Carpenter, M. (2007). Shared intentionality. *Developmental Science* 10, 121-5.
- Tomasello, M.; Carpenter, M.; Call, J.; Behne, T. & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28, 675-90.
- Tononi, G. & Edelman, G.M. (1998). Consciousness and complexity. *Science* 282, 1846-51.
- Tononi, G.; Sporns, O. & Edelman, G.M. (1999). Measures of degeneracy and redundancy in biological networks. *Proceedings of the National Academy of Sciences* 96, 3257-62.
- Turing, A. (2004). Intelligent machinery (1948). In: Copeland, B.J. (ed.). *The essential Turing*. Oxford: Clarendon, p. 395-432.
- Tyler, T. (2011). *Memetics: Memes and the science of cultural evolution*. Lexington

(KY): Mersenne Publishing.

Vacca, J.R. & Hawking, S.W. (2005). *The world's 20 greatest unsolved problems*. Upper Saddle River (NJ): Prentice Hall PTR.

Wallach, W. & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. New York: Oxford University Press.

Watkins, M. (1989). The knowledge argument against 'the knowledge argument'. *Analysis* 49, 158-60.

Wilson, D.S. & Wilson, E.O. (2007). Rethinking the theoretical foundation of sociobiology. *Quarterly Review of Biology* 82, 327-48.

Wilson, E. O. & Holldobler, B. (2009) *The Superorganism: The beauty, elegance and strangeness of insect societies*. New York: Norton.

Woods, V. (2010). *Bonobo handshake: A memoir of love and adventure in the Congo*. Collingwood (Australia): Black.

Yalowitz, S. (1997). Rationality and the argument for anomalous monism. *Philosophical Studies* 87, 235-58.

Yudkowsky, E. (2003). An intuitive explanation of Bayesian reasoning. Online: <http://www.yudkowsky.net/bayes/bayes.html>.

——— (2004). Coherent extrapolated volition. Online:<http://singinst.org/upload/CEV.html>.

——— (2006). World's most important math problem. Online: <http://pt.scribd.com/doc/2327578/Worlds-Most-Important-Math-Problem-Eliezer-Yudkowsky-Future-Salon>

——— (2007). Levels of organization in general intelligence. In *Artificial General Intelligence* (pp. 389-501). Springer Berlin Heidelberg.

——— (2008a). Cognitive biases potentially affecting judgment of global risks. In: Bostrom, N. & Čirković, M.M. (eds.). *Global catastrophic risks*. Oxford: Oxford University Press, p. 91-119.

——— (2008b). The simple truth. Online: <http://yudkowsky.net/rational/the-simple-truth>.

——— (2010). Timeless decision theory. Online: <http://intelligence.org/files/TDT.pdf>.

—— (s/d) A technical explanation of technical explanation. Online:
<http://yudkowsky.net/rational/technical>

Zimmer, C. (2003). *O livro de ouro da evolução: o triunfo de uma ideia*. Rio de Janeiro:
Ediouro.