



UNIVERSIDADE FEDERAL DO CEARÁ
INSTITUTO DE CULTURA E ARTE
CURSO DE FILOSOFIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

HENRIQUE ALVES COSTA

**UM CONCEITO DE INFORMAÇÃO PARA A TEORIA REPRESENTACIONAL DA
MENTE**

FORTALEZA

2020

HENRIQUE ALVES COSTA

UM CONCEITO DE INFORMAÇÃO PARA A TEORIA REPRESENTACIONAL DA
MENTE

Dissertação apresentada ao Programa de Pós-Graduação em Filosofia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Filosofia. Área de concentração: Filosofia da Linguagem e do Conhecimento.

Orientador: Prof. Dr. Cícero Antônio Cavalcante Barroso.

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

C872c Costa, Henrique Alves Costa.
Um conceito de informação para a Teoria Representacional da Mente / Henrique Alves Costa Costa. –
2020.
92 f. : il.

Dissertação (mestrado) – Universidade Federal do Ceará, Instituto de cultura e Arte, Programa de Pós-
Graduação em Filosofia, Fortaleza, 2020.
Orientação: Prof. Dr. Cícero Antônio Cavalcante Barroso.

1. Informação. 2. Computação. 3. Representação. 4. Conteúdo mental. I. Título.

CDD 100

HENRIQUE ALVES COSTA

UM CONCEITO DE INFORMAÇÃO PARA A TEORIA REPRESENTACIONAL DA
MENTE

Dissertação apresentada ao Programa de Pós-Graduação em Filosofia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Filosofia. Área de concentração: Filosofia da Linguagem e do Conhecimento.

Aprovada em: 10/03/2020.

BANCA EXAMINADORA

Prof. Dr. Cícero Antônio Cavalcante Barroso (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Carlos Eduardo Fisch de Brito
Universidade Federal do Ceará (UFC)

Profa. Dra. Maria Eunice Quilici Gonzalez
Universidade Estadual Paulista (UNESP)

A todos aqueles que se dedicaram e se dedicam
à pesquisa no Brasil.

AGRADECIMENTOS

Ao Prof. Dr. Cícero Antônio Cavalcante Barroso, que me orientou nesta dissertação e que também coordenou, nos meus tempos de graduação, as atividades de iniciação científica que realizei sob sua tutela. Não seria exagero dizer que, sem seu auxílio, tanto ao ensinar-me o ofício da pesquisa quanto boa parte dos assuntos debatidos aqui (através da disciplina Tópicos Especiais em Filosofia da Mente I e das discussões sobre filosofia da informação na iniciação científica e em outras disciplinas) durante a graduação, sem sua orientação e seu constante apoio e exemplo, este trabalho jamais teria surgido.

Aos professores Ralph Leal Heck, participante da banca examinadora de qualificação, Maria Eunice Quilici Gonzalez, participante da banca examinadora de defesa, e Carlos Eduardo Fisch de Brito, participante de ambas. Ao primeiro, agradeço a minúcia na leitura do texto e nas sugestões e correções. Aos demais, agradeço, além das sugestões de correção, a instigante discussão sobre o tema e as sugestões de pesquisa, que certamente abriram perspectivas futuras.

Ao Grupo de Pesquisa Filosofia, Metafísica e Cognição (antigo Pensamento, Cognição e Linguagem), na figura dos professores coordenadores Luís Estevinha e Cícero Barroso, por manter um vivo e instigante ambiente de debate em filosofia analítica na Universidade Federal do Ceará.

Aos colegas da turma de mestrado, em especial a Yuri Rodrigues, Ícaro Martins, Victória de Oliveira, Marcos Caetano, Octavio García, Renan Esteves e provavelmente outros que esquecerei de nomear (mas não por maldade), pelas valiosas discussões e sugestões de pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) - Código de Financiamento 001.

“Synapse to synapse / The possibilities thin...”
(COMPANY CALLS EPILOGUE, 2000, faixa
8).

RESUMO

Nesta dissertação, abordamos a noção de que computação é processamento de informação a fim de cumprir o objetivo de verificar se esta noção, por mais difundida que seja, é uma concepção apropriada de computação, ao menos no âmbito da ciência cognitiva. Precisamente, buscamos uma definição de informação no contexto da mente humana, ou seja, realizar uma análise daquilo que denominamos informação mental. Para cumprir estes objetivos, defenderemos uma versão da Teoria Computacional da Mente, a Teoria Representacional da Mente, e forneceremos um conceito de informação para ela a partir da ideia de que os portadores de informação na mente são representações mentais. Além disso, evidenciaremos o caráter semântico da informação mental a partir de sua relação com representações mentais e mostraremos como, a nosso ver, ela é processada. Abordaremos também algumas noções de informação e de computação diferentes da nossa, apontando em que pontos nos aproximamos e em nos afastamos delas e, por fim, responderemos a um dos principais problemas colocados às Teorias Computacionais da Mente: a crítica de que toda atribuição computacional é trivial (isto é, dependente de um observador ou agente externo). No decorrer da nossa resposta, apontaremos para outros problemas colocados à inteligência artificial e à ciência cognitiva, como o problema do quarto chinês e o problema do homúnculo, além de discutirmos brevemente defesas de requisitos semânticos para a computação e sua relação com nossa defesa do requisito representacional para a mente.

Palavras-chave: Informação. Computação. Representação. Conteúdo mental.

ABSTRACT

In this thesis, we approach the definition of computation as information processing to check if this definition, although broadly accepted, is an appropriate notion of computation, at least to the cognitive sciences. More precisely, we seek a definition of information as applied to the human mind, i.e., an analysis of what we call mental information. In order to achieve these goals, we will support one of the Computational Theories of Mind, the Representational Theory of Mind, and will offer a concept of information to it from the idea that the bearers of information in mind are mental representations. Moreover, we will make explicit the semantic character of mental information (through a justification of its particular character as information) from its relation to mental representations and will show how, in our view, it is processed. We will also approach some alternative concepts of information and computation, showing in which points we agree and in which points we disagree with them. At last, we will give an answer to one of the main problems put to Computation Theories of Mind: that all computational attributions are trivial (i.e., dependent on an observer or external agent). As we put forward our answer, we will briefly deal with some other problems put to artificial intelligence and cognitive science, such as the Chinese Room argument and the Homunculus problem, and will briefly discuss defenses of semantic requirements to computation and their relation to our defense of a representational requirement to mind.

Keywords: Information. Computation. Representation. Mental content.

LISTA DE FIGURAS

Figura 1	–	O sistema de comunicação de Shannon	38
Figura 2	–	Síntese esquemática de nossa aplicação do sistema comunicacional de Shannon à mente	46
Figura 3	–	Resumo esquemático (<i>tower bridge</i>) de concepções funcionais da computação	59
Figura 4	–	Versão do esquema <i>tower bridge</i> para a concepção de computação de Dietrich	60

SUMÁRIO

1	INTRODUÇÃO	11
2	POR QUE PENSAR NA MENTE COMO UM COMPUTADOR?	14
2.1	Teorias computacionais da mente	25
3	O QUÊ, O PORQUÊ E O COMO DA INFORMAÇÃO MENTAL	37
3.1	Aplicações do conceito de informação à mente	37
3.1.1	<i>Informação de Shannon</i>	37
3.1.2	<i>Informação representacional</i>	40
3.1.3	<i>Informação mental</i>	44
3.2	A relação de nossa visão da informação mental com outras abordagens do tema	47
3.2.1	<i>A ontologia da informação de Radu Bogdan</i>	47
3.2.2	<i>Concepções funcionais da informação mental</i>	58
4	UMA DEFESA DA ATRIBUIÇÃO COMPUTACIONAL À MENTE	64
4.1	O argumento do quarto chinês	64
4.2	Searle contra o cognitivismo	66
4.3	O problema do homúnculo	68
4.4	Demais críticas de Searle ao cognitivismo	70
4.5	Um pouco de Chalmers sobre computação e cognição	73
4.6	O que é uma atribuição de computação não-trivial? A exigência representacional	77
4.7	Computação semântica, representação, informação	81
5	CONCLUSÃO	84
5.1	O problema do homúnculo outra vez: descarregando os <i>homunculi</i>	84
5.2	Em que sentido computações requerem representações?	86
5.3	Computação é processamento de informação?	88
	REFERÊNCIAS	90

1 INTRODUÇÃO

As teorias que defendem que a mente humana funciona como um computador se enquadram sob a denominação geral de Teorias Computacionais da Mente (TCM). A principal representante delas é a Teoria Representacional da Mente (TRM), cujo principal expoente é Jerry Fodor. Deve-se entender aqui que esta teoria defende que o funcionamento da mente compartilha diversas características com o funcionamento de máquinas de Turing – o modelo abstrato de todos os computadores existentes –, mas não necessariamente que nossas mentes são máquinas de Turing. Certas características destas, como o processamento em série e a memória não-endereçável, por exemplo, não parecem descrever como nossa mente opera (RESCORLA, 2017)¹. Por outro lado, a TRM, ao menos enquanto defendida por Fodor, incorpora também uma linguagem de pensamento².

A princípio, queremos compreender uma noção pressuposta por toda e qualquer TCM: a noção de computação. Falar em uma TCM implica entender operações mentais como computações, mas o que é uma computação? Uma maneira comum de defini-la é dizer que computação é processamento de informação; porém, isso não resolve o problema, uma vez que nessa definição ambos os termos também carecem de esclarecimento, em especial o termo ‘informação’. Desta forma, a informação se torna o novo objeto de nossa pesquisa³. Em

- 1 Cabe menção aqui a uma teoria computacional alternativa à TRM, a saber, o conexionismo. Ambas diferem no que diz respeito ao modelo computacional adotado: enquanto na TRM o modelo é a máquina de Turing, no conexionismo são as redes neurais. Uma rede neural é composta por um conjunto de nós divididos em três camadas: nós de entrada, nós de saída e nós ocultos, que fazem o intermédio entre os nós de entrada e os nós de saída. Nós têm valores de ativação dados por números reais, ou seja, quando a corrente que atravessa o circuito atinge determinado limiar, o nó é ativado. Eles estabelecem conexões entre si (daí o nome “conexionismo”) que possuem determinada força ou peso; estas conexões também têm valores de ativação. Os valores de ativação de nós ocultos e de nós de saída é uma função da soma de seus pesos (isto é, das conexões que estabelecem com os nós de entrada), a que se chama ativação total de entrada (RESCORLA, 2017). TRM e conexionismo também diferem quanto ao requerimento de representações para a computação: comumente se afirma que, diferente de computadores clássicos, redes neurais não empregam representações em suas operações computacionais (apesar de isto não necessariamente dever ser o caso). O melhor exemplo que podemos dar de uma rede neural, até mesmo sugerido por seu nome (apesar de não poder ser reproduzido em todas as suas operações pela mesma) é o cérebro humano: nós podem ser facilmente vistos como neurônios, pesos como sinapses etc.
- 2 Por isso, talvez, seja mais exato denominar a teoria de Fodor não apenas de RTM, mas sim de RTM + LoTH (sigla para Language of Thought Hypothesis), o que se expressa na seguinte definição: “...a atividade mental envolve computação estilo-Turing [ou seja, operações sobre representações] sobre a linguagem do pensamento [que compõe o conjunto de representações em questão].” (RESCORLA, 2017).
- 3 Existem outras concepções de computação além da que apresentamos: a funcional-valorativa (EGAN, 1991, 1992, 1999), a estrutural (CHALMERS, 1995, 1996, 2011, 2012), a mecanicista (PICCININI, 2008) e até mesmo uma concepção pluralista (EDELMAN, 2008). Não as abordamos aqui (a não ser quando necessário no contexto de alguma discussão) porque é apenas na concepção de computação apresentada por nós que vemos as noções de informação e computação intrinsecamente interligadas; uma vez que a informação é o

especial, o que poderíamos chamar de informação mental, isto é, a informação envolvida em operações mentais.

Acreditamos que o caráter semântico da informação mental é uma de suas características essenciais, pois, se a mente operasse com informação sem caráter semântico, não seríamos capazes de explicar a capacidade de representações mentais serem, efetivamente, representações, isto é, possuírem capacidade de representar objetos e estados de coisas do mundo. Por outro lado, temos bons motivos para pensar que boa parte da atividade mental, senão toda ela, não é representacional, motivos amplamente apresentados por defensores da teoria computacional conexionista da mente (ao menos no âmbito da TCM). Mais uma vez vemos como nosso tema, inicialmente desligado de outras questões filosóficas da TCM, exige uma resposta a pelo menos uma delas antes de ser tratado em particular. É importante saber qual das duas versões da TCM, a representacional ou a conexionista, a descreve melhor, pois isto determina se a mente opera ou não com representações mentais; por sua vez, é importante termos uma resposta a esta última questão para sabermos se a informação mental tem caráter semântico ou não: se a mente não é representacional, então podemos falar em informação, com relação à ela, no sentido usual; de fato, o próprio termo “informação mental” perderia seu sentido, uma vez que poderíamos falar simplesmente em informação (assim como não pressupomos nenhum caráter – nem denominação – particular à informação processada, por exemplo, pelos nossos computadores). Mas, se a mente é efetivamente representacional, há sentido em falar em informação mental, isto é, informação com caráter semântico processada pela mente.

Em resposta a estas questões, seguiremos aqui um percurso que permitirá, entre outras coisas: 1) introduzir os conceitos básicos à compreensão da TCM e das suas áreas científicas correlatas, como a inteligência artificial e a ciência cognitiva; 2) defender a TRM (na versão elaborada por Fodor) como teoria computacional da mente; 3) mostrar o elo entre representação e informação a partir de uma teoria causal da representação; 4) elaborar um modelo do processamento de informação mental para a TRM.

Acreditamos que a representação deve ser compreendida de maneira causal, ou seja, que a relação entre representante e representado é tal como uma relação de indicação ou de significação natural (GRICE, 1991). A partir desta concepção da representação, podemos estabelecer o elo entre representação e informação e, assim, explicar como a mente pode ser

nosso principal objeto de estudo, esta concepção de computação, no presente trabalho, ganha prioridade.

vista como um dispositivo de processamento de informação (mental). Além disso, podemos mostrar o porquê de a informação mental ser semântica. O modelo de processamento da informação mental, por sua vez, visa mostrar como ela é processada.

No decorrer da dissertação, sustentaremos melhor a teoria causal da representação de certos ataques a ela, como o problema do erro e o problema da disjunção, compararemos nosso modelo do processamento da informação mental e nossa concepção dela com concepções alternativas da informação e da computação para a mente e, por fim, formularemos uma resposta tentativa a um dos principais problemas para o empreendimento pressuposto e desenvolvido por nós, o problema da trivialidade de atribuições computacionais à mente (SEARLE, 2006), mostrando como conceber a mente como representacional é a melhor maneira de defender a não-trivialidade desta atribuição.

2 POR QUE PENSAR NA MENTE COMO UM COMPUTADOR?

Poder-se-ia compreender a introdução da Teoria Computacional da Mente (TCM) no cenário filosófico como um corolário dos avanços recentes na área da computação, assim como houve teorias da mente que decorreram de outros desenvolvimentos. Esta concepção foi expressa, em filosofia, pelo filósofo norte-americano John Searle⁴:

Porque não compreendemos muito bem o cérebro, somos constantemente tentados a usar a última tecnologia como um modelo para o tentar compreender. Na minha infância, asseguravam-nos que o cérebro era um quadro telefônico. (...) Sherrington, o grande neurocientista britânico, pensava que o cérebro trabalhava como um sistema telegráfico. Freud comparou muitas vezes o cérebro a sistemas hidráulicos e electromagnéticos. Leibniz comparou-o a um moinho e disseram-me que alguns dos antigos gregos pensaram que o cérebro funciona como uma catapulta. Hoje em dia, como é óbvio, a metáfora é o computador digital. (SEARLE, 1984, p. 55-56).

Conforme veremos, o fato de sermos “constantemente tentados a usar a última tecnologia como um modelo para... tentar compreender [a mente]” nem é um dos motivos pelos quais filósofos concebem a mente como um computador, nem mesmo poderia fazer parte deles, devido a sua inadequação. Mas deixemos as razões para rejeitar tal raciocínio para um outro momento.

Quando falamos das relações entre mente, cérebro e computador, é importante ter em mente que podemos formular diversas perguntas a respeito delas. Isso é necessário se quisermos chegar a respostas claras e corretas, pois a confusão entre estas perguntas é responsável por mais de um mal-entendido filosófico a respeito da TCM e muitas das críticas dirigidas a ela têm aí sua origem. Conforme entendemos (com CRANE, 2003, p. 83-84), podemos nos perguntar por três coisas ao lidar com aquelas relações:

- a) Um computador é capaz de modelar a mente humana?;
- b) Quais são as condições necessárias e suficientes para dizermos que algo tem uma mente e, a partir disto, atribuímos mentalidade a computadores?;
- c) Quais são as condições necessárias e suficientes para dizermos que a mente humana é um computador?

4 Desde que compreendamos ‘cérebro’ como ‘mente’ em suas diversas aplicações ao longo do trecho que citamos (ou como termos correferenciais).

As perguntas que nos interessam mais de perto são as perguntas 2 e 3. Com efeito, e apesar de serem constantemente aplicados em ciência cognitiva com grande proveito, modelos da mente humana não constituem o fenômeno a ser estudado aqui – modelos econômicos não são a economia, modelos de crescimento populacional não são o próprio crescimento populacional, etc⁵. É claro que utilizamos modelos nos estudos empíricos que consideram a mente um computador, mas o que está em questão em nossa pesquisa é exatamente o fato de a mente humana ser um computador. Poder-se-ia dizer, entretanto, que parece que temos no computador um caso especial: um caso em que o modelo computacional também executa a atividade que pretende modelar (BODEN, 2008, p. 742). Uma comparação permite entender melhor este ponto: modelos meteorológicos não fazem chover, mas modelos de busca efetivamente realizam buscas. Por isso, a réplica diz que devemos levá-los em consideração no estudo a ser desenvolvido aqui. Mas há aqui uma petição de princípio na medida em que se usa de uma resposta afirmativa à questão 2 para estabelecer a importância em respondê-la. O correto é que, neste caso, ainda nos perguntemos: o computador tem uma mente de fato, ou somente a emula? Por essas e outras razões, acreditamos que a resposta à questão 1, por maior que seja o interesse em obtê-la, não é nem leva a uma resposta às perguntas filosoficamente interessantes sobre a TCM.

As perguntas 2 e 3, entretanto, são repletas de interesse filosófico. Algumas considerações adicionais a elas são necessárias. A pergunta 2 é provavelmente melhor entendida como uma versão do problema das outras mentes, em que buscamos saber sob que condições podemos afirmar que algo possui uma mente. Para respondê-la, é necessário estabelecer as condições necessárias e suficientes envolvidas e a que se aplicam; em especial, contemporaneamente, o debate gira em torno de saber se computadores cumprem ou não essas condições. Quem procura responder à pergunta 3 não está necessariamente interessado na capacidade de atribuir ou não a existência de mentes a outros seres ou objetos, pois o objeto de seu estudo é a mente humana, a qual cada um pode saber que existe. Seu objetivo é saber se essa mente pode ser compreendida computacionalmente.

Por outro lado, há pelo menos um ponto em que as duas perguntas se interseccionam: se a pressuposição da pergunta 2 estiver correta, isto é, se um computador

5 Há porém casos em que uma modelagem computacional bem-sucedida é apontada como exemplo do sucesso explicativo da TCM. É o que acontece, por exemplo, em Boden (2008).

cumpra as condições necessárias e suficientes para possuir uma mente, então não precisamos ser céticos quanto a outras mentes considerando a mente humana como o paradigma para o mental; de fato, o paradigma é o computador e nossas operações mentais só recebem essa qualificação em relação a ele. Pelo contrário, na pergunta 3 pressupõe-se que a mente é, por excelência, a mente humana e a partir daí se busca explicá-la e se chega a razões para pensar que a melhor explicação de suas propriedades é computacional. A partir daí, acredita-se que a mente humana é computacional, mas isto não diz nada a respeito da mente em geral. De qualquer forma, a distinção que realizamos serve apenas ao propósito de clarificar as questões que podem ser colocadas a respeito da TCM e de determinar os problemas que nos interessam mais de perto; na literatura da área, porém, tudo é muito mais confuso do que fizemos parecer aqui e em muitos casos é difícil determinar a qual das perguntas um filósofo pretende responder, ou qual dos empreendimentos um filósofo tem em vista ao fazer uma crítica à TCM. Para citar um exemplo disso, o bastante conhecido argumento do quarto chinês, formulado por John Searle (1984) - que abordaremos em detalhe adiante -, por exemplo, pode ser rejeitado por um filósofo que procura responder à pergunta 3 a partir do pressuposto de que é possível respondê-la afirmativamente (isto é, que a mente humana é um sistema computacional e que podemos estabelecer critérios necessários e suficientes para caracterizá-la como tal), pois questiona o pressuposto aceito por filósofos que buscam uma resposta à pergunta 2 de que computadores cumprem os requisitos necessários e suficientes da mentalidade em geral (não necessariamente aceito por quem busca uma resposta à pergunta 3).

Toda essa digressão serve ao propósito de explicar uma diáspora em nosso objetivo inicial – mostrar por que há filósofos que entendem a mente como um computador – que ocorre devido à própria divisão de objetivos presente no campo da TCM. Portanto, para responder a essa questão, precisamos estabelecer:

- a) quais razões suportam a tese de que computadores cumprem as condições necessárias e suficientes para possuir uma mente;
- b) quais razões suportam a tese de que a mente humana cumpre as condições necessárias e suficientes para ser considerada um computador⁶.

6 Tentar entender ambas as questões em termos de condições necessárias e suficientes pode ser desencaminhador. A primeira delas certamente pode ser entendida assim, mas a segunda recebeu um tratamento filosófico impactante que fez a pergunta pela melhor explicação assumir esse papel.

Começemos pelo primeiro ponto.

Em linhas gerais, podemos dizer que o que motiva filósofos a dizerem que computadores cumprem as condições necessárias e suficientes para possuir uma mente é o fato de que eles são, por definição, capazes de executar qualquer conjunto bem definido de instruções que visa um objetivo específico e que eles são efetivos em cumpri-lo. Como qualquer atividade pode ser descrita como um tal conjunto de instruções (desde que esse conjunto seja finito)⁷, as atividades mentais certamente também o podem. Logo, atividades mentais, uma vez descritíveis nesses termos, são executáveis por um computador.

Certa terminologia é necessária para uma compreensão melhor e mais precisa dos motivos que apresentamos logo acima para defender a ideia de que computadores possuem mentes. O “conjunto bem definido de instruções que visa um objetivo específico e é efetivo em cumpri-lo” pode ser chamado, em uma palavra, de algoritmo. Uma definição formal mais precisa para algoritmo é “um método... para calcular o valor de uma função” (CRANE, 2003, p. 87). Uma função pode ser entendida como uma composição cujos elementos são entradas, uma caixa-preta e saídas, fornecidas a partir das entradas por meio de um processo realizado pela caixa preta (CARNIELLI & EPSTEIN, 2009, p. 46). Esta recebe esse nome porque, comumente, não nos interessa o procedimento que realiza para produzir as saídas. Por sua vez, damos o nome de domínio ao conjunto de entradas de uma função e o nome de imagem ao conjunto de suas saídas (idem, p. 48).

Duas caixas-pretas podem representar uma mesma função; isso ocorre quando geram as mesmas saídas a partir das mesmas entradas. Nem toda composição do tipo que apresentamos é, entretanto, uma função. Para que seja considerada como tal, ela deve cumprir o requisito de não ser ambígua, isto é, não gerar mais de um resultado a partir de uma mesma entrada (idem, p. 47).

Podemos definir a noção de função mais formalmente da seguinte forma: considere dois conjuntos, X e Y, não-vazios. Uma função de X em Y é um conjunto de pares

7 Se o conjunto de instruções que descreve a atividade é finito, a atividade é executável. É possível que uma atividade seja descritível apenas através de um conjunto infinito de instruções; neste caso, ela não é executável, assim como no caso em que ela não é descritível. O termo ‘executável’, por motivos que veremos adiante, deve ser compreendido como ‘computável’; ele comparece aqui apenas por motivos didáticos.

Talvez esta afirmação seja melhor compreendida como um pressuposto da resposta afirmativa à pergunta 2: toda atividade mental é descritível e “executável”.

ordenados $\langle x, y \rangle$ em que x pertence a X , y pertence a Y e, se $\langle x, y \rangle$ e $\langle x, z \rangle$ pertencem ao conjunto, então $y=z$.

Mas qual é a relação da matemática com aquilo que estamos trabalhando aqui? Essa definição não ajuda a entendermos qual é a importância da noção de algoritmo para a compreensão das atividades mentais; afinal, não as compreendemos como meros cálculos de funções matemáticas. Por outro lado, essa reação é fruto de uma compreensão do conceito de função restrita ao campo da matemática; entretanto, pelo que vimos, funções podem ser aplicadas a diversos âmbitos.

Em nossa definição informal de algoritmo, dissemos também que ele é efetivo em cumprir um objetivo específico. Um outro nome pelo qual se costuma chamá-lo, aliás, é procedimento efetivo, exatamente porque proceder de acordo com um algoritmo deve levar invariavelmente ao resultado esperado. Certamente se pode criar um procedimento que não cumpra os requisitos para ser um algoritmo, mas neste caso ou ele não implementa uma função, ou implementa uma função diferente da desejada, gerando resultados inesperados⁸.

De qualquer forma, a efetividade do algoritmo é garantida através de duas condições: 1) seus procedimentos são claros. Os estágios que compõem o algoritmo não deixam margem à nenhuma dúvida e a execução das transições entre eles não exige nenhuma interpretação especial; 2) o conjunto de seus procedimentos é finito (ou: ele pode ser descrito em uma quantidade finita de passos)⁹.

Dizer que um computador cumpre as condições necessárias e suficientes para ter uma mente, ou que a mente pode ser compreendida como um computador, implica dizer que a mente opera através de computações. Mas computações são operações formalmente definidas como execuções de algoritmos, daí a importância de toda a explicação anterior a respeito deles. Há também uma definição formal de ‘computador’: uma máquina capaz de executar algoritmos. A descrição geral desta máquina (ou melhor, uma descrição de uma máquina capaz de executar qualquer algoritmo) foi formulada por Alan Turing em 1936 e leva seu

8 A título de exemplo: na função $y=x+1$, o argumento $x=2$ gera o valor $y=3$.

9 Desde já se apresentam motivos para uma das principais críticas contra a TCM: a pressuposição do caráter sistemático da mente. Afinal, um algoritmo é um procedimento sistemático e, se as operações da mente podem ser descritas em algoritmos, a mente é também sistemática. Falaremos melhor disso adiante, mas de antemão pode-se dizer que essa suposição é, na verdade, uma inferência para a melhor explicação: entendemos melhor o funcionamento da mente se pressupormos que seu funcionamento é sistemático.

nome.

Turing não o fez de caso pensado. Apesar de, durante a Segunda Guerra Mundial, haver discutido sobre máquinas pensantes com Claude Shannon, pioneiro da Teoria da Informação, nos Laboratórios Bell (GLEICK, 2013, p. 212-213), ele não anteviu todo o progresso científico que levou os computadores a serem o que são hoje. Não, pelo menos, no estágio inicial de seu pensamento, em que o objetivo era fornecer uma solução ao *Entscheidungsproblem* (problema da decisão), um problema que assolava a matemática há anos. Este problema foi colocado por David Hilbert em 1928 e consiste na pergunta acerca da decidibilidade da matemática, isto é, na existência (ou inexistência) de “... um rigoroso procedimento passo a passo por meio do qual, dada uma linguagem formal de raciocínio dedutivo, seria possível realizar automaticamente uma demonstração.” (GLEICK, 2013, p. 214)¹⁰. Em sua resposta ao problema da decisão, Turing perguntou-se a respeito da computabilidade dos números. Para Turing, “Os números ‘computáveis’ podem ser descritos de forma breve como os números reais cujas expressões decimais são calculáveis por meios finitos. (...) De acordo com minha definição, um número é computável se seu decimal pode ser escrito por uma máquina” (TURING, 1936, p. 116, apud CARNIELLI; EPSTEIN, 2009, p. 105). Assim, Turing considerou uma computação um procedimento mecânico e forneceu o modelo de uma máquina capaz de realizar qualquer operação matemática possível, assim como a linguagem na qual programá-la.

Como nenhuma máquina existente à época fornecia um modelo à sua noção de computação, ele criou uma máquina abstrata (idem, p. 216). Essa máquina abstrata, porém, está atualmente instanciada em boa parte dos computadores concretos que conhecemos hoje. A máquina de Turing (daqui em diante, MT) é talvez melhor compreendida como “uma especificação teórica e abstrata de uma máquina possível... seu objetivo não é, em primeiro lugar, ser construída, mas ilustrar algumas propriedades gerais de algoritmos e computações.” (CRANE, 2003, p. 92-93). As principais delas eram: 1) esse dispositivo deve ser capaz de executar qualquer computação possível, independente de sua complexidade; 2) sua descrição deve ser a mais simples possível. (idem, p. 92).

Uma MT é uma máquina composta por uma fita “unidimensional” (isto é, uma

¹⁰ Retomaremos o problema em mais detalhes a partir da p. 24.

fita na qual só se escreve símbolos em um de seus lados) dividida em quadrados e potencialmente infinita. O conjunto de símbolos que podem ser escritos pela máquina é finito e sequências de símbolos podem ser tratadas como outros símbolos (compostos)¹¹. Além da fita, a máquina é também composta por um computador, ao qual se atribui as seguintes operações: observar o símbolo escrito em um quadrado da fita, escrever um símbolo em um dos quadrados, apagar um símbolo da fita, mudar sua distribuição (movê-la para a esquerda ou para a direita, um quadrado por vez) e assumir um estado mental, cujo número é finito, por operação¹². O conjunto de operações realizáveis pela máquina se chama tabela da máquina. O símbolo observado pelo computador em dado momento e seu estado mental determinam completamente seu comportamento e o estado posterior da máquina. Uma instrução de uma MT é uma quádrupla ordenada na forma $\langle q_i, S, Op, q_j \rangle$, em que q_i é o estado inicial da máquina, S é o símbolo escrito no quadrado observado por ela neste estado, Op é a operação realizada por ela e q_j é o novo estado que ela assume (CARNIELLI; EPSTEIN, 2009, p. 109). A instrução de uma MT também pode ser descrita como um conjunto de sentenças condicionais, da seguinte maneira: “se a máquina está no estado X e lendo o símbolo S , então ela realizará uma certa operação... e mudará para o estado Y (ou permanecerá no mesmo estado) e moverá a fita para a direita/esquerda.” (CRANE, 2003, p. 93). Um conjunto de instruções constitui a tabela da máquina, ou seu programa.

Desde já, há uma possível objeção ao uso da MT para a compreensão da mente humana, que, por ser brevemente solucionável e ser mais fruto de uma estranheza ou confusão que uma objeção propriamente dita, gostaríamos de abordar já aqui. Pode parecer que quem argumenta que a MT incorpora elementos mentais ou possui uma mente incorre em uma petição de princípio, pois sua formalização assume, de antemão, que ela pode assumir “estados mentais”. Seria por essa capacidade de “incorporar um fantasma” dentro de si que a máquina seria capaz de realizar as operações que realiza e, assim, servir para a compreensão da nossa própria mente; já pressupomos que ela é capaz de assumir estados mentais para demonstrar que ela é capaz disso. Como resposta a isso, precisamos esclarecer que não há nada de fantasmagórico nos estados mentais assumidos pela máquina. O próprio Turing

11 Apesar disso, trabalharemos aqui diretamente com símbolos e operações simples, isto é, que não podem ser decompostas em outros símbolos e operações mais simples. Logo, assumimos que o computador é capaz de observar, escrever ou alterar apenas um símbolo simples da máquina por vez, e é capaz de assumir apenas um estado mental por operação.

12 As restrições quanto ao número de quadrados manipuláveis pela MT e quanto ao número de estados que ela pode assumir por operação servem a fins de simplificação.

comenta a respeito disso em seu artigo clássico sobre o tema:

É sempre possível, para o computador, parar seu trabalho, ir embora e esquecer tudo sobre ele e mais tarde voltar e continuá-lo. Se o computador faz isso, ele deve deixar uma nota com instruções (escritas em alguma forma padrão) explicando como o trabalho deve ser continuado. Esta nota é a contrapartida do ‘estado da mente’. (...) o estado do progresso da computação, em qualquer estágio, é completamente determinado pela nota de instruções e pelos símbolos da fita. Isto é, o estado do sistema pode ser descrito por uma expressão única (sequência de símbolos), consistindo dos símbolos na fita seguidos por Δ (...) e então pelas instruções. Esta expressão pode ser chamada de ‘fórmula de estado’. (TURING, 1936, p. 140, *apud* CARNIELLI; EPSTEIN, 2009, p. 108).

Podemos, portanto, entender o “estado mental” de uma MT como aquilo que determina como a máquina deve proceder a partir de dado estado. Não há nada de fantasmagórico nele: trata-se apenas de uma instrução para que ela execute o passo seguinte.

A título de simplificação, consideremos que $S = \{0, 1\}$, ou seja, assumamos que a máquina apenas é capaz de escrever o símbolo ‘1’ na fita e que o símbolo ‘0’ representa um quadrado vazio (CARNIELLI; EPSTEIN, 2009, p. 108-109). Assumir uma notação binária para a MT traz uma consequência interessante: ela nos permite representar qualquer operação nela, independente do nível de complexidade (CRANE, 2003, p. 97). Uma consequência disso é que a própria MT é codificável e pode servir de entrada para outra MT. Isso é possível se codificarmos sua tabela em zeros e uns, o que pode ser feito ao se atribuir números - codificados em notação binária - às operações e estados da máquina. Como a MT é completamente definida por sua tabela, ao fazermos isso também a estamos codificando. Uma vez codificada, ela pode servir como entrada para outra MT e pode ser observada por ela, desde que esta máquina possua um método de converter as operações daquela em operações próprias - ou, por assim dizer, imitar seu comportamento (*idem*, p. 98). Uma consequência disso é que não é necessário que façamos uma MT para cada operação matemática existente: basta que coloquemos a máquina capaz de realizar certa operação como entrada de outra máquina capaz de imitá-la. Mas, se toda MT pode ser codificada em notação binária, há uma máquina capaz de imitar o comportamento de todas as outras, e assim, em vez de definir diversas MTs - uma para cada operação matemática existente, executando seu algoritmo -, precisamos apenas de uma MT que aceite como entrada tabelas de todas estas MTs, codificadas em S. Esta máquina é chamada de Máquina de Turing Universal (daqui em diante, MTU) e é o modelo de boa parte dos computadores atuais.

Como Turing respondeu ao problema da decisão? Em termos mais precisos, o problema da decisão é o problema de saber se podemos decidir, para qualquer proposição P da matemática, se P é V ou F . Isso pode ser transformado num problema de lógica, pois equivale ao desafio de saber, para qualquer proposição P da matemática, se P é uma consequência lógica de certos axiomas. Isso, por sua vez, só podemos determinar, em geral, se produzimos uma prova de P a partir dos axiomas. Nesse ponto, vemos que a decidibilidade da matemática (via lógica), depende de acharmos provas de teoremas. Mas buscar uma prova é executar um procedimento que pode parar (achando a prova de P ou a prova de $\sim P$) ou não. Assim, a matemática (e a lógica de predicados) só é decidível se existe um mecanismo que determina, para cada procedimento de busca de uma prova, se ele para ou não. Turing nos deu o modelo geral deste procedimento e mostrou, a partir de seu teorema da parada, que esse mecanismo não existe. Logo, a matemática (e a lógica de predicados) não é decidível.

Pensar que uma MTU pode servir como descrição da nossa mente é, por um lado, algo completamente contraintuitivo: parece haver um abismo entre o tipo de operações para as quais ela foi pensada – cálculos matemáticos – e o conjunto das operações que nós, humanos, somos capazes de realizar¹³. Por outro lado, somos diariamente confrontados com exemplos de MTs capazes de realizar diversas operações que jamais seríamos capazes de imaginá-las realizando: os computadores pessoais. De qualquer forma, a explicação acima nos permite entender como é possível que MTs desempenhem estas funções: qualquer conjunto de instruções pode ser codificado em uma tabela de uma MT e alimentado em uma MTU.

A partir de tudo o que foi dito, uma exigência à resposta à primeira questão que apresentamos (quais são as condições necessárias e suficientes para dizermos que algo tem uma mente - e, portanto, para dizermos que computadores pensam?) se coloca: “A questão sobre computadores pensantes é (em parte) sobre se o processamento de informação que *computadores* realizam pode ter algo a ver com o ‘processamento de informação’ envolvido no *pensamento*.” (CRANE, 2003, p. 111, grifos do autor). Um problema em especial é a pluralidade de definições do termo ‘informação’, o que torna bastante provável que o processamento de informação envolvido em um caso seja diferente daquele envolvido no outro (DODIG-CRNKOVIC, 2006). Por outro lado, enquanto há um conceito bastante preciso

13 Esta, em última instância, é uma das principais críticas postas à Inteligência Artificial: o problema da criatividade.

de informação aplicável a computadores - o conceito de informação de Shannon (SHANNON, 1948; SHANNON; WEAVER, 1972) –, não temos nada parecido no que diz respeito à mente, apesar de haver alguns candidatos - que mencionaremos no decorrer do trabalho.

Gostaríamos de responder, neste momento, à crítica feita à Teoria Computacional da Mente com a qual iniciamos este capítulo: ela seria apenas mais um exemplo de como a mente foi vista à luz da tecnologia de uma época. Em resposta a isso, pode-se dizer duas coisas. A primeira é que o contexto histórico em que uma ideia foi desenvolvida não diz nada a respeito de sua correção. A segunda é que, por um lado, diferente dos casos citados por Searle anteriormente, as etapas percorridas no desenvolvimento dos computadores caminharam lado a lado com tentativas de sistematizar o conhecimento humano e compreender suas capacidades intelectuais através da sistematização ou codificação do pensamento humano (CRANE, 2003, p. 112, 114). Por outro lado, nenhum defensor desta ideia afirma que nossas teorias computacionais atuais são o melhor modelo para compreendermos a mente. Eles afirmam que, qualquer que seja a melhor teoria computacional a ser desenvolvida, ela será também uma teoria da mente (BODEN, 2008, p. 756).

Boa parte do que foi dito até aqui se refere ao segundo dos três empreendimentos que apresentamos acima, a saber, aquele que se pergunta pelas condições necessárias e suficientes para dizermos que algo tem uma mente – e, portanto, para dizermos que computadores pensam. Cientificamente, este empreendimento se identifica com a área de investigação da inteligência artificial. Inteligência, aqui, não necessariamente se identifica com o que é conhecido como processos de alto nível, como o pensamento; antes, engloba coisas como reconhecimento de texto, visão tridimensional, ou tarefas específicas (sistemas experts) como diagnósticos médicos simples, por exemplo (CRANE, 2003, p. 114-115). Para se realizar tais empreendimentos, geralmente se parte do pressuposto de a mente, ou ao menos algumas tarefas mentais, podem ser modeladas computacionalmente, partindo-se daí à construção destes modelos. Em outros empreendimentos, parte-se da ideia de que a mente seja efetivamente um computador para, a partir daí, construir computadores pensantes. No primeiro caso, estamos nos deparando com o que ficou conhecido como inteligência artificial fraca, ou simplesmente IA-fraca. No segundo, estamos nos deparando com a inteligência artificial forte, ou IA-forte¹⁴.

14 Segundo John Searle, o autor desta distinção, a IA-forte é definida como “a concepção de que tudo que tem

É chegado o momento de considerar mais detalhadamente as questões que levantamos no início deste capítulo. De fato, como afirmamos acima, um defensor da IA-forte – alguém que responde afirmativamente à pergunta 2 - acredita que construir um computador pensante nos diz algo a respeito do que o pensamento é em geral. Por outro lado, alguém que responde afirmativamente à pergunta 3 pode defender que pensar é algo que humanos fazem; não conhecemos nenhum outro ser capaz disso. Se não temos razões para dizer, por exemplo, que outros animais pensam, porque achamos que estamos em melhor condição ao imputar pensamento a máquinas? Assim, o critério da pergunta é invertido: só podemos dizer que uma máquina pensa se ela for capaz de realizar as operações que constituem nosso pensamento. Além disso, quem responde afirmativamente à pergunta 3 acredita ser necessário recorrer a um tipo de psicologia para conhecer nosso pensamento. De fato, o estudo desta psicologia engloba um conjunto de disciplinas - linguística, psicologia, computação, entre outras –, que compõem o que se conhece como ciência cognitiva.

Por outro lado, nada garante que o pensamento humano é o único tipo de pensamento existente. Algo mais pode ser capaz de pensar, mas não como nós. Para defensores da IA-forte, computadores seriam algo assim e, portanto, o conhecimento do pensamento humano é dispensável na busca por uma teoria do pensamento. Para eles, podemos dizer que algo pensa quando é capaz de executar ações que, se executadas por humanos, requereriam pensamento (CRANE, 2003, p. 116)¹⁵. A uma possível réplica de que isto seria uma simulação de pensamento em vez da “coisa mesma”, teóricos da IA-forte respondem com o Teste de Turing. No famoso artigo “Computing machinery and intelligence” (*apud* BODEN, 1990), Turing elaborou este teste numa resposta à questão ‘Máquinas podem pensar?’, que achou melhor reformular como ‘Sobre que circunstâncias uma máquina seria confundida com uma pessoa pensante real?’. O teste consiste numa conversa entre uma pessoa com outra pessoa e com uma máquina. Se o sujeito experimental deste caso não consegue diferenciar entre ambas as conversas, podemos dizer que esta máquina pensa (CRANE, 2003, p. 116-117).

O Teste de Turing mostra como é possível imputar pensamento a uma máquina

uma mente é um programa”, enquanto a IA-fraca é definida como “a concepção de que processos cerebrais (e processos mentais) podem ser simulados computacionalmente” (SEARLE, 2006, p. 288). Nossa definição destes termos deve ser vista apenas como uma adaptação da distinção de Searle a nossos propósitos.

15 Não podemos deixar de pensar aqui que o pensamento humano, por mais que se afirme o contrário, não deixa de ser o critério para imputarmos pensamento a máquinas.

sem concebê-lo como concebemos nosso próprio pensamento. Naturalmente, se ele é ou não apropriado é uma questão em aberto, mas é inegável que este é seu objetivo. Empreendimentos em IA-forte podem trabalhar em revisões do teste de Turing ou em outros tipos de teste, mas certamente são encorajados por ele a não se basear necessariamente no pensamento humano para definir o pensamento em geral.

2.1 Teorias computacionais da mente

A aproximação de filósofos a teorias computacionais para compreender a mente deu-se, de início, num contexto de críticas a teorias que dominaram o cenário da Filosofia da Mente em meados do século XX, a saber, o behaviorismo lógico e a teoria da identidade de tipos. Estas teorias surgiram como respostas naturalistas ao dualismo dominante até então, soterrado por sua impossibilidade de solucionar satisfatoriamente o problema mente-corpo (FODOR, 1981a, p. 2-3). Em outras palavras, o dualismo não fornecia uma resposta adequada à pergunta: como é possível a interação causal entre um ente imaterial (a mente) e um ente material (o corpo)?¹⁶ Tal problema se coloca de forma diferente em teorias materialistas porque, nelas, a interação mente-corpo é compreendida como um tipo especial de interação causal – e toda interação causal, conforme defende o naturalismo, é caracterizada pela tese do fechamento causal do mundo físico, isto é, a tese de que “todos os efeitos físicos têm causas totalmente físicas” (PAPINEAU, 2016). Em teorias naturalistas, o problema reside em explicar que tipo de interação é essa e como ela se dá.

Acreditamos ser útil, como introdução teórica e histórica das teorias computacionais da mente, mostrar, ainda que brevemente, as limitações teóricas de suas antecessoras, a saber, o behaviorismo lógico e o funcionalismo, limitações estas que também serviram (ao menos em parte) como motivações para sua elaboração.

Uma das respostas materialistas ao problema mente-corpo, como já apontamos, foi o behaviorismo lógico. Para o behaviorismo lógico, atribuições de estados ou processos mentais a organismos são semanticamente equivalentes a atribuições de propriedades

16 O vocabulário pouco materialista em que formulamos a questão serve ao propósito de salientar que o problema se coloca mesmo para quem efetivamente acredita que a mente é uma substância distinta do corpo. O problema da interação mente-corpo não é apenas uma crítica externa de filósofos contrários ao dualismo; ele é um problema que surge internamente a esta doutrina.

disposicionais a eles (mais especificamente, disposições comportamentais) (FODOR, 1981a, p. 3). Um organismo possui uma disposição comportamental se, e somente se, satisfaz um conjunto indefinido (potencialmente infinito) de hipotéticos comportamentais, que compõem a análise da disposição. Hipotéticos comportamentais, por sua vez, são sentenças (condicionais)¹⁷ cujo antecedente é formulado em termos de parâmetros de estímulo e o consequente é formulado em termos de parâmetros de resposta (idem). Estes parâmetros, porém, carecem de definição na teoria behaviorista.

Para clarificar os termos acima, consideremos o seguinte exemplo: dado um certo organismo S, seu estado mental “S está com sede” corresponde, no behaviorismo lógico, ao que é expresso pela sentença “se houvesse água próximo a S, ele a beberia”¹⁸, uma disposição comportamental composta de apenas um hipotético comportamental. Seu parâmetro de estímulo é “haver água próximo a S” e seu parâmetro de resposta é “beber água”.

Uma vantagem do behaviorismo lógico, em especial por se tratar de uma teoria materialista, era incorporar, pelo menos em certo sentido, as explicações da psicologia popular, isto é, a ideia de que estados mentais causam o comportamento, já que, nele, estados mentais não passam de disposições comportamentais (FODOR, 1981a, p. 4).

Contudo, o behaviorismo lógico ignorou – e isto acabou por ser a sua derrocada filosófica – as relações entre estados mentais na causação de um mesmo comportamento. Muitas vezes, só é possível entender a referida relação causal como tal ao se considerar a relação dos estados e processos mentais entre si. Mas o behaviorismo não possuía uma explicação das sentenças que expressavam estas relações; nenhuma análise era possível de disposições comportamentais compostas de diversos (potencialmente infinitos) hipotéticos comportamentais¹⁹. Para citarmos um exemplo do problema em questão, considere uma análise do estado mental “ter uma dor de cabeça”, causa do comportamento “tomar uma aspirina”. Para entendermos a relação entre ambos como causal, isto é, para que este estado mental seja efetivamente a causa deste comportamento, ele deve vir acompanhado de uma série de outros estados mentais: o desejo de se livrar da dor de cabeça, a crença de que exista

17 Entre estas implicações estão inclusas sentenças contrafactuais.

18 Esta relação não é biunívoca. A mesma disposição comportamental pode comparecer na análise de outro estado mental – por exemplo, “S sente um gosto forte de pimenta”.

19 A própria incapacidade da teoria de definir em que casos o número de hipotéticos era infinito e em que casos ele não o era contribuiu para que ela fosse abandonada com base na constatação de que este problema era internamente insolúvel.

algo como remédios analgésicos, a crença de que eles reduzem a dor de cabeça, a crença de que seus efeitos colaterais não são piores que a dor de cabeça (que o remédio pretende fazer cessar ou, ao menos, reduzir), e assim por diante (FODOR, 1981a, p. 5). Nem todos estes estados mentais podem ser traduzidos em sentenças com vocabulário behaviorista (sem introdução de vocabulário mentalista, que é precisamente o que o behaviorismo procura evitar), mas certamente todas elas concorrem para a produção do comportamento de tomar aspirina.

A teoria da identidade de tipos surgiu como uma alternativa preferível ao behaviorismo lógico porque, nela, uma vez que tudo o que é mental (eventos, processos, estados etc.) é identificado com um estado, processo ou evento físico, não há nenhuma dificuldade teórica em explicar cadeias de eventos mentais concorrendo na causa de um único efeito comportamental, uma vez que a explicação da inter-relação entre eles se dá na mesma forma que a explicação da inter-relação causal entre processos físicos. De fato, uma vez que eles são processos físicos, esta é a única explicação que possuem e não há nada que impeça que um efeito físico tenha diversas causas, também físicas (idem, p. 6-7). Porém, a teoria da identidade de tipos cometia um grave erro: definia estados, eventos e processos físicos apenas em termos da fisiologia humana. Um exemplo que se tornou famoso foi a identidade “dor é o disparo de fibras-C”. Ora, seria então o caso de ser impossível que todo ser vivo natural, ou até mesmo máquinas, que não possuem tal tipo de estrutura nervosa não sejam capazes de sentir dor? Parece bastante implausível que a resposta a isso seja “sim”.

Em certo sentido, o funcionalismo é um aprimoramento de ambas as teorias apresentadas e um prosseguimento natural delas. Por um lado, ele preserva o caráter relacional do behaviorismo lógico, rejeitando porém a exigência de que estados mentais sejam explicados somente em termos comportamentais. Eles passam a ser explicados por seus papéis funcionais, isto é, por suas relações com *inputs* sensoriais, *outputs* motores e com outros estados mentais – ou seja, com sua organização funcional (RESCORLA, 2017). Isso desfaz o problema de explicar as relações entre diversos estados mentais na causação de um comportamento, uma vez que o problema consistia em reduzi-las a termos comportamentais, uma tarefa impossível. Por outro lado, o funcionalismo reconhece o caráter causal dos processos mentais, sem porém reduzir seu domínio à fisiologia humana (FODOR, 1981a, p. 9-11). A principal tese em favor deste ponto foi a Tese da Realizabilidade Múltipla do Mental,

defendida por Putnam (1975): assim como diversos tipos de hardware podem executar o mesmo software, a mente poderia sobrevir a diversos tipos de estruturas físicas²⁰.

Putnam especificou que tipo de explicação contava como uma explicação funcionalista de um processo mental: apenas aquelas que apelavam a (e somente a) 1) entradas da máquina, 2) saídas da máquina, 3) alguma de suas operações elementares, 4) outros estados de programa da máquina. Ou seja, “os mecanismos conceituais explorados em tais definições deveriam ser idênticos àqueles empregados na especificação dos estados de programa de (uma classe restrita de) computadores.” (FODOR, 1981a, p. 13). Assim, evitava-se o problema de confundir explicações funcionais com outros tipos de explicações e resolvia-se o problema de sua exequibilidade: sem especificarmos um mecanismo para elas, seria muito fácil apelar para papéis funcionais para explicar processos mentais, mas agora, uma vez que estados mentais estão especificados como estados de programa de uma Máquina de Turing, isto se torna bem mais simples, uma vez que esta, em virtude da trivialidade de execução de suas operações, pode ser facilmente reproduzida.

O funcionalismo, porém, é uma teoria insuficiente por uma série de motivos. Alguns deles só poderão ser melhor entendidos no contexto da exposição de certos temas, adiante, mas um deles pode ser abordado desde já. Apesar de afirmar que a mente tem uma estrutura causal, a teoria funcionalista não diz propriamente que esta estrutura é computacional, mas apenas que ela pode ser descrita como tal. Isso nos leva a um problema de atribuição de computacionalidade à mente: o modo como dizemos que a mente computa seus estados e processos é diferente do modo como poderíamos dizer que planetas, por exemplo, também computam as trajetórias que realizam? Ou seja, a atribuição de computacionalidade depende ou não de um atribuidor externo?²¹

Não apenas esta crítica como outras levaram pensadores, em especial Jerry Fodor, a proporem uma nova teoria, que ficou conhecida como Teoria Representacional da Mente ou Hipótese da Linguagem do Pensamento. A Hipótese da Linguagem do Pensamento (daqui em diante apenas LoTH, de *Language of Thought Hypothesis*) pode ser brevemente apresentada nos seguintes termos: a cada atitude proposicional com conteúdo entretida por um sujeito (por

20 Pode-se dizer que tanto o funcionalismo quanto outros tipos de Teorias Computacionais da Mente defendem um tipo de superveniência fraca, isto é, estados mentais sobrevivem a certas estruturas físicas, mas não apenas a elas.

21 Cf. o capítulo 3 de nossa dissertação para mais detalhes.

exemplo, a crença de Bernardo expressa pela sentença “Bernardo crê que Fortaleza é a capital do Ceará”), há, em sua cabeça, um conjunto de símbolos que se organizam de forma a representar esta sentença. Estes símbolos são palavras e se organizam em sentenças em uma linguagem (privada) do pensamento, que correspondem à sentença em linguagem natural (FODOR, 1981a). Este conjunto de símbolos e suas combinações é o que forma a LoT, ou “mentalês” (CRANE, 2003, p. 135; RESCORLA, 2017).

Não haveria nenhuma vantagem nas Teorias Computacionais da Mente sobre outras teorias se não ficasse claro em que sentido elas dizem que há símbolos ou até mesmo sentenças “na cabeça”. Poder-se-ia sempre objetar que, uma vez que não está claro como isso acontece, teorias concorrentes (teoria da identidade de tipos, behaviorismo etc.) são preferíveis a teorias computacionais. Mas teóricos computacionais da mente têm uma noção muito clara de como isto acontece.

É necessário esclarecer o sentido de alguns termos que serão usados no decorrer da explicação deste processo. Os primeiros são os conceitos de tipo (*type*) e de exemplar (*token*) (CRANE, 2003, p. 136). Acreditamos que a melhor maneira de apresentar esta distinção é através de um exemplo. Esta dissertação, por exemplo, possui pelo menos dois exemplares: um em meu computador e outro que você, leitor, lê neste momento. Porém, ela possui apenas um tipo. Outro exemplo se aplica a sentenças: quando um comandante, por exemplo, exorta seus soldados gritando repetidamente “Avante! Avante! Avante!”, temos três exemplares de um mesmo tipo de palavra.

Outra distinção importante é a distinção entre meio e veículo. Utilizando-se da distinção anterior, podemos dizer que as diferentes formas de armazenar os diferentes exemplares de um mesmo tipo de sentença constituem o meio no qual ela é transmitida. Por outro lado, quando o próprio tipo da mensagem muda e o que permanece entre suas diferentes manifestações é apenas seu conteúdo (ou seja, ela é realizada de modo diferente), estamos falando de seu veículo de representação (*idem*). Por exemplo: a sentença “Não fume” pode ser representada por diversos meios: avisos sonoros, palavras escritas no verso de carteiras de cigarro, avisos falados na TV, etc. Ela também pode ser representada por um veículo pictórico (um círculo vermelho dividido diagonalmente em dois ao redor da representação visual de um cigarro) em locais onde fumar é proibido, além do veículo linguístico (a sentença “Não fume”) apresentado nos exemplos anteriores.

De posse destes conceitos, fica mais fácil falar de “símbolos na cabeça”, como fizemos anteriormente, e formular de maneira mais precisa a LoTH. Isto pode ser feito da seguinte maneira:

... quando um pensador tem uma crença ou desejo [ou um pensamento, poderíamos acrescentar]²² com o conteúdo P, há escrito em sua cabeça uma sentença (isto é, uma representação com estrutura sintática e semântica) que significa P. Os veículos de representação são linguísticos [sentenças da LoT (idem, p. 137)], enquanto o meio de representação é a estrutura neural do cérebro. (CRANE, 2003, p. 140).

Assim como portas lógicas, em um computador, espelham as propriedades lógicas do código que ele executa, a estrutura neuronal humana espelha nosso funcionamento mental, que é, segundo Fodor, “...computação estilo Turing sobre a linguagem do pensamento.” (RESCORLA, 2017).

Esta é a formulação da Teoria Representacional da Mente, ou LoTH. A mera existência e formulação da teoria, porém, não serve para sustentá-la. Uma teoria não se sustenta por si mesma; são também necessários argumentos para tal. O argumento de Fodor a favor de sua teoria pode ser resumido nos seguintes termos: a existência de uma LoT é a melhor forma de explicar a presença de características sintáticas e semânticas da mente²³. Ora, se a mente possui uma sintaxe e uma semântica, possui uma linguagem, sobre a qual opera. A sistematicidade da mente nos permite falar sobre sua sintaxe; sua composicionalidade e produtividade, por sua vez, sobre sua semântica. Para compreender melhor este argumento, porém, precisamos entender do que tratam estes termos.

22 Preservamos a ambiguidade do termo ‘pensamento’ presente no livro de Crane, mas por motivos, ao menos em parte, diferentes dos dele. ‘Pensamento’, aqui, é uma categoria ampla, que engloba toda e qualquer atitude proposicional: crenças, desejos, esperanças, medos etc. Isso porque, segundo Crane, todos esses estados são representacionais: “... toda esperança, crença ou desejo é dirigida a algo. Se você espera, você deve esperar por algo; se acredita, deve acreditar em algo; se deseja, deve desejar algo.” (2003, p. 23). Uma vez que o motivo para mantermos este termo aqui, embora ambíguo, se relaciona com a adoção ou não do representacionalismo (tese que apresentaremos adiante), deixaremos em aberto nossos motivos para preservá-lo, mas de antemão é possível notar que manter sua ambiguidade é interessante por, com ela, ser possível assumir a LoTH como verdadeira sem assumir, propriamente, que ela é uma teoria do pensamento no sentido amplo assumido por Crane. Se esta categoria compreende todas as atitudes proposicionais relevantes, não vemos problema em se afirmar que a Teoria Representacional da Mente, ou LoTH, é uma teoria do pensamento, porque isso significaria o mesmo que dizer que ela é uma teoria da mente. Mas aí o sentido de diferenciar os dois termos se perde. De qualquer forma, manter o termo ‘pensamento’ significa que não é necessária uma revisão terminológica seja qual for o resultado desta querela.

23 Efetivamente, à época de Fodor a Teoria Representacional da Mente era a *única* teoria da mente capaz de explicar estes fenômenos, o que fez Fodor citar a famosa frase de London B. Johnson (vice de John Kennedy) em relação à sua teoria: “eu sou o único presidente que vocês têm”. Certamente isto não é mais o caso, mas ainda se debate se a teoria de Fodor é a melhor entre as existentes.

A composicionalidade se refere à seguinte característica de certas linguagens: enquanto termos individuais adquirem seu sentido por referirem-se a objetos no mundo, sentenças adquirem seu sentido através dos termos que as compõem e de seu modo de organização (CRANE, 2003, p. 138-139). Ou seja, a composicionalidade mostra como os termos e sentenças de uma linguagem adquirem seu sentido; em uma palavra, trata de sua semântica. A produtividade é um corolário da composicionalidade: uma vez que o sentido de uma sentença é determinado por seus componentes e seu modo de organização, a partir da compreensão de um número finito de termos individuais somos capazes de formar e entender um número infinito de sentenças formadas a partir deles (CRANE, 2003, p. 140; RESCORLA, 2017). Por fim, a sistematicidade refere-se à capacidade de se inferir sentenças a partir de outras sentenças. Isto se dá por meio de uma sintaxe, que determina quais sentenças da linguagem são bem formadas e como se pode inferir outras sentenças a partir de certo conjunto de sentenças previamente dadas.

Alguns exemplos certamente ajudarão a compreender estes conceitos. Tomemos como exemplo a sentença “Romeu ama Julieta”. Se invertemos seus termos e formamos a sentença “Julieta ama Romeu”, o resultado é uma sentença com sentido diferente, mesmo que ambas as sentenças contenham os mesmos termos. O responsável, portanto, por sua diferença de sentido é sua ordem. Por outro lado, por mais que sejam diferentes entre si, nada a mais é requerido para compreender a segunda do que aquilo que é requerido para compreender a primeira.

Como dissemos, o argumento de Fodor se sustenta em duas ideias: 1) o pensamento possui uma sintaxe e uma semântica, o que é evidenciado por suas características de composicionalidade, produtividade e sistematicidade; 2) a melhor explicação para isso é postular uma LoT. Começemos pela primeira ideia: o que significa dizer o pensamento possui essas características? Significa dizer que:

- a) Pensamentos têm propriedades semânticas. Eles significam algo por referirem-se a coisas no mundo, por estarem no lugar de coisas do mundo ou, simplesmente, por representarem objetos e estados de coisas (CRANE, 2003, p. 140);
- b) O significado dos pensamentos é determinado por seus componentes e por seu

modo de organização (idem). Por exemplo: para pensar que Romeu ama Julieta, eu não preciso pensar em mais nada além de Romeu, Julieta, e na relação de amar. O conteúdo (ou significado) do meu pensamento é determinado por seus termos componentes e seu modo de organização (para ver isso, basta notar que ‘Julieta ama Romeu’ é um pensamento diferente, mas com os mesmos componentes);

- c) Um sujeito pensante pode pensar uma quantidade indeterminada de coisas (ter um número infinito de pensamentos) a partir de um número finito de representações.

A partir dos pontos acima, vemos que o pensamento também possui as propriedades de composicionalidade e produtividade. A partir daí, podemos inferir²⁴, conforme a ideia 2 acima, que a melhor explicação para isso é a existência de uma LoT. Mas há uma propriedade cuja constatação requer um pouco mais de cuidado: a sistematicidade.

Devemos ter claro o que significa, propriamente, dizer que o pensamento é sistemático. Para isso, é interessante compará-lo com um exemplo de pensamento assistemático, assim como mostramos o que é composicionalidade a partir da comparação de uma linguagem composicional com uma linguagem não-composicional. Considere, portanto, um exemplo de pensamento assistemático:

... e os gloriosos crepúsculos e as figuras nos jardins da Alameda sim e todas as ruazinhas estranhas e as casas rosa e azuis e amarelas e os jardins-de-rosas e os jasmims e os gerânios e cactos e Gibraltar quando eu era uma mocinha onde eu era uma Flor da montanha sim quando eu pus uma rosa no meu cabelo como as moças andaluzas usavam ou será que eu vou usar uma vermelha sim e como ele me beijou debaixo do muro mouresco e eu pensei bem tanto faz ele como um outro e então eu lhe pedi com meus olhos que pedisse novamente sim e então ele me pediu se eu queria sim dizer sim minha flor da montanha e primeiro eu pus meus braços à sua volta sim e o arrastei para baixo sobre mim para que ele pudesse sentir meus seios todos perfume sim e seu coração disparou como louco e sim eu disse sim eu quero Sim. (JOYCE, 2005, p. 814-815)

Compare-o com este exemplo de pensamento sistemático:

O único mérito que posso reivindicar é o de ter reconsiderado minha posição

24 Perceba que esta é uma inferência para a melhor explicação, não uma dedução propriamente dita. De fato, a principal concorrente da teoria de Fodor, a Teoria Computacional Conexionista da Mente (que abordaremos brevemente adiante), se contrapõe a ela precisamente alegando ser uma melhor explicação para estes fenômenos. Veremos também adiante porque eu, seguindo Crane (2003), acredito que isto seja falso.

imediatamente quando... ficou claro para mim que o perigo que teria podido ameaçar um ocupante daquele quarto, fosse qual fosse, não poderia ter entrado nem pela janela nem pela porta. Minha atenção foi rapidamente atraída... para aquele vão e para a corda de campainha que descia até a cama. A descoberta de que ela era uma simulação, e de que a cama estava pregada no assoalho, fez surgir instantaneamente a suspeita de que a corda estava ali como uma ponte, para permitir que alguma coisa passasse pelo buraco e chegasse à cama. A ideia de uma cobra ocorreu-me de imediato, e, quando juntei a ela a informação de que o médico era abastecido com animais da Índia, tive a impressão de que provavelmente estava na pista certa. (DOYLE, 2011, p. 272)²⁵.

A decisão de Molly Bloom de aceitar Leopold Bloom como seu namorado, exposta na primeira citação acima, não é sistemática, em especial em sua primeira parte²⁶. Molly decide com base em percepções do ambiente (crepúsculos, ruas e jardins de Gibraltar) e em acontecimentos (o beijo sob o muro mouresco, o repetido pedido de Leopold), mas não em raciocínio (talvez tenha decidido com base em associação livre, por exemplo)²⁷. Por outro lado, as ações de Sherlock Holmes são claramente guiadas por raciocínios e inferências: Sherlock crê que, se o perigo que ameaçava o ocupante do quarto não poderia ter entrado pela janela nem pela porta, então deve ter entrado pelo vão entre os dois quartos; o perigo não poderia ter entrado pela janela nem pela porta; logo, o detetive acredita que ele entrou pelo vão (uma instância de um *modus ponens*); Sherlock acredita que o perigo que ameaçava o ocupante do quarto era uma cobra; o médico (que dormia no quarto ao lado) tinha um suprimento de animais da Índia; logo, o detetive é levado a pensar que o médico é o responsável pelo assassinato da ocupante do quarto contíguo²⁸ (FODOR, 1990, p. 21). Podemos dizer, portanto, que o pensamento de Molly Bloom não é sistemático, enquanto o de Sherlock Holmes o é. É sistemático porque possui propriedades sintáticas, assim como a linguagem. As transições entre nossos estados mentais se assemelham a inferências lógicas, isto é, às transições entre sentenças de uma linguagem, descritas por sua sintaxe.

É hora de retomar a exposição e acrescentar alguns pormenores. A propriedade compartilhada entre linguagem e pensamento, no caso de realizarmos uma inferência lógica, é a preservação da verdade. Dizemos que inferências lógicas preservam a verdade porque nelas

25 Exemplo retirado de FODOR, 1990.

26 Os repetidos sins da personagem podem fazer o leitor pensar que ela já havia tomado sua decisão e que estava realmente decidindo entre acatá-la ou não. Não cremos, porém, que isto prejudique o exemplo.

27 Conforme detalharemos adiante, a Teoria Representacional da Mente convive pacificamente com a ideia de que nem todos os estados e processos mentais são sistemáticos.

28 Neste segundo exemplo, naturalmente, trata-se de uma inferência para a melhor explicação e não de um argumento dedutivo. Mas pensamento sistemático não é sinônimo (estritamente) de inferência dedutiva.

é garantido que, se partimos de premissas verdadeiras, chegamos a conclusões também verdadeiras (quando isto ocorre, temos um argumento válido)²⁹. Ora, esta é uma propriedade sintática, presente no pensamento apenas por ele ser sistemático. Isto aponta para o fato de que ele possui uma sintaxe.

Um último argumento a favor da existência de uma linguagem do pensamento é retirado de uma característica do aprendizado linguístico: não é possível aprender uma linguagem sem possuir uma linguagem na qual representá-la. Fodor sustenta esta característica com o seguinte argumento:

Aprender uma linguagem (incluindo, é claro, uma primeira linguagem) envolve aprender o que os predicados da linguagem significam. Aprender o que os predicados de uma linguagem significam envolve aprender um modo de determinar a extensão desses predicados. Aprender um modo de determinar a extensão dos predicados envolve aprender que eles caem sob certas regras (i.e., regras de verdade). Mas ninguém pode aprender que P cai sob R a menos que tenha uma linguagem na qual P e R possam ser representados. Assim, ninguém aprende uma linguagem a menos que tenha uma linguagem. (FODOR, 1975, p. 63-64)

A defesa da linguagem do pensamento se torna clara quando pensamos como o aprendizado da primeira língua é realizado se o aprendizado linguístico realmente tem esta característica, isto é, se todo aprendizado de uma linguagem requer uma linguagem prévia. Se isto realmente é o caso, só poderíamos aprender nossa primeira língua caso possuíssemos, de antemão, uma linguagem na qual representá-la. Como não temos nenhum aprendizado linguístico antes disso, a linguagem na qual nossa língua materna é representada só pode ser uma linguagem inata. Esta linguagem inata é a linguagem do pensamento.

Todas essas considerações apontam para o fato de que o pensamento possui uma sintaxe e uma semântica. Possuir ambas é o requisito para considerar um conjunto de símbolos uma linguagem. Podemos dizer, portanto, que há algo como uma linguagem do pensamento.

Podemos agora apresentar o segundo motivo pelo qual a LoTH é preferível ao funcionalismo. As características da mente que nos permitem afirmar que ela é computacional – como sua produtividade e sistematicidade, por exemplo –, não obstante ser possível questionar se elas realmente fornecem o elo entre o *explanandum* da mente e o *explanans*

²⁹ Naturalmente, nem sempre isso ocorre. É o que acontece nas falácias. Mas falácias, lembremos, são argumentos que parecem válidos, mas não o são.

computacional, são efetivamente propriedades mentais e, como tais, demandam explicação. Ora, o funcionalismo não fornece uma tal explicação. Uma vez que a Teoria Representacional da Mente o faz, está certamente em vantagem em relação a ele, ao menos neste aspecto.

A partir das conclusões acima, surge uma dúvida: o que, exatamente, se quer dizer com ‘pensamento’? Isso retoma a ambiguidade que apresentamos anteriormente sobre o termo. Esta ambiguidade tem raízes na aceitação ou rejeição do representacionalismo, isto é, a ideia de que todo e qualquer estado mental é representacional ou, em outros termos, intencional. Um estado mental é intencional quando tem direcionalidade, ou, em outras palavras, quando é sobre algo (por exemplo, uma crença é sempre uma crença em *algo*, um desejo é sempre um desejo *de algo* etc.), e representacional quando ocupa o lugar de algo ou está no lugar de algo.

Portanto, uma representação, por definição, é intencional: aquilo que ela representa é seu objeto intencional. Ora, se o representacionalismo é verdadeiro, segue-se que todo estado mental seria também intencional³⁰. Se a LoTH é uma teoria que explica o pensamento por meio de representações, então ela é, de fato, uma teoria da mente, pois não haveria nada mais para ela explicar – uma vez que não haveria nada de não-representacional na mente. Por outro lado, se nem todo estado mental é representacional, não se pode dizer que a LoTH é uma teoria exaustiva da mente, mas é possível preservar parte de seu poder explicativo afirmando que ela explica uma parte relevante de seu funcionamento, a saber, o pensamento. Ou seja, a LoTH seria capaz de explicar a parcela representacional da mente, permanecendo neutra sobre a parcela não-representacional. Ela seria uma teoria do pensamento, ou, para usar o termo corrente nos dias atuais, da cognição.

Como fica, porém, o debate sobre a tese representacionista? Uma das pedras no sapato do representacionalismo são os *qualia*: estados mentais sensoriais (visuais, auditivos, olfativos etc.). É razoável pensar que eles não são estados representacionais e há quem argumente a favor disso. O próprio Fodor previu dificuldades, na LoTH, em lidar com os *qualia*. Ele afirmou:

Parece-me que... de fato, não temos nenhuma explicação materialista adequada de conteúdos qualitativos. (...) nossa perplexidade é tão profunda que não podemos sequer dizer precisamente o que é isto que gostaríamos que uma teoria explicasse

³⁰ Essa conclusão constitui parte do que, em Filosofia, ficou conhecido como Tese de Brentano (LECLERC, 2018).

(FODOR, 1988, p. 17-18).

Como encaminhamento à resolução deste problema, ou aceitamos que a LoTH tem essa restrição e é uma teoria da cognição (em vez de uma teoria da mente) ou mostramos como *qualia* podem ter propriedades representacionais.

3 O QUÊ, O PORQUÊ E O COMO DA INFORMAÇÃO MENTAL

Nosso principal objetivo nesta dissertação é encontrar um conceito de informação para a mente que seja tão preciso quanto aquele aplicado na teoria da computação³¹. Como podemos, porém, chegar a ele? O capítulo anterior prepara nossa estratégia de resposta a essa questão no seguinte sentido: nele, introduzimos os conceitos fundamentais de teorias computacionais da mente e fizemos uma breve defesa de uma delas, a Teoria Representacional da Mente ou Hipótese da Linguagem do Pensamento. Portanto, visamos, neste capítulo, formular uma definição de informação que se aplique a esta teoria. Mas, uma vez que é impossível falar de informação sem falar de seu processamento, ou seja, de computação, devemos também mostrar como, em seu âmbito, é possível dizer que a mente processa informação. Além disso, também pretendemos defender que a informação envolvida nestes processos, isto é, a informação mental, tem caráter semântico. Em poucas palavras, defendemos que esta informação é semântica porque envolve representações, entendidas aqui como estados mentais intencionais (isto é, com capacidade de referir-se a estados de coisas). Os três objetivos apresentados aqui se referem aos termos interrogativos do título deste capítulo. O quê da informação mental é seu conceito ou sua definição; o porquê se refere à nossa defesa de seu caráter semântico; por fim, o como da informação mental é a explicação de seu processamento, ou seja, de como a mente computa. Antes de partir às nossas respostas a estas questões, porém, parece-nos necessário apresentar brevemente algumas respostas ao problema que contribuirão na formulação de nossa própria proposta.

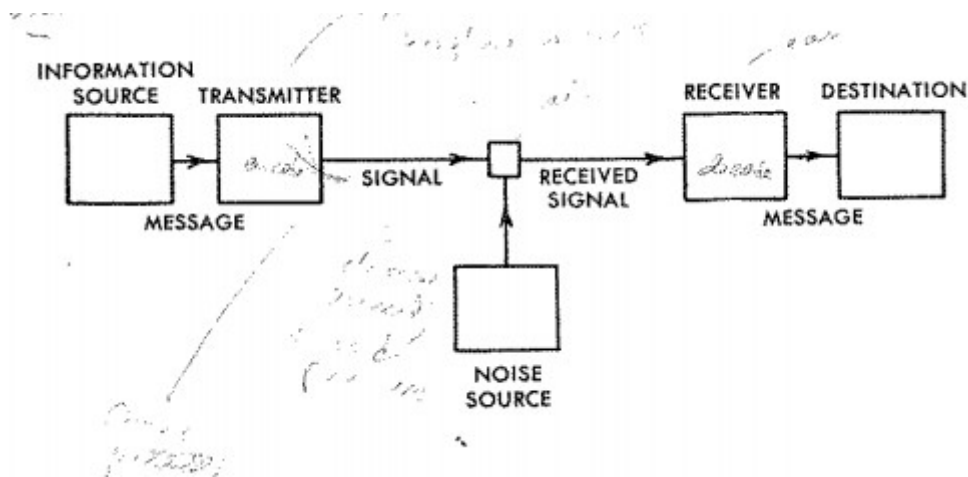
3.1 Aplicações do conceito de informação à mente

3.1.1 *Informação de Shannon*

31 É bastante problemático oferecer um exemplo de um conceito (totalmente) preciso, seja ele filosófico ou não. Portanto, o “tão” presente aqui deve ser lido num sentido meramente comparativo: qualquer que seja o grau de precisão do conceito de computação utilizado na teoria da computação, é dele que queremos nos aproximar. Isto porque, comparativamente, quaisquer que sejam as dificuldades presentes no conceito de computação desta área, elas parecem ser menores e em menor número que aqueles presentes no conceito de computação usado pela ciência cognitiva. É deste fenômeno que partimos em nossa investigação. Naturalmente, o conceito de computação simpliciter não é tão consensual assim quanto talvez fazemos parecer aqui, o que traz consequências interessantes para a tese de que a mente/cérebro é um computador. Este debate será aprofundado adiante.

Não há como evitar iniciar pela teoria canônica da informação (ou, ao menos, seu tratamento matemático mais difundido e aceito), a teoria matemática da comunicação desenvolvida por Claude Shannon (SHANNON, 1948; SHANNON; WEAVER, 1972). Para Shannon, informação é algo que ocorre em um sistema de comunicação. Tal sistema é composto de uma fonte de informação - “a pessoa ou a máquina geradora da mensagem” (GLEICK, 2013, p. 31) –, de um transmissor – responsável por codificar a mensagem para produzir o sinal –, um canal - “o meio utilizado para transmitir o sinal” (idem) –, o receptor – responsável por decodificar a mensagem do transmissor – e o destinatário (idem). Ele pode ser basicamente descrito nos seguintes termos: a fonte de informação seleciona uma mensagem a partir de um conjunto de mensagens possíveis (SHANNON; WEAVER, 1972, p. 7). Esta mensagem é enviada ao transmissor e é codificada em um sinal enviado através de um canal de comunicação ao receptor, que a decodifica em uma mensagem e a envia ao destinatário. Em uma transmissão ideal, a mensagem emitida seria a mesma mensagem recebida, sem quaisquer elementos extrínsecos, mas, empiricamente, constata-se que elementos indesejados à fonte de informação sempre são inseridos no sinal. Em conjunto, estes elementos são denominados ruído (idem, p. 7-8).

Figura 1 - O sistema de comunicação de Shannon



Fonte: SHANNON; WEAVER (1972)

Shannon não formulou um conceito explícito de informação, uma vez que seu principal objetivo era o de quantificar sua transmissão através de sistemas de comunicação. Entretanto, podemos depreender de sua teoria que informação é a liberdade de escolha de uma fonte em selecionar uma mensagem (idem, p. 9). Em outras palavras, o conceito de

informação que podemos depreender da teoria matemática da comunicação é o de redução de possibilidades. A situação mínima que atende a estes requisitos é aquela em que uma fonte de informação tem de escolher entre duas mensagens; neste caso, ela gera uma unidade de informação, denominada bit (abreviação de *binary digit*). Como fica claro pelas nossas considerações até aqui, esta concepção de informação se afasta consideravelmente da concepção de senso comum. De maneira geral, consideramos que algo é informativo quando lhe podemos extrair sentido a partir de alguma das nossas faculdades cognitivas (o exemplo mais ilustrativo disto, talvez, seja o jornal, do qual dizemos que nos informa sobre os acontecimentos). Implicitamente, no senso comum a informação é considerada como algo subjetivo, dependente de seres humanos. Por outro lado, para torná-la objeto de estudo, Shannon precisou depurar tais elementos subjetivos da informação, considerando-a naquilo que ela possui de objetivo. Ao fazê-lo, ele manteve algumas das intuições presentes na noção popular de informação, mas trouxe à tona outras relações insuspeitas até então. Um ponto central de sua noção de informação é a relação entre informação e incerteza. De maneira geral, dizemos que um evento é informativo quando a fonte escolhe a mensagem a ser transmitida entre diversas outras mensagens possíveis; a incerteza (e, por sua vez, a quantidade de informação) é proporcional ao número de alternativas que são dadas para a escolha da fonte.

Dizíamos que a informação, tal como entendida por Shannon, não tem a ver com significado, ou seja, não tem caráter semântico. Uma maneira de mostrá-lo é através da comunicação entre um casal fictício, Alice e Bob. De acordo com Shannon, era indiferente que Alice, ao fazer uma ligação para Bob, dissesse “eu te adoro” ou “eu te odeio”, uma vez que ambas as mensagens poderiam ser codificadas com a mesma quantidade de bits (idem, p. 372; VEDRAL, 2010, p. 31). Segundo Weaver (SHANNON; WEAVER, 1972, p. 4), Shannon se preocupou mais detidamente com o chamado nível A dos problemas de comunicação, que tenta responder à pergunta: “quão precisamente os símbolos de comunicação podem ser transmitidos?”. Esta questão é também chamada de “o problema técnico”. Por outro lado, o nível dos problemas de comunicação que está mais diretamente associado a aspectos semânticos da informação é o nível B, aquele que procura responder ao chamado “problema semântico”: “quão precisamente os símbolos transmitidos carregam o sentido desejado?” (idem). Para Shannon, porém, “os aspectos semânticos da comunicação são irrelevantes aos aspectos de engenharia.” (idem, p. 8). Ou seja: informação, na teoria matemática da

comunicação, é um termo não intencional, isto é, não orientado a sentido (BODEN, 2008, p. 743-744).

Como a teoria de Shannon foi aplicada ao estudo da mente? Isso foi efetivamente realizado nos primórdios da ciência cognitiva, que, então, foi “incitada pela teoria da informação [de Shannon], pela noção dos ciberneticistas de *feedback* e por ideias sobre modelos (analógicos) no cérebro” (BODEN, 2008, p. 742). O principal atrativo da aplicação da teoria da informação de Shannon à mente era tanto comparar quanto medir as capacidades e limitações do processamento de informação da mente/cérebro, assim como se utilizar da noção de redução de diferenças para iluminar aspectos do comportamento com propósito e para entender como ocorre a codificação e transmissão de informação em agentes cognitivos (idem, p. 743). Assim, embora tivesse alguma promessa de explicação de como ocorrem nossos processos cognitivos, a teoria de Shannon tinha apenas uma parcela da explicação do caráter da informação envolvida neles. Era difícil ver como se poderia realizar a transição entre o contexto não intencional da informação na teoria de Shannon e o contexto intencional da informação em agentes cognitivos humanos.

Conforme veremos adiante, alguns destes prognósticos podem – e, a nosso ver, devem – ser revistos. Acreditamos que, com acréscimos, tanto o conceito de informação de Shannon ainda é um forte candidato a explicar o quê da informação mental quanto a teoria matemática da comunicação ainda é uma forte candidata a explicar seu como.

3.1.2 Informação representacional

O que nos impede de aplicar diretamente a informação de Shannon à compreensão da informação é sua ausência de caráter semântico. Portanto, o próximo passo em nossa investigação é mostrar de onde a informação mental retira tal caráter. Intuitivamente, a resposta inicial a esta questão parece estar na consideração de que a mente é representacional, ou seja, um mecanismo que opera sobre representações³², uma vez que se acredita que estas entidades sejam precisamente as responsáveis por nos informar sobre o mundo. Mas alguma consideração de como elas fazem isso – ou mesmo se são capazes de tanto – é necessária para sustentar esta intuição.

32 Daqui em diante, usaremos simplesmente ‘representações’ para nos referirmos a representações mentais.

Não é nosso objetivo entrar no debate sobre representações, mas uma das candidatas parece nos mostrar um caminho a partir do qual elas podem ser consideradas informativas³³. Trata-se da teoria causal da representação. Para apresentá-la propriamente, é necessária uma breve digressão para introduzir alguns conceitos necessários à compreensão da proposta, como o de indicação. Este conceito tem origem na teoria do signo de Charles Sanders Peirce, a qual abordaremos brevemente³⁴.

Para Peirce, significar é algo que fazemos conforme nosso interesse em coisas; é um interesse mediado em algo com o objetivo de trazer uma ideia à mente. O que traz a ideia à mente nesta mediação é o signo (também chamado de representação) (PEIRCE, *apud* MARCONDES, 2010, p. 73). Há três maneiras de significar: 1) por semelhança, em que a ideia é trazida à mente por o signo imitar a coisa que representa; neste caso, o signo é um ícone. Um exemplo de ícone é uma fotografia (em relação à pessoa fotografada); 2) por indicação, em que a ideia é trazida à mente através da existência de uma relação física entre o signo e a coisa representada; neste caso, o signo é um índice. Sintomas de doença são bons exemplos de índices, pois há uma relação causal entre seu aparecimento e a presença de doença em um organismo. Por exemplo: pontos vermelhos na pele são sintomas de sarampo (indicam que alguém está com esta doença) porque há uma relação causal entre elas e a doença: é o sarampo que causa os pontos vermelhos na pele; 3) por denotação, em que a ideia é trazida à mente por existir uma relação convencional (estabelecida pelo uso da linguagem) entre o signo e a coisa representada; neste caso, o signo se chama símbolo. A relação entre a luz vermelha de um semáforo e a ação de parar por parte dos motoristas é deste tipo. Destes, os tipos de significação e signo que nos interessam mais de perto são a indicação e o índice,

33 Um esclarecimento: conforme podemos depreender a partir de Pitt (2018), a existência da relação entre representação e informação não é colocada em questão nos âmbitos da Ciência Cognitiva e das Teorias Computacionais da Mente, uma vez que elas as entendem como “estruturas que carregam informação (*information-bearing structures*)”. Sobre como esta relação se dá, porém, há muita controvérsia (cf. *op. cit.*). O tratamento desta relação que fazemos nesta seção e na seguinte, a teoria causal da representação, não está livre destas controvérsias; na seção posterior a estas, procuramos, no contexto de comparação de nossa teoria com outras teorias da informação mental, rebater algumas delas, mas tratar do tema com maior profundidade foge ao escopo do nosso projeto.

34 Nossa exposição das teorias do sentido de Peirce e Grice (adiante) toma algumas liberdades para melhor aplicá-las aos propósitos que temos em mente. Dadas possíveis diferenças entre estas teorias e o que expusemos aqui, deve-se preferir, para os propósitos de leitura e compreensão deste texto, a versão exposta por nós. Não se trata de presunção, mas apenas da constatação do fato de que, aqui, não pretendemos fazer uma exegese precisa delas, mas aproveitá-las no que são capazes de esclarecer os fenômenos dos quais tratamos. Por outro lado, tampouco queremos dar a entender, com seu uso, que os autores endossavam uma abordagem mecanicista da mente. Peirce, por exemplo, se opõe radicalmente a concepções mecanicistas da evolução e também da cognição. Sua abordagem semiótica expande também o universo mecanicista das representações. Agradeço à prof^a. dr^a. Maria Eunice Gonzalez por haver me apontado isto.

respectivamente.

A teoria peirceana do signo foi posteriormente desenvolvida por Grice (1991) em uma teoria do sentido, aplicada a sentenças. Grice distinguia entre dois tipos de sentido, o natural e o não-natural. Podemos retomar uma das sentenças mencionadas acima como exemplo de sentença com sentido natural, com uma leve alteração:

(1) Pontos vermelhos na pele significam sarampo.

A alteração consiste na substituição do termo ‘indicam’ (ou da locução ‘são sintomas de’) por ‘significam’. É trivial dizer que esta substituição não traz prejuízo, uma vez que indicar é uma forma de significar. Chamarei de indicativa a este tipo de sentença. Compare-a com a seguinte sentença, também retomada e adaptada de um exemplo anterior:

(2) A luz vermelha no semáforo significa ‘pare’.

Chamarei de denotativa a este tipo de sentença.

As novidades trazidas por Grice à concepção peirceana do signo que nos interessam mais de perto são duas: 1) sua constatação de que, de uma sentença indicativa *p*, a expressão “*S* quer dizer que *p*” implica *p*, enquanto o mesmo não pode ser dito de sentenças denotativas (1991, p. 213-214); 2) se, na expressão “*S* quer dizer que *p*”, *S* indica *p*, a expressão possui sentido natural, isto é, “quer dizer” deve ser lido em sentido natural. Se *S* denota *p*, a expressão possui sentido não-natural (idem, p. 214)³⁵.

Em que sentido as noções de indicação e de sentido natural são necessárias à compreensão da teoria causal da representação? Se fornecer uma teoria da representação é o mesmo que fornecer condições necessárias e suficientes para que algo represente algo (formalmente, para que *x* represente *y*), então, na teoria causal da representação, estas condições estão na noção de indicação ou sentido natural. Em outras palavras, a teoria causal da representação diz que *x* representa *y* se, e somente se, *x* indica (significa naturalmente) *y* (CRANE, 2003, p. 177).

35 Como dissemos acima, nossa exposição da teoria do sentido de Grice toma algumas liberdades para melhor aplicá-la aos propósitos que temos em mente e, além disso, acomodá-la à teoria peirceana do sentido. Por exemplo, existem diversos outros critérios para as noções de sentido natural e não-natural, por um lado, e, por outro, diversas outras aplicações.

Vemos esta teoria como uma forte candidata a mostrar como representações mentais podem ser informativas porque as relações de indicação que a sustentam também são informativas. Intuitivamente, por exemplo, pode-se dizer que sou informado de que alguém está com sarampo ao ver (a indicação de) que está com pontos vermelhos na pele. Poderíamos também estabelecer a relação entre representação e informação a partir do seguinte argumento:

- a) x representa y sse x indica y (TCR);
- b) Se x indica y, então x informa sobre y;
- c) Se x representa y, então x indica y (DD 1)³⁶;

Conclusão: Se x representa y, então x informa sobre y³⁷ (TI 3, 2)³⁸.

Este tipo de informação deve ser semântico, pois possui sentido, isto é, se refere a estados de coisas no mundo (é sobre coisas³⁹): minha informação de que alguém possui sarampo (a partir dos pontos vermelhos em sua pele) se refere à doença que esta pessoa efetivamente possui, sarampo. Transfere-se também a ela o caráter factivo presente na significação natural e na representação: se, segundo estes critérios, tenho informação sobre um estado-de-coisas descrito por uma proposição p, então p é o caso (PICCININI; SCARANTINO, 2010, p. 241-242). Através disto, justifica-se também nossa opção pela teoria causal da representação: pelos motivos acima, acreditamos que esta teoria da representação é a que melhor explicita o caráter informativo das representações, a intuitividade deste caráter e a relação entre a informação mental e sua fonte, isto é, o mundo percebido (ou seja, seu caráter semântico).

Entretanto, o argumento acima permanece frágil enquanto não oferecermos uma melhor defesa de suas premissas, em especial a premissa 2. Apesar de parecer intuitivo que relações de significação natural sejam informativas, para estabelecermos isto devemos apontar como e por que representações possuem informatividade. É o que faremos na seção seguinte.

36 Abreviação da regra de inferência de definição da dupla implicação.

37 Esta definição pode ser acusada de petição de princípio na medida em que, quer S seja uma variável para sujeito ou para sistema cognitivo, ela pretende explicar a mente introduzindo, na explicação, a própria mente como critério. Mas, por outro lado, talvez esta preocupação seja um tanto exagerada. Podemos pensar que se, por um lado, a informação mental está sempre atrelada a representações, por outro sua manipulação é algo que ocorre em pelo menos um nível acima do nível básico, isto é, do processamento de representações estilo-Turing. Não estaríamos, assim, introduzindo um homúnculo mais básico que este nível para encaixar a informação na explicação canônica da computação mental. Sobre o problema do homúnculo, cf. adiante, cap. 4 e Conclusão.

38 Abreviação da regra de inferência de transitividade da implicação.

39 Estabeleceremos melhor este caráter de aboutness da informação mental adiante.

3.1.3 Informação mental

A aplicação do sistema comunicacional de Shannon à mente nos parece possível, mas peca como explicação na medida em que deixa de lado elementos semânticos. Isso, porém, não é obstáculo a uma incrementação ao sistema que os inclua. Aqui, pretendemos mostrar ser possível integrar uma compreensão semântica da informação mental à aplicação do sistema comunicacional de Shannon à mente a partir da Teoria Representacional da Mente, compreendendo representações tal como a teoria causal da representação as define.

A informação mental semântica pode ser integrada ao sistema comunicacional de Shannon desde que o interpretemos nos seguintes termos: um estado-de-coisas S^{40} , cuja fonte pode ser considerada o mundo, é uma mensagem codificada⁴¹, por sua vez, em uma sentença em mentalês, Ψp , em que p é a proposição que descreve o estado-de-coisas e Ψ é uma atitude proposicional. Dizer que aquilo em que um estado-de-coisas é codificado é uma proposição é uma forma de expressão: a linguagem do pensamento é composta de representações – é um Sistema Representacional Interno (FODOR, 1981a, p. 194) - que podem ou não ser linguísticas. Entretanto, qualquer que seja a forma que tomem, são algo que se organiza como uma linguagem e é precisamente isso o que justifica que nossos processos mentais compartilhem características com linguagens, como vimos antes. P é o conteúdo da sentença, em mentalês, correspondente ao estado de coisas representado pela mente; é também o sinal transmitido pelo canal comunicacional. Ψp pode ser decodificada para cumprir diversos propósitos, definidos pelo sistema receptor. Por exemplo, a decodificação pode ocorrer para uma sentença da linguagem natural que descreve um estado mental e, a partir disso, ele pode ser tornado consciente para fins de inspeção; ou pode sequer passar pela consciência e ser expressa linguisticamente. Mas, rotineiramente, Ψp é decodificada em uma instrução para um organismo comportar-se de certa maneira e, por sua vez, destinada a (um ou mais) sistemas efetores, isto é, sistemas responsáveis por executar o comportamento. O canal comunicacional aqui é, naturalmente, a estrutura neuronal e as degradações na transmissão dos sinais

40 Entendido em um sentido ontologicamente amplo em que a relação entre representação e estado-de-coisas se dá entre “coisas do tipo que possuem valores de verdade e coisas do tipo pelo qual valores de verdade são determinados” (FODOR, 1990, p. 32), respectivamente.

41 Naturalmente, é o cérebro quem efetua esta operação (ou seja, é o transmissor da mensagem). Por sua vez, dizer aqui que um estado-de-coisas é uma mensagem traduz a ideia bastante plausível de que nossos órgãos dos sentidos sempre captam parcelas (aspectos) das situações.

transmitidos por ele corresponderiam ao ruído.

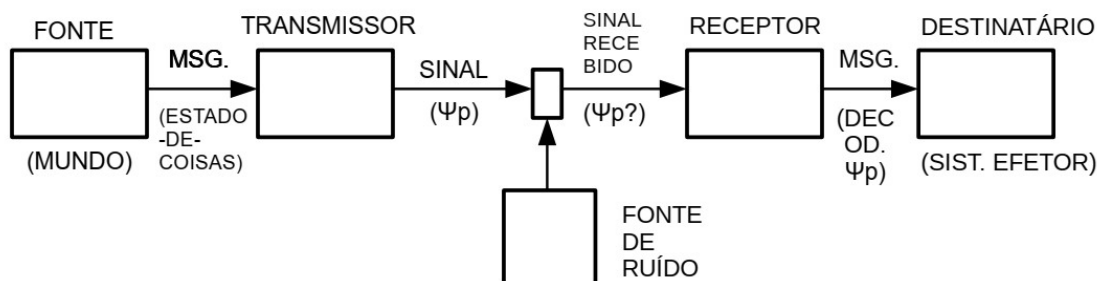
Como isto pode ser entendido como processamento de informação? P é composta de representações, com as quais a mente opera. Em tal operação, as representações contidas na mente são compostas para formar p ⁴², em que p é uma representação do estado-de-coisas tomado em conjunto. A mente processa informação nestas operações de composição e na aposição de uma atitude proposicional a p . Nestas operações de seleção, as que ocorrem primeiro são aquelas envolvidas na formação da proposição p e, em seguida, a aposição da atitude proposicional Ψ ⁴³. Assim como em nossas linguagens naturais existem probabilidades envolvidas nas composições de sentenças (a maioria das sentenças em linguagens naturais possuem a forma sujeito-verbo-objeto e apenas certas categorias de palavras podem ocupar estas funções – portanto, um número bem definido delas –, de tal forma que realizamos seleções ao formar sentenças), o mesmo ocorre na composição das sentenças Ψp . Naturalmente, tais sentenças são compostas a partir dos conceitos (entendidos como representações primitivas) que um sujeito possui que se aplicam à situação a ser representada, cujo número é finito, e dos possíveis direcionamentos que a sentença p pode assumir no sistema (cujo número também acreditamos ser finito). Além disso, nestas seleções a informação gerada por uma seleção posterior depende da probabilidade envolvida na seleção anterior⁴⁴ (isto ocorre parcialmente em linguagens naturais: por exemplo, ainda no paradigma da sentença sujeito-verbo-objeto, se o verbo da sentença é um verbo de ligação, o objeto muito provavelmente será um predicativo do sujeito – categoria na qual apenas certas classes gramaticais se enquadram). É assim que, em nossa visão, a mente processa informação. Um esquema representativo da nossa teoria do processamento da informação mental pode ser visto abaixo:

42 Que fique claro: chamamos de representações tanto os elementos que compõem p quanto o *output* deste processo, ou seja, p .

43 Isto pode parecer estranho ao leitor, uma vez que atitudes proposicionais são, por excelência, os elementos que caracterizam uma proposição como um estado mental. O que temos em mente em separar, aqui, atitudes proposicionais de proposições que são conteúdos de sentenças do mentalês é que o papel de atitudes proposicionais nestas sentenças é primariamente instrucional, ou seja, indicar o tipo de comportamento que devem gerar; do que se pode depreender que a capacidade representacional da mente não está necessariamente atrelada a estas instruções.

44 Tecnicamente, trata-se de uma cadeia de Markov.

Figura 2 – Síntese esquemática de nossa aplicação do sistema comunicacional de Shannon à mente.



Elaborado por influência de Fodor (1981a).

Alguns esclarecimentos a respeito deste esquema são necessários. Em primeiro lugar, o sinal recebido é “ $\Psi p?$ ”, com interrogação, pois, devido ao ruído, é certo que o sinal recebido nunca será equivalente ao transmitido. Em segundo lugar, os correspondentes de certos elementos do sistema que não possuem correspondentes (por exemplo, o transmissor e o receptor) não são explicitados por acreditarmos que isto se trate de um resultado de investigação empírica⁴⁵.

As sentenças p do mentalês, como dissemos anteriormente, são representações, *truth-bearers* cujos *truth-makers* são os estados-de-coisas S . Aí está o porquê de a informação mental ser semântica: ela possui, conforme a compreensão acima, uma referência. Por fim, seu conceito é o de produto (*output*) de um processo de redução de possibilidades.

Podemos inferir também que a informação assim produzida é transferida entre estados mentais através de relações lógicas que se sustentam entre seus conteúdos, o que, por sua vez, pode ser visto como uma explicação da sistematicidade da mente: a mente é sistemática porque é, em grande parte, um mecanismo para transmissão de informação. Assim, se explicaria tanto como estados mentais podem interagir entre si, quanto por que um paradigma informacional⁴⁶ é superior a um paradigma causal em explicar estas relações.

45 Seguir este caminho pode nos colocar na posição de inserir *homunculi* em nossa explicação do processamento mental da informação. Sucintamente, o problema do homúnculo diz respeito a como explicar racionalidade sem se utilizar de racionalidade, isto é, sem já pressupor racionalidade no sistema a ser explicado e, assim, cair em uma petição de princípio. Uma vez que buscamos uma explicação mecânica (não-intencional) do processamento da informação mental, o problema é um empecilho a nosso empreendimento. Exposições mais detidas e nosso tratamento do problema serão realizados nas seções 4.3 e 5.1.

46 Também temos motivos para pensar que um paradigma informacional seria ao menos explicativamente superior a um paradigma causal neste ponto porque podemos, neste caso, procurar por relações causais

Na medida em que as escolhas envolvidas na seleção do conteúdo de uma sentença na LoT produz informação com carácter semântico, podemos chamar a informação gerada por este processo de informação-sobre, no sentido amplo de ser informação sobre estados-de-coisas, de possuir *aboutness*. Quanto ao papel evolutivo, é a informação gerada na seleção da atitude proposicional que acompanha uma proposição que o tem intrinsecamente. É ela que, ao menos de maneira indireta, determina a ação decorrente dele: crer que p (dependendo do conteúdo de p) pode gerar uma ação bastante diferente de desejar que p, esperar que p, etc; são estas ações, por sua vez, que contribuem para a sobrevivência do organismo. A atitude proposicional está, em certa medida, relacionada com o conteúdo da sentença da qual é parte; por sua vez, sua escolha está relacionada com o comportamento decorrente dela. O papel evolutivo da informação-sobre, por sua vez, é provindo da intrínseca relação entre representação e informação apresentada por nós anteriormente. Portanto, pensamos que, em última instância, é a representação que dá à informação-sobre seu papel evolutivo; ela só possui esse papel de maneira vicária.

Já apresentamos, acima, algumas visões da informação mental distintas da defendida por nós com o intuito de utilizar de alguns de seus conceitos na construção de nossa teoria. Neste momento, passamos à apresentação de propostas que, apesar de divergirem da apresentada por nós, concordam conosco – em certa medida – sobre o carácter semântico da informação mental. Em nossa apresentação delas, tomaremos como complemento à nossa proposta os pontos com os quais concordamos e, nos pontos de discordância, mostraremos como nossa proposta é distinta delas.

3.2 A relação de nossa visão da informação mental com outras abordagens do tema

3.2.1 A ontologia da informação de Radu Bogdan

Nossa noção de informação-sobre aproxima-se, em grande medida, da defendida

somente no cérebro humano, no qual não há nada que nos indique uma relação causal entre a percepção de, por exemplo, um molho de tomate e a consequente esperança por uma refeição (no caso, espaguete). Por outro lado, há certos casos em que há relação causal e informacional na relação entre estados mentais: por exemplo, entre a percepção de um animal, a crença de que ele seja um animal feroz e a consequente ação de fuga. A diferença entre os dois casos apresentados por nós parece estar no papel evolutivo para a espécie humana do comportamento causado pelo estado mental.

por Bogdan (1988)⁴⁷: informação-sobre é informação material – ou seja, informação “entendida em termos de tipos materiais de fatos e sujeita apenas a restrições e leis materiais” (p. 88), também chamada por ele de informação-de por, nela, importar apenas a “interação emissor-receptor com resultados estruturais e efeitos causais estruturalmente sensíveis no receptor” –, mas com “fatos e restrições adicionais garantidores de *aboutness*” (idem). Tudo o que o receptor pode dizer sobre o emissor é, por assim dizer, a sua origem. A distinção pode ser melhor apresentada com um exemplo: uma fechadura de carro tem informação *da* chave deste carro – podemos descobrir o padrão da chave a partir do padrão da fechadura, uma vez que apenas este padrão de chave abre esta fechadura –, mas não tem informação *sobre* a chave do carro. É comum afirmar que, neste caso, a fechadura não possui informação sobre a chave do carro porque não possui a capacidade de representá-la e, portanto, não possui uma representação dela. Mas Bogdan vê esta questão de uma forma um pouco diferente.

Para Bogdan, a informação-sobre tem um caráter semântico porque possui determinado papel, determinado, por sua vez, por uma teleologia ou função, e uma meta ou objetivo. É porque a informação-sobre é também uma informação-para (uma vez que é algo que um sistema utiliza para algo) que ela possui *aboutness*. Por sua vez, o que o sistema faz com a informação deve se relacionar com (ou aplicar-se a) aspectos do mundo para que o sistema seja efetivo na tarefa para a qual utiliza a informação, de formas que expliquem seu comportamento. Na medida em que é teleológica, pode-se dizer que a informação-sobre cumpre um papel evolutivo para o organismo. Portanto, explica-se assim que esta informação se refira ao mundo e seja semântica (BOGDAN, 1988, p. 89).

Nossa concepção afasta-se da de Bogdan na medida em que, a nosso ver, exigimos menos de um sistema para considerar que a informação processada por ele seja considerada semântica. Não construímos uma nova ontologia para mostrar como sistemas processam informação – em termos de modificações geradas por interações causais entre estruturas (idem, p. 81) - e como a informação (inicialmente) material processada por eles pode passar a ser considerada semântica, mas utilizamos a teoria matemática da comunicação, tida como a teoria canônica da informação para diversas áreas das ciências naturais, e a aplicamos à mente humana com alguns acréscimos e modificações. Nossa visão da

47 Devemos notar que as noções de informação-sobre e de informação-para (da qual trataremos adiante) e a distinção entre ambas, ainda que em um sentido relativamente distinto do de Bogdan (e do nosso), aparecem em Barroso (2010; 2015).

informação mental, portanto, tem ao menos a vantagem de ser mais econômica e de dialogar melhor com mais áreas do conhecimento. Uma outra vantagem é a de evitarmos o problema de explicar a passagem entre informação material e informação semântica, sobre o qual boa parte do texto de Bogdan se dedica. O problema está em que, uma vez que concordamos que há dois tipos de informação, a material e a semântica, e uma vez que entendemos a mente como um sistema material que lida com informação semântica, devemos explicar como ela possui esta capacidade peculiar, uma vez que, via de regra, sistemas materiais lidam apenas com informação material. O apelo aos “fatos e restrições adicionais garantidores de *aboutness*” (idem, p. 88), dos quais a informação material deveria ser dotada para se tornar semântica, nos diz quais requisitos a mente deve cumprir para tanto, mas uma série de questões permanecem: quais fatos e restrições? Como eles são capazes disto? Como a mente cumpre estes requisitos? Como evoluiu para tanto? etc.

Também nos afastamos de Bogdan por acreditarmos que não só a informação-sobre tem papel evolutivo para um organismo, mas também a informação gerada na seleção da atitude proposicional que acompanha uma proposição que é conteúdo de um estado mental o tem, como afirmamos antes. Por outro lado, acreditamos que o papel evolutivo que Bogdan atribui à informação-sobre é, na verdade, provindo da intrínseca relação entre representação e informação, como afirmamos antes. Antes, porém, de explicitarmos melhor o que temos em mente aqui, é pertinente apresentar com detalhes o que Bogdan pensa sobre representações para sabermos se nosso tratamento de representações é comensurável com o seu.

Para Bogdan, as noções de informação semântica e representação estão dissociadas. Um sistema pode ter, codificar e agir sobre informação semântica sobre⁴⁸ algo sem representá-lo (BOGDAN, 1988, p. 109). Quando o faz, tem informação semântica por sua própria arquitetura, em vez de tê-la por representação. Portanto, para Bogdan, representar é apenas outra maneira (além da arquitetural) de obter informação semântica (idem, p. 112). A informação semântica, portanto, precede a representação – no sentido de que esta é uma maneira mais sofisticada de obtê-la que a arquitetural.

Informação semântica, aqui, deve ser entendida como informação material submetida a filtros ontológicos de outro tipo do que os que atuam sobre uma ontologia

48 Tenha em mente a distinção entre informação-sobre e informação-de apresentada aqui. O uso de ‘sobre’ é deliberado e também presente no autor.

rarefeita para gerá-la. A noção de filtro ontológico, grosso modo, é constituída pelos fenômenos e leis de uma ontologia densa (*thick*) e significa algo que atua sobre ela para incluir (“filtrar”) certas propriedades e relações e excluir outras (idem, p. 83).

A noção de ontologia densa, por sua vez, contrapõe-se à de ontologia rarefeita (*thin*), que indica

um mundo de *tokens* materiais nus em organizações e interações, sobre tipos e leis, cuja natureza concreta (ou densa) é simplesmente mantida sem especificação... Uma ontologia rarefeita, assim entendida, é um tipo de esquematismo em carência de um domínio de instanciação, um tipo de sistema de *place-holders* ou fendas (*slots*) esperando ocupantes atuais, densos. (BOGDAN, 1988, p. 83)

Para formar uma ontologia densa, a ontologia rarefeita deve ser especificada “tanto pela tipologia de configurações, propriedades e relações de seus itens chave (tais como partículas elementares ou moléculas ou células), quanto pelas leis governando a composição estrutural de, e as interações causais entre, tais itens.” (idem). Podemos perceber, portanto, que em um certo sentido a noção de filtro ontológico está contida na noção de ontologia densa: é ele o responsável pela especificação que a produz. Em sentido prático, podemos ver como resultado de filtros ontológicos (ou seja, como ontologias densas) as divisões entre as diversas áreas das ciências da natureza. A natureza é uma só (uma espécie de ontologia rarefeita), mas dividida em diversas áreas conforme seus itens chave e suas propriedades, configurações e relações: física, se o item chave são partículas elementares; química, se são átomos e moléculas; biologia, se é a célula (para cada área, com suas respectivas propriedades, relações e configurações).

Para Bogdan, os filtros ontológicos que incluem informação semântica são tipos teleológicos. São eles os fatos garantidores de *aboutness* mencionados acima. Exemplos de tipos teleológicos são necessidades, objetivos e metas, os quais Bogdan subsume à categoria de meta. Neste sentido, uma meta é “uma maneira de identificar e racionalizar uma classe de comportamentos que mostram sistemática sensibilidade a aspectos selecionados do ambiente distal” (idem, p. 92). Portanto, para Bogdan, apenas deve ser considerada semântica a informação que um organismo obtém e utiliza para cumprir uma meta.

A noção de tipo teleológico relaciona-se de perto com a noção de tipo semântico. Tipos semânticos são, resumidamente, “aspectos semânticos de interesse teleológico para um organismo” (idem, p. 94). Por exemplo, para um leão faminto em uma savana, os aspectos

semânticos de uma zebra são, por exemplo, sua grande concentração de carne, mas não o padrão de cores em seu pelo. A relação entre tipos teleológicos e tipos semânticos consiste precisamente em que os tipos semânticos de um organismo são identificados por seus tipos teleológicos, uma vez que eles identificam os aspectos do ambiente que importam a ele (idem).

Aqui surge a primeira de nossas críticas à teoria de Bogdan. Apesar de não terem caráter semântico para o leão no sentido desejado por Bogdan, elas certamente o tem de alguma outra forma. Do contrário, não seriam eficazes mecanismos de sobrevivência para a zebra ao iludi-lo. Se o confundem, é porque o fazem representar algo a que, de fato, não correspondem. Mas isto as torna representações falsas (ilusões), isto é, entidades que preservam algum caráter semântico.

Uma segunda crítica é a de que, assim como é possível pensar que as características semânticas dos organismos surgem a partir de sua teleologia, também podemos pensar o contrário: que características semânticas são necessárias a ela. Por exemplo, no caso citado acima, podemos pensar que é porque, para um leão faminto, uma zebra significa (ainda que em um sentido muito rudimentar de significação) um animal com grande quantidade de carne e não um animal com cores e um padrão de formas muito peculiar em seu pelo que seu comportamento em relação a ela é o de atacar e não o de observar; caso ela significasse algo diferente para ele, ele se comportaria de maneira diferente. Este sentido de significação, por sua vez – no mesmo sentido usado por nós – depende de representação. Portanto, isto mostra que podemos pensar em uma explicação alternativa à oferecida por Bogdan para a origem do caráter semântico da informação a partir da teleologia.

Para Bogdan, porém, esta saída não é viável, uma vez que seu conceito de representação, como já mencionamos brevemente acima, é bastante distinto do nosso. Para ele, a informação semântica precede a representação porque representar é apenas mais uma maneira de lidar com ela, ou seja, de “codificar, processar e utilizar tipos semânticos” (BOGDAN, 1988, p. 94), em contraposição à maneira arquitetural de fazê-lo. Sua diferença está em sua maleabilidade, em contraposição à rigidez do design arquitetural. Para clarificar a diferença entre ambos, é relevante apontar os exemplos utilizados pelo autor, a saber, dois daqueles que ele chama sistemas de informação semântica, DOP e DOPPEL.

DOP (de *door opener*) é, resumidamente, um mecanismo para abrir portas. DOP é

apresentado por Bogdan como um dispositivo que cumpre as condições elencadas por ele para que um sistema manipule informação de tipo semântica (ainda que de forma arquitetural), a saber, sensibilidade de codificação – por exemplo, DOP é capaz de ler padrões de luz a partir de uma fotocélula e obter informação deles (mas não sobre eles) –, comportamento apropriado – DOP possui uma parte motora que abre uma porta ao detectar certos padrões de luz –, seletividade interna – os inputs recebidos não permanecem restritos a aspectos proximais dos estímulos, mas são convertidos em informação sobre aspectos distais dos mesmos (por exemplo, identificar formas humanas a partir de certos tipos de padrões de luz) –, alinhamento intencional – funções que alinham sensações e comportamento cuja explicação só faz sentido a partir do apelo a tipos semânticos –, comportamento teleológico – que o comportamento do DOP seja motivado teleologicamente de tal modo que as metas do comportamento identifiquem, na fonte, aspectos que importam ativamente (e, em última instância, vitalmente) a ele (BOGDAN, 1988, p. 98-99)⁴⁹ – e, por fim, semântica por todo o sistema (*system-wide semantics*) - isto é, suas diversas partes interagem e cooperam entre si (em especial percepção e comportamento) em sua cognição (idem, p. 101). Para Bogdan, bactérias são um exemplo no mundo vivo de seres que cumprem estes requisitos (idem, p. 103-105).

É pertinente notar duas coisas a respeito de DOP e do tratamento dado por Bogdan a este mecanismo. Por um lado, parece-nos que ele o simplifica demais em seus níveis mais básicos. Enquanto DOP apenas cumpre os requisitos de ter sensibilidade de codificação e de se comportar, é certo que não temos “padrões de codificação da informação sobre restrições que refletem, de modo sistemático, aquilo sobre ou para o qual o sistema deve agir” (idem, p. 98) - aquilo que caracteriza a seletividade interna –, mas também não temos uma ligação direta, mecânica, entre estímulos e comportamento. A ligação é computacional: DOP realiza algum tipo de seleção sobre os padrões de luz captados por sua fotocélula, portanto, processa informação; se processa informação, computa. Por outro lado, Bogdan complexifica demais DOP nas adições necessárias aos requisitos mais elevados, a ponto de fazê-lo deixar de ser um mecanismo exemplar do tipo de arquitetura que um sistema semântico de informação deve possuir; ora, como tornar compreensível a ideia de uma meta vital (necessária ao requisito de comportamento teleológico) para um mecanismo que.... abre portas?

49 Para a distinção entre metas vitais e metas ativas, cf. BOGDAN, 1988, p. 92.

DOP é um exemplo de sistema arquitetural. Bogdan compara a ele outro sistema, de tipo representacional, chamado DOPPEL. A diferença essencial entre ambos, como apontado anteriormente por nós, está no modo como manipulam a informação que recebem. Uma vez que manipula informação de modo diferente, DOPPEL, em relação a DOP, sofre uma série de mudanças estruturais: mobilidade, visão tridimensional (e *templates* visuais, também tridimensionais, com os quais compara formas de objetos vindos em sua direção), audição (e *templates* análogos aos apresentados acima para a visão) e, por fim, um braço e mão mecânicos, com o objetivo de dar um aperto de mão em quem estiver cruzando o portal em dado momento. Portanto, além de estar exposto a uma gama mais rica de *inputs* e ter maior variedade de respostas e de manipulação deles, DOPPEL também é capaz de integrar e cruzar a informação recebida, inclusive para *feedback* e correção do tipo de resposta que pretende realizar. Por exemplo, apenas com seu sistema visual, por mais avançado que seja, DOPPEL poderia, por exemplo, deixar passar pela porta um urso ou qualquer animal grande, com altura semelhante à de um ser humano. Mas, ao cruzar seus dados visuais com seus dados auditivos, DOPPEL consegue corrigir (aproximativamente) sua resposta e não deixar animais como ursos passarem pela porta, uma vez que o som de seus passos é distinto do som de passos humanos (BOGDAN, 1988, p. 110-111). A adição de um braço mecânico com uma mão para dar um aperto de mão em quem passar pela porta é o apogeu deste desenvolvimento: após cruzar toda a informação obtida, DOPPEL a submete a mais um teste para se certificar de que está permitindo apenas humanos passarem pela porta; ele cruza toda esta informação com a informação obtida por uma inferência: “se o que está passando pela porta é um humano, então serei capaz de apertar sua mão”⁵⁰. Para Bogdan, é nestas conclusões dos processos que DOPPEL realiza (mas também DOP) que se encontra a informação semântica que ele processa (idem, p. 110); por sua vez, podemos dizer que as representações de DOPPEL são os produtos do seu processamento de informação semântica – assim como “informação semântica é uma forma de codificar e usar informação material, representação é uma forma de codificar e computar informação semântica” (idem, p. 112). Em síntese, é na

50 Naturalmente, como diz Bogdan (1988, p. 110), esta é a nossa maneira de nos referirmos à cognição de DOPPEL, mas ela não se dá assim. Conforme a inclinação de alguém em ciência cognitiva, isto pode ocorrer por dois motivos, que não podem ser simultaneamente sustentados: 1) a cognição de DOPPEL não envolve elementos linguísticos, portanto não pode ser formulada neles; 2) a cognição de DOPPEL se dá em termos análogos a termos linguísticos (em linguagem do pensamento, por exemplo), portanto pode ser formulada nestes (de fato, esta é a melhor explicação para esta possibilidade); mas, nesse caso, a formulação da cognição de DOPPEL em termos da linguagem natural fornece apenas uma aproximação. Assumir 2 nos permite acesso à cognição de DOPPEL: ela é o programa que diz como ele deve se comportar.

quantidade de *inputs*, na capacidade aumentada de sua integração na geração da resposta desejada e no modo que eles são manipulados que, para Bogdan, um sistema representacional se distingue de um sistema arquitetural, e não em algo que aquele tipo de sistema possui e este não. Em certo sentido, para este autor deveríamos falar na atividade de representar como uma maneira mais rica e flexível de manipular informação semântica, não em representações (como objetos).

Pelo que foi exposto, é plausível pensar que não cabe comparação entre a noção de representação apresentada aqui e a apresentada por Bogdan, uma vez que, para este autor, representar é outra maneira de lidar com informação semântica: é como sistemas com arquitetura representacional o fazem. Assim, o conceito de representação comparece ao fim da análise, como algo que surge de transformações sobre a informação semântica. Nós, pelo contrário, fazemos o conceito de representação anteceder a informação semântica, pois acreditamos que ele é o requisito para caracterizá-la como tal (ao menos se a compreendemos a partir de uma teoria causal). É porque fumaça é um sinal de fogo (indica fogo) que ela representa fogo; é por fumaça indicar e representar fogo que sou informado, ao vê-la, que algo pega fogo. Portanto, a informatividade do sinal depende da relação de representação que estabelece com aquilo que representa, que, por sua vez, depende de uma relação causal entre ambos. Parece-nos que, por recusar tratar representações como entidades que precedem a informação semântica, Bogdan não vê o papel que elas cumprem, no lugar da noção de informação-sobre, em explicar o caráter evolutivo da informação semântica (e como, de fato, são elas que dão à informação-sobre este caráter).

Cabem mais algumas palavras sobre a noção de representação defendida neste trabalho, não só para salientar seu contraste com a noção de Bogdan no contexto desta seção, mas também para mostrar em que sentido acreditamos que ela possui um papel evolutivo. Neste sentido, recorreremos à ideia de Dretske (1985) de função teleológica da representação. Originalmente, Dretske formulou esta noção para responder a dois problemas da teoria causal da representação, o problema da má representação (*misrepresentation problem*) e o problema da disjunção. Uma vez que são instâncias de um mesmo problema, podem ser apresentados em conjunto.

O responsável por gerar estes problemas é o requisito de facticidade da teoria da indicação de Grice: se x indica y, então y. Logo, uma vez que adotamos a teoria causal da

representação, isto é, defendemos que x representa y se, e somente se, x indica y , a consequência é a de que se x representa y , então y . Mas há casos desencaminhadores de indicação natural, responsáveis por erros; são justamente estes casos que colocam a teoria causal da representação em maus lençóis para explicá-los. Retomemos o exemplo apresentado acima. Pontos vermelhos na pele não são apenas sinais de sarampo, mas também de catapora. A diferença entre ambas é que, na catapora, os pontos vermelhos se desenvolvem e formam pequenas vesículas, enquanto no sarampo permanecem como estão. Pela exigência colocada à indicação, pontos vermelhos na pele devem indicar sarampo; nos casos em que pensaríamos que eles indicam outra coisa, não há, de fato, uma relação de indicação. Em outras palavras, a teoria do sentido natural não abre margem a erros. Pode-se tentar fugir a este problema ao reformular esta relação de indicação da seguinte maneira: pontos vermelhos na pele indicam sarampo ou catapora. Mas, uma vez que um índice não pode indicar confiavelmente duas coisas, somos obrigados a dizer que pontos vermelhos na pele indicam um estado de coisas descrito por uma disjunção (“sarampo-ou-catapora”). Entretanto, esta parece ser uma consequência indesejável para o sentido natural se queremos que a indicação seja confiável, uma vez que, assumindo esta versão da teoria, pontos vermelhos na pele podem indicar, em vez de um, três estados de coisas: sarampo, catapora, ou “sarampo-ou-catapora” (CRANE, 2003, p. 178-180).

O problema está em como se vê a indicação como uma condição para a representação. Se a tomamos como uma condição necessária (x não pode representar y se x não indica y , ou seja, se y não é o caso), caímos no problema da má representação, uma vez que há casos em que a condição não é cumprida (x representa y ainda que y não seja o caso). Para escapar ao problema da má representação, transformamos a condição necessária em uma condição suficiente (se x indica y , x representa y), mas caímos aí no problema da disjunção: se x indica y ou z (um estado de coisas disjuntivo), devemos também dizer que representa y ou z e perde-se parte da confiabilidade da indicação, uma vez que ela agora pode indicar três estados de coisa: y , z ou “ y ou z ”. Elimina-se a possibilidade do erro, quer porque eliminamos da representação os estados de coisa não indicados (problema da má representação), quer porque os incluímos (problema da disjunção) (idem, p. 180-181).

Como dissemos, Dretske fornece uma solução a estes problemas a partir da ideia

de função teleológica⁵¹ da representação. Em resumo, esta teoria reformula a condição apresentada por nós para a representação da seguinte maneira: x representa y se, e somente se, é a função de x indicar y. A principal vantagem desta reformulação é a ausência de exigência de que y seja o caso; com efeito, algo pode ter a função de indicar um estado de coisas ainda que este não seja o caso. É precisamente através desta vantagem que esta concepção da representação fornece uma resposta ao problema da má representação: nos casos em que, aparentemente, um estado de coisas distinto é indicado por um mesmo índice, a relação de representação permanece, uma vez que a função de indicação do índice não mudou. Por exemplo, no caso do sarampo, ainda que pontos vermelhos possam ser confundidos com sinais de catapora, devemos dizer que eles efetivamente indicam sarampo uma vez que sua função é indicar esta doença (idem, p. 182). Mas é ela também capaz de responder ao problema da disjunção? Há quem pense que não, uma vez que o problema pode ser igualmente formulado para funções teleológicas (idem, p. 182-183). Dretske apresenta um exemplo onde isto fica visível: bactérias que se utilizam de ímãs internos (chamados de magnetossomos) para se locomover conforme o campo magnético da Terra. Uma explicação para a função deste tipo de locomoção (chamado de magnetotaxis) é que ela faz com que as bactérias, que não conseguem sobreviver na presença de oxigênio, evitem águas superficiais (1985, p. 26 apud CRANE, 2003, p. 182-183). A questão é que, neste caso, há uma indeterminação quanto à função dos magnetossomos nestas bactérias: seria a função deles impulsioná-las para o norte magnético da Terra (como qualquer outro ímã) ou fazê-las evitar águas superficiais e, conseqüentemente, oxigênio? Podemos argumentar que os magnetossomos, de fato, podem ter ambas as funções; mas aí elas estariam novamente submetidas ao problema da disjunção, uma vez que, neste caso, o estado de coisas que magnetossomos têm a função de indicar é “impulsionar para o norte magnético da Terra ou evitar águas superficiais” (CRANE, 2003, p. 183).

De início, a situação apresentada acima não coloca um problema para a teoria da informação mental apresentada por nós, uma vez que não nos comprometemos com que bactérias sejam capazes de processar informação semântica. De fato, isto seria até implausível

51 A título de esclarecimento, função teleológica não é exatamente o mesmo que função evolutiva, mas as tratamos aqui como se fossem porque aceitamos, em certa medida, a subordinação que Bogdan faz das metas ativas às metas vitais. Ou seja, funções teleológicas, para seres vivos (por exemplo, a função do coração de bombear o sangue), servem a papéis evolutivos (um organismo não consegue se manter vivo se não tiver seus nutrientes distribuídos através da circulação sanguínea).

segundo nossa teoria, pois é difícil acreditar que bactérias sejam capazes de representar estados de coisas proposicionalmente. Mas acreditamos que um segundo equívoco é cometido ao mostrar que a função teleológica da representação não é capaz de responder ao problema da disjunção, a saber, a ausência de distinção entre metas ativas e metas vitais.

De maneira geral, para Bogdan, uma meta é “uma maneira de identificar e racionalizar uma classe de comportamentos que mostram sensibilidade sistemática a aspectos selecionados do ambiente distal” (BOGDAN, 1988, p. 92). ‘Meta’ é uma categoria que se aplica a organismos e que engloba coisas como necessidades e objetivos, por exemplo. Ainda segundo Bogdan, há dois tipos de metas, vitais e ativas. Metas vitais são metas gerais necessárias à sobrevivência do organismo, como comer, se reproduzir, evitar perigo, entre outras. Por sua vez, metas ativas são especificações das metas vitais, adaptações destas ao ambiente em que o organismo se encontra (idem). Podemos dizer, portanto, que metas ativas são o modo como um organismo, com sua conformação específica e frente as especificidades do ambiente em que se encontra, cumpre suas metas vitais.

Aplicando estas noções ao exemplo apresentado, a meta ativa deste tipo de bactéria é locomover-se conforme o campo magnético da Terra, enquanto sua meta vital é evitar águas superficiais (para evitar oxigênio e, no limiar, sobreviver). Em última instância, porém, organismos têm apenas metas vitais; metas ativas, uma vez que estão subordinadas às vitais, são apenas instrumentos para explicarmos como o organismo as implementa. Portanto, isso favorece a leitura de que os magnetossomos destas bactérias têm apenas uma função: a de fazê-las evitar águas superficiais.

Para concluir esta seção, é necessária ainda uma palavra sobre a concepção de Bogdan da informação semântica. Talvez este autor concordasse com parte do que apresentamos aqui, pois ele mesmo não acreditava que a informação semântica, conforme descrita por ele, cumpria todos os requisitos para servir como combustível da cognição. Ele o diz textualmente:

A tipificação semântica da informação.... não reflete (porque não está sujeita a) os tipos de fatos e restrições que são responsáveis pelo tipo de informação manipulada em acreditar, planejar, inferir ou agir. Há uma terceira forma de informação, informação mental.... Informação mental é informação semântica fazendo trabalho cognitivo e comportamental real. As leis e restrições sobre este trabalho, e portanto a tipificação da informação que o faz, podem ser mostradas como essencialmente diferentes de, e irreduzíveis a, aquelas responsáveis pela tipificação de informação

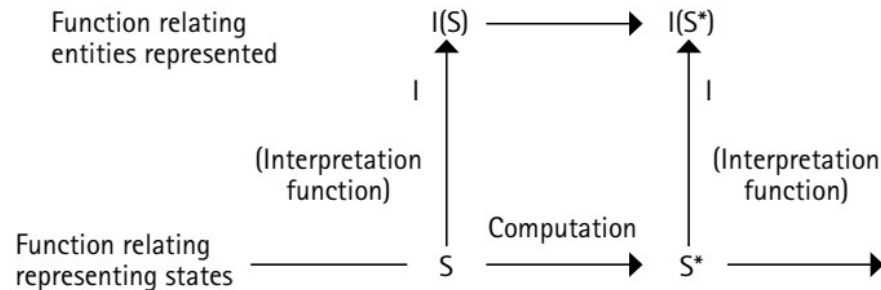
material e semântica. (BOGDAN, 1988, p. 121)

3.2.2 *Concepções funcionais da informação mental*

Nossa visão da informação mental não é incompatível com concepções funcionais da computação mental, como, por exemplo, as de Dietrich (1989), Shagrir (2006) e de Ladyman (2009), que, por sua vez, se baseiam na abordagem de Cummins (1989). De uma maneira geral, estas concepções se baseiam em uma visão não-reduativa da representação.

Na Teoria Causal da Representação, R depende de uma relação causal com b para ser uma representação de b, enquanto na Semântica Interpretacional (SI) de Cummins, R depende de uma interpretação para ser uma representação de b. Essa interpretação não é arbitrária, mas também não é um nexos fixo e natural com b como é na Teoria Causal da Representação (daí o problema do erro). Na SI, toda representação faz parte de um processo (simulação) que simula (*matches*) outro processo (processo-alvo) no sentido de que, para cada entrada do processo-alvo que é mapeada numa saída s do processo-alvo há uma entrada da simulação I(S) que é mapeada numa saída I(S*) da simulação, sendo I uma função de interpretação que associa cada elemento do processo-alvo a um elemento da simulação. Com isso, pode-se dizer que I(S) é uma s-representação de e e que I(S*) é uma s-representação de s. Em suma, na SI, não há representações naturais, no sentido de representações geradas naturalmente por um processo cognitivo. O que existe são os processos cognitivos e suas estruturas de dados, sendo que tais estruturas de dados podem ser usadas para representar os objetos intencionados durante os processos cognitivos. Essas estruturas não têm significado fora do processo; pelo contrário, é o processo que lhes atribui um conteúdo representacional. É por isso que Cummins chama essas estruturas de s-representações em vez de representações *simpliciter*: elas só ganham conteúdo dentro de uma simulação. Por fim, todo o processo é tomado como uma computação. Uma visualização gráfica desta visão da computação pode ser encontrada no esquema da *tower bridge*, de Cummins (1989, *apud* CRANE, 2003, p. 204):

Figura 3 – Resumo esquemático de concepções funcionais da computação (*tower bridge*)



Fonte: CRANE, 2003.

A visão abordada por nós, a teoria causal da representação, é redutiva porque tenta defini-la de tal forma que não haja nada no *definiendum* que contenha, explícita ou implicitamente, menção ao *definiens* (CRANE, 2003, p. 201). Esta abordagem implica que, para tratar o problema da informação mental (e, por consequência, o da computação mental), devemos primeiro tratar o problema da representação, isto é, fornecer-lhe uma definição redutiva. Pelo contrário, representações são definidas em abordagens não-redutivas da representação em termos de uma noção de computação que as antecede, através da função de interpretação.

A atribuição de uma função a um dado comportamento, naturalmente, obedece alguns requisitos, que também cumprem a função de barrar sua aplicação aos contextos em que ela não se aplica de modo a não trivializar sua aplicação. Esta concepção de computação tem a vantagem de dar uma definição bastante precisa à ideia de espelhamento, que utilizamos antes para descrever o tipo de relação que o *hardware* e o *software* de um computador estabelecem entre si.

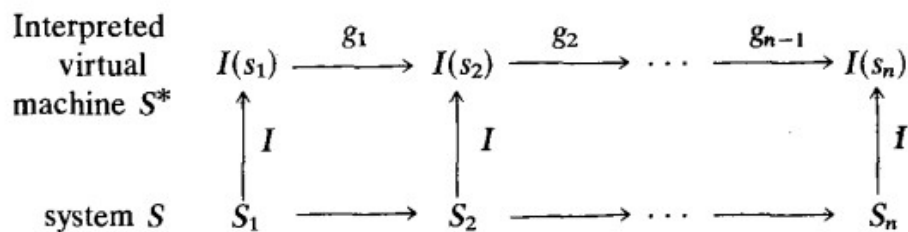
Em particular, Dietrich formula sua concepção de computação mental como uma resposta à crítica de Stich (1983) à atribuição de conteúdos semânticos a estados psicológicos. Em resumo, Stich argumenta que, uma vez que não há método cientificamente respeitável para fazer tal atribuição, o conteúdo semântico é cientificamente ocioso nelas: tudo o que elas explicam recorrendo à noção de conteúdo semântico, poderiam explicar sem ele (*apud* DIETRICH, 1989, p. 119, 125)⁵². O objetivo de Dietrich ao se contrapor a Stich, portanto, é

52 Segundo Dietrich (1989), Stich aponta duas estratégias para mostrar a ociosidade do conteúdo semântico em explicações psicológicas: a estratégia da similaridade e a estratégia causal. O leitor interessado em se aprofundar nesta questão pode consultar Dietrich (1989, p. 121-125) ou mesmo Stich (1983).

refutá-lo fornecendo tal método cientificamente respeitável, que ele vê em sua estratégia computacional. Por sua vez, a estratégia computacional não é nada mais nada menos que uma atribuição de uma função de interpretação nos termos apresentados acima.

Sendo a atribuição de uma função de interpretação, a estratégia computacional de Dietrich também deve cumprir certos requisitos para poder ser realizada. Tais requisitos podem ser formalmente apresentados da seguinte maneira: a computação de um sujeito ou sistema físico S de uma função F nos termos de uma sequência de funções $\langle g_1, \dots, g_n \rangle$ é bem-sucedida se, e somente se: 1) F é uma função composta das funções $\langle g_1, \dots, g_n \rangle$ - que, por sua vez, é um algoritmo expresso em linguagem formal (idem, p. 138, n. 10); 2) a sequência de estados pelos quais S passa é tal que cada estado corresponde ou ao domínio ou à imagem de uma das funções g_i e a imagem de uma função g_i é também o domínio da função g_{i+1} ; 3) nós conhecemos previamente as funções g_i individuais (idem, p. 126). Para Dietrich, pode-se mesmo dizer que é necessário que os parâmetros da função sejam interpretados para que o sistema S possa ser entendido como computando uma função F , uma vez que são eles que determinam a função computada pelo sistema. Além desta consequência, uma vez que a teoria se aplica a computações em geral ela também teria a consequência de que requisitos semânticos se aplicam à computação e que, portanto, ela não deveria ser vista apenas como manipulação formal-simbólica (idem, p. 133).

Figura 4 – Versão do esquema *tower bridge* para a concepção de computação de Dietrich



Fonte: DIETRICH, 1989, p. 128.

Na mesma linha, Ladyman concebe a computação como a implementação de transformações lógicas⁵³ por sistemas físicos. Para que haja computação, segundo ele, são

⁵³ Ladyman reserva o termo 'processo' para mudanças de estados físicos e 'transformação' para as correspondentes transições entre estados lógicos.

necessários “um dispositivo físico [D], uma especificação de que estados físicos deste dispositivo correspondem aos possíveis estados lógicos... e um operador de evolução temporal deste dispositivo.” (LADYMAN, 2009, p. 380). Os estados físicos em associação com estados lógicos são chamados por ele de estados representativos. O operador temporal em questão é concebido de tal maneira a garantir que “[para] *quaisquer* dos estados físicos representativos em que o dispositivo está, ele acaba no estado representativo apropriado.” (idem). Para Ladyman, portanto, uma computação é uma função cujo valor y é obtido através de um argumento x , em que x e y são estados lógicos; por sua vez, a esta transformação lógica corresponde um processo físico em que se parte do estado físico (representativo) $D_{in}(x)$ e se chega ao estado físico (representativo) $D_{out}(y)$ através de um operador temporal Λ_L definido da seguinte maneira:

$$\forall x \in X, \Lambda_L (D_{in} (x)) = D_{out} (L(x))$$

em que $y = L(x)$. Assim, uma máquina-L é formada pelo conjunto $\{D, \{D_{in} (x)|x \in X, D_{out}(y)|y \in Y\}, \Lambda_L\}$ e é implementada quando, sob operação de Λ_L , sai do estado de *input* $D_{in} (x)$ (correspondente ao *input* lógico x) e chega ao estado *output* $D_{out} (y)$, correspondente ao *input* lógico $y = L(x)$ (LADYMAN, 2009, p. 381).

Cabe aqui também uma breve menção a Shagrir (2006). Este autor faz uma abordagem do problema mais próxima à neurociência (em vez de tratá-lo em termos da noção de computação *simpliciter*) e procura entender, portanto, em que sentido podemos dizer que o cérebro realiza computações. Uma maneira de fazê-lo é definir que tipos de problemas o cérebro resolve por meio de computação. Segundo Shagrir, estes problemas são problemas “formulados em termos de regularidades” (2006, p. 402); são, em outras palavras, funções. Dizer que o cérebro computa, para Shagrir, é, mais uma vez, dizer que ele calcula funções. Estas funções, porém, são especificáveis somente através dos conteúdos de suas entradas e saídas (que são, *ipso facto*, vistas como estados representacionais). É por serem funções com este caráter específico que Shagrir as chama de tarefas semânticas.

A razão pela qual nossa concepção da informação mental não é incompatível com concepções como a de Dietrich, Shagrir e Ladyman (ao menos em parte) é que é pacífico, para nós, conceder que a redução de possibilidades operada pela mente se dá entre os elementos de um conjunto, na qual é selecionado um argumento para uma função. Uma

função é, intuitivamente, um tipo de operação na qual um elemento de um conjunto (domínio) é tomado e mapeado para outro elemento de outro conjunto (imagem) a partir de uma lei de formação. Porém, antes disso, deve ocorrer uma seleção entre os elementos do domínio para a escolha daquele que será o argumento da função, a partir do qual o restante da operação se desdobrará. É nesta seleção que há redução de possibilidades (pois, de todos os elementos do conjunto, apenas um é escolhido, ou alguns⁵⁴) e, conseqüentemente, geração de informação.

Também nos parece pacífica a possibilidade de interpretar funcionalmente nosso esquema de como a informação mental é processada: as sentenças S que representam estados-de-coisas podem ser vistas como argumentos de uma função que as computa e fornece como valor uma sentença na forma Ψp , que, por sua vez, é argumento de uma função que fornece como valor uma sentença em linguagem natural. A principal diferença entre nossa concepção da informação mental e a concepção funcional está, porém, na própria ideia de computação. Na concepção funcional, computar é calcular uma função (fornecer os valores apropriados para os argumentos apropriados) e tudo o mais (representação e informação) se segue disso. Pensamos, pelo contrário, que só é possível entender o que é computação a partir das noções de representação e, em especial, de informação.

Computar é um tipo de operação feita com informação; por sua vez, ao menos no que diz respeito à mente, representações são as portadoras da informação. Segue-se disso que, se faço a representação depender de computação, a dependência se estende à informação. Mas, por sua vez, computar é uma operação definida em termos de informação. Logo, chega-se a um *catch 22*: o que define a operação de computar é, por sua vez, definido em função dela. Pode-se contra-argumentar que a concepção funcional não necessariamente aceita uma definição de computação como processamento de informação, o que parece ser uma crítica justa. No pormenor, porém, cabe a pergunta: ela forneceria uma boa definição do que é computação se o cálculo de funções não pudesse ser interpretado como redução de possibilidades e, portanto, processamento de informação?

Outra crítica que pode ser levantada a esta concepção é a de que a atribuição de computação a um sistema é trivializada. A teoria prevê que todo sistema que pode ser

54 Virtualmente, é possível que todos os elementos de um conjunto sejam escolhidos em uma dada operação - por exemplo, em uma função constante. Mas isto não invalida o que dissemos aqui; este procedimento certamente não geraria informação, mas não se poderia dizer que ele não é informacional.

interpretado como computando dada função efetivamente a computa; mas todo sistema pode, sob dada descrição, ser interpretado como computando algo (SEARLE, 2006); logo, todo sistema computa alguma função, o que parece absurdo (é contraintuitivo pensar que, por exemplo, planetas calculem as trajetórias que descrevem). Os autores mencionados nesta seção, porém, discordam disso (CRANE, 2003; DIETRICH, 1989; LADYMAN, 2009) por acreditarem haver restrições causais a esta aplicação. Essas restrições fazem com que nem toda transição de estados $s \rightarrow s'$ possa ser interpretada como o cálculo de uma função; se isto é o caso, então há um ganho explicativo em encontrar uma função F que explique determinado fenômeno cognitivo. Como bem sintetiza Crane, “enquanto a aplicação de uma teoria a um organismo é claramente uma questão de decisão humana, não é uma questão humana se a teoria o caracteriza corretamente [ou não].” (2003, p. 207). Retornaremos a esta questão adiante, mas de antemão endossamos a crítica levantada aqui – nossos motivos para tanto, porém, só os apresentaremos a seguir⁵⁵.

55 No capítulo 3.

4 UMA DEFESA DA ATRIBUIÇÃO COMPUTACIONAL À MENTE

Concebemos, inicialmente, apresentar e responder no capítulo final desta dissertação às diversas objeções levantadas à Teoria Computacional da Mente. Entre elas, temos o argumento Lucas-Penrose (LUCAS, 1996; PENROSE, 1993), teorias concorrentes à Teoria Representacional da Mente (como a teoria conexionista e teorias de cognição incorporada, por exemplo), o debate entre o representacionalismo e o antirrepresentacionalismo e suas implicações para a teoria defendida aqui, críticas à capacidade da Teoria Computacional da Mente de fornecer uma resposta ao problema da consciência e, por fim, concepções alternativas (não-semânticas e não-representacionais) de computação aplicáveis à mente, como a funcional-valorativa (EGAN, 1991, 1992, 1995, 1999) e a mecanicista (PICCININI, 2006). Porém, escolhemos abordar aqui em particular uma das principais críticas a teorias computacionais da mente; se não sua crítica mais forte, certamente uma delas: a de que há arbitrariedade em escolhê-las como as teorias explicativas da mente por excelência porque atribuir computacionalidade aos fenômenos mentais é trivial – tudo pode ser interpretado computacionalmente. O principal partidário dessa crítica é Searle (2006), mas uma crítica similar também pode ser encontrada em Putnam (1988). Abordaremos aqui em maior detalhe a crítica elaborada por Searle.

4.1 O argumento do quarto chinês

Searle formula parte de sua crítica nos seguintes termos: a mente não pode ser considerada um computador tanto porque não é um programa quanto porque seu meio de implementação, o cérebro, não é um computador digital. Seu argumento para defender o primeiro ponto é o famoso argumento do quarto chinês (SEARLE, 1984), que pode ser brevemente exposto nos seguintes termos: imagine um quarto com duas janelas, no qual está John Searle. Folhas de papel com ideogramas chineses são jogadas dentro do quarto por uma das janelas e, conforme um conjunto de instruções contidas em um livro dentro do quarto (escrito em inglês), Searle faz algo com a folha recebida e lança outra folha pela outra janela, já contida numa pilha de papéis dentro do quarto. A alguém do lado de fora do quarto a jogar folhas de papel com ideogramas chineses pela janela, pareceria ocorrer ali alguma espécie de conversação com a pessoa dentro dele, pois ela lança folhas de papel com ideogramas chineses que simbolizam frases interrogativas e recebe, pela outra janela, folhas com

ideogramas que simbolizam suas respostas. Ou seja, pareceria que a pessoa dentro do quarto compreende o que significam os ideogramas chineses das perguntas. Mas o fato é que Searle não compreende nada de chinês; ele apenas segue um conjunto de regras preestabelecidas para respondê-las.

Naturalmente, nesta experiência de pensamento, o quarto é uma alegoria para um programa de computador. Visto de fora, ele parece realizar tarefas que envolvem semântica, ou seja, ele parece interpretar os símbolos que manipula, mas, por dentro, é puramente sintático, isto é, opera apenas conforme regras de manipulação de símbolos. Uma vez que a mente é intrinsecamente semântica (todo estado mental é caracterizado por um conteúdo) e o programa não, a mente não é um programa. O argumento pode ser melhor estruturado da seguinte maneira:

1. Programas de computador são puramente formais ou 'sintáticos': *grosso modo*, eles são sensíveis apenas às 'formas' dos símbolos que eles processam.
 2. O entendimento genuíno (e, por extensão, todo o pensamento) é sensível ao sentido (ou à 'semântica') dos símbolos.
 3. A forma (ou a sintaxe) nunca pode constituir, ou ser suficiente para, o sentido (ou a semântica)
- Conclusão: (...) rodar um programa de computador nunca pode ser suficiente para o entendimento ou o pensamento (CRANE, 2003, p. 124).

Duas observações podem ser feitas a este argumento no contexto do presente trabalho: 1) acreditamos que o argumento não atinge o empreendimento almejado aqui, uma vez que se coloca contra a possibilidade da IA-forte, não defendida aqui. Ao contrário, abraçamos uma versão da IA-fraca; 2) ainda que a sintaxe não seja suficiente para constituir a semântica, ela contribui, em grande parte, para explicar o funcionamento da mente humana. Ou seja, ainda que Teorias Computacionais da Mente não pudessem ser capazes de explicar como estados mentais adquirem seus significados, ainda assim poderiam explicar bastante coisa na medida em que a mente opera conforme computações. Basta sustentar que estas computações são realizadas sobre símbolos não-interpretados; necessitaríamos, portanto, de outra explicação para entender como eles adquirem seu sentido (*idem*, p. 126). Mas Searle rejeita também esta possibilidade, como veremos a seguir. De qualquer forma, o argumento do quarto chinês não conta como ponto contra a ideia de que o cérebro é um computador digital pois se dirige contra a ideia de que a mente é um *software*, não contra a ideia de que o cérebro é o *hardware*

desse *software*⁵⁶.

4.2 Searle contra o cognitivismo

O primeiro ponto de crítica de Searle à ideia de que explicações sobre a mente podem ser (ao menos parcialmente) computacionais é uma estranheza em relação à tese da realizabilidade múltipla do mental. Ele nos apresenta o que parece ser uma consequência indesejável desta tese:

...nessa concepção, você pode construir um ‘cérebro’ que funcione exatamente como o seu ou o meu a partir de gatos e ratos e queijo, ou alavancas, ou canos de água, ou pombos, ou qualquer outra coisa... Você simplesmente precisaria de uma quantidade incrível de gatos, ou pombos, ou canos de água, ou o que quer que fosse. (SEARLE, 2006, p. 295).

A consequência indesejável surge quando a realizabilidade múltipla não é definida em termos de produção de efeitos causais, mas de atribuição computacional. Por exemplo, assim como para computadores, a realizabilidade múltipla é válida para termostatos. Se definimos um termostato como um instrumento para medir a temperatura, então há um enorme conjunto de coisas capazes de cumprir esta função (ou seja, de produzir os mesmos efeitos causais); todas elas serão termostatos. A função também parece restringir este conjunto, pois nem tudo pode cumprir o papel causal de medir temperatura. Por outro lado, para computadores basta que algo seja capaz de cumprir as operações básicas de uma máquina de Turing para que seja considerada um computador. A atribuição é, segundo Searle, feita em termos meramente sintáticos.

Já aqui, como em outras passagens, nos parece que Searle simplifica questões complexas. Reavaliemos os exemplos anteriores. Primeiramente: “medir a temperatura” é um efeito físico? Temperatura é o grau de agitação das moléculas de um sistema. Atribuímos um número em uma escala (Celsius, Fahrenheit ou Kelvin) que corresponde a um determinado grau de agitação. A agitação das moléculas, propriamente falando, é física, mas sua divisão

56 Apesar de, naturalmente, contribuir para esta conclusão na medida em que, se não mais estamos obrigados a ver a mente como um *software*, também não mais estamos obrigados a ver no cérebro o *hardware* que o implementa. Isto, entretanto, parece implausível na medida em que, uma vez que sempre podemos interpretar qualquer órgão do corpo humano como que realizando computações, o cérebro não seria exceção a isso (ainda que possamos dizer, seguindo Chalmers (2011), que implementar computações é diferente de ser um computador). Os motivos para essa possibilidade de interpretação ficarão mais claros adiante (seção 4.5, p. 76-80).

em graus (maiores ou menores) ou mesmo a atribuição de um número a certo grau de agitação não o são; são atribuições dependentes de um ser humano, que não só mede como interpreta estes dados (o motivo para estas considerações ficará mais claro adiante). Por outro lado, considerar algo um computador não é uma atribuição totalmente desvincilhada de exigências físicas. Minimamente, um computador é, como o próprio Searle reconhece, algo capaz de realizar as operações de uma máquina de Turing: mover uma fita para a esquerda ou para a direita, escrever e apagar símbolos na fita, etc. (2006, p. 293). Ora, estes são requisitos físicos! Sua extrema simplicidade, porém, parece fazer Searle ignorar isto.

As dificuldades no tratamento de Searle destas questões parecem surgir dos tipos de requisitos que ele procura na Teoria Computacional da Mente – ou no Cognitivismo, como ele nomeia a tese de que o cérebro é um computador digital (*idem*, p. 288). Ele diz:

Queríamos saber se não havia algum sentido no qual os cérebros fossem *intrinsecamente* computadores digitais, de certa forma como as folhas verdes intrinsecamente realizam fotossíntese, ou os corações intrinsecamente bombeiam sangue. Não é uma questão de arbitrariamente ou ‘convencionalmente’ atribuímos a palavra ‘bomba’ a corações ou ‘fotossíntese’ a folhas. Há uma ocorrência efetiva do fato. (SEARLE, 2006, p. 297; grifos do autor)

Concordamos com Searle: é possível que, intrinsecamente, o cérebro não seja um computador. Mas Teorias Computacionais da Mente (ao menos a Teoria Representacional da Mente) nunca pretenderam que ele “intrinsecamente” o fosse. A Hipótese da Linguagem do Pensamento é atribuída à mente única e exclusivamente com base em um argumento para a melhor explicação: há propriedades compartilhadas entre a mente e a linguagem que não são explicadas por teorias melhores. Claro, pode-se sempre jogar a teoria pela janela e escolher ficar com explicação nenhuma se se está tão preocupado com o verdadeiro caráter intrínseco do funcionamento do cérebro.

De qualquer forma, Searle defende a ideia de que a tese da realizabilidade múltipla desemboca numa tese da realizabilidade universal, da qual surgem os problemas apresentados. Ele sustenta esta passagem numa noção de computação de aceitação aparentemente ampla, expressa em duas ideias: a de que cada objeto possui uma descrição sob a qual ele é um computador e a de que a cada programa corresponde um objeto suficientemente complexo tal que ele o implementa sob alguma descrição (SEARLE, 2006, p. 297-298). Curiosamente, ele menciona a possibilidade que acreditamos ser uma possível

resposta a este problema, a saber, impor restrições à noção de computação, mas apenas para logo deixá-la de lado porque “...o problema realmente profundo é que a sintaxe é essencialmente uma noção relativa ao observador.” (SEARLE, 2006, p. 299). Perceba que somar esta definição de computação com a tese da realizabilidade universal torna toda atribuição computacional trivial.

O problema aqui nos parece muito claro: Searle supõe haver uma definição canônica ou padrão de computação (e, mais que isso, que ela pode ser encontrada em livros-texto de computação), quando, na verdade, não há. Encontrar tal definição tem sido um problema filosófico que se arrasta já há alguns anos; portanto, não só é suspeito aceitar tão prontamente definições de computação de livros científicos como também o é tomar a questão como resolvida sem nem mesmo considerá-la, ainda mais estando ciente de autores que, à sua época, já o faziam.

Acreditamos que nem tudo das críticas de Searle até aqui deve ser rejeitado ou descartado, mas que o problema está mal colocado. Colocada de outra maneira, a tese da múltipla realizabilidade parece ser um sério problema para teorias computacionais da mente, mais do que Searle talvez faça parecer. Se queremos defender que a mente humana é uma espécie privilegiada de computador por operar computações de tipo semântico (processar informação semântica), ou seja, que deve interpretar os símbolos com os quais opera para poder operar computações, então muito do que nos permitia caracterizá-la como um computador se perde. Seria jogar fora a principal vantagem teórica das Teorias Computacionais da Mente: a ideia de que ela não está atrelada a estruturas físicas particulares. Porém, estamos longe de saber por que tal concepção seria defensável, isto é, por que devemos pensar que a mente é um tipo privilegiado de computador. Se não temos uma defesa adequada para esta afirmação, então bem podemos afirmar que é a caracterização computacional da mente que está errada de saída. E podemos defender, sem prejuízo para esta concepção, que o acesso aos itens processados, apesar de ser essencial à mente humana, não é essencial para a caracterização de suas operações, isto é, de suas computações⁵⁷.

4.3 O problema do homúnculo

57 Defenderemos esta concepção com mais detalhes na parte final deste capítulo.

Como veremos, os demais pontos da crítica de Searle ao cognitivismo dependem da ideia de que a atribuição de computação a um sistema é sempre dependente de um observador. Searle trata, por exemplo, do problema do homúnculo. O problema do homúnculo surge quando tentamos explicar a racionalidade de um sistema sem pressupô-la na própria explicação. Em outras palavras, quando tentamos explicar eventos, estados ou estruturas intencionais em termos não-intencionais, isto é, a partir do que Dennett chama de posição de *design*. Dennett (1997) chama de intencional a explicação de um sistema que atribui a ele racionalidade, em algum nível. O critério de racionalidade adotado por ele não é exatamente estrito, mas abarca tudo aquilo cujo comportamento pode ser explicado e previsto atribuindo racionalidade. Nós mesmos somos exemplos de sistemas intencionais. Ao tentar explicar e prever o comportamento de alguém, nós partimos do pressuposto de que esta pessoa age racionalmente – do contrário, nenhuma explicação seria possível. É através desta via, como já vimos, que explicações como a Hipótese da Linguagem do Pensamento funcionam – assumimos a racionalidade da mente para atribuir a ela sistematicidade, o que é uma parte fundamental do argumento a favor da hipótese. Mas, segundo Dennett, também computadores que jogam xadrez são intencionais (pensemos, por exemplo, em *Deep Blue*, o computador da IBM que venceu, nos anos 90, o campeão mundial de xadrez à época, Garry Kasparov). Isto porque o que interessa aqui à atribuição é que o comportamento do sistema seja explicado através de uma atribuição de racionalidade, não necessariamente que ele possua os estados mentais correlatos – crenças, desejos etc. Tais explicações, segundo Dennett, são feitas a partir da posição intencional.

Por sua vez, explicações a partir da posição de *design* são aquelas que “são geradas por assumir que cada parte funcional funcionará apropriadamente” (idem, p. 4), dado que tenhamos uma decomposição das partes do sistema que atribua a cada uma delas uma função (daí se falar em partes funcionais). Esta decomposição pode ser realizada com níveis de granularidade maiores ou menores e, assim, as explicações e previsões do comportamento do sistema podem possuir níveis de abstração maiores ou menores: a explicação de um computador, por exemplo, terá um nível de abstração maior se o explicarmos a partir do código que executa, e um nível de abstração menor se o explicarmos a partir de seus transistores e *switches* (idem). Em suma, portanto, o problema do homúnculo é problema de reduzir a explicação a partir da posição intencional à explicação a partir da posição de *design*.

Poderíamos pensar que há um regresso ao infinito nesta estratégia, uma vez que poderíamos sempre levar a explicação a níveis cada vez menores de abstração, mas, para Dennett, se conseguimos reduzir a explicação em termos da posição intencional a uma explicação em termos de *design*, não há exatamente um problema do homúnculo. A explicação deve parar em algum lugar – provavelmente, ali onde não mais conseguimos entender os componentes do sistema funcionalmente.

O ponto de Searle sobre este argumento é que, de fato, jamais é possível “descarregar” os homúnculos, pois a atribuição de computação ao sistema é dependente de um observador que a executa, em vez de ser intrínseca. Como afirma o próprio Searle, “a tentativa de eliminar a falácia do homúnculo através da decomposição recursiva fracassa, porque a única forma de tornar a sintaxe intrínseca à Física é colocar um homúnculo dentro da física.” (SEARLE, 2006, p. 305).

Pensamos que, se pensar um sistema como um computador nos ajuda a compreendê-lo face a outras abordagens, deve haver algo, em sua própria constituição física, que torne possível pensá-lo assim. Isto não implica que haja uma única explicação computacional do sistema, mas implica que seja possível explicá-lo em termos computacionais. A atribuição não é estritamente determinada pelo físico, mas também não é tão arbitrária quanto Searle faz parecer. Portanto, não é necessário “inserir um homúnculo dentro da Física” (idem) para tornar possível a atribuição computacional; pelo contrário, a própria Física torna possível a atribuição.

4.4 Demais críticas de Searle ao cognitivismo

Outra crítica de Searle ao cognitivismo que se baseia na arbitrariedade de atribuições de computação é sua ideia de que a sintaxe não tem poder causal, ou melhor, retira seu poder causal do meio que a implementa. O cognitivismo afirma, porém, que é justamente a partir dela que a cognição é causada. Para Searle, o poder causal que atribuímos à sintaxe, de fato, só é de fato considerado causal “nos olhos de quem vê”, pois depende de que façamos uma atribuição computacional ao cérebro (idem, p. 308). Mas para ele, conforme vimos, esta atribuição é arbitrária. Aqui cabe um comentário extra, de nossa parte, a respeito de como Searle vê a relação entre *hardware* e *software*. Não é nenhum absurdo afirmar que o poder causal de computadores está, de fato, nos processos físicos que seu *hardware* executa, mas há

algo aqui que não pode deixar de ser considerado: estes processos não ocorreriam sem a condução do *software* e das operações que ele descreve. Um *hardware* sem *software* é como uma orquestra sem maestro: todos ali estão individualmente aptos a tocar, mas precisam de alguém que os guie para que possam tocar em conjunto, como uma orquestra. Naturalmente, o maestro também faz parte da orquestra. O que queremos dizer com isso é que a relação entre *hardware* e *software*, a coordenação entre ambos, faz parte do que acontece quando um computador computa e, longe de ser completamente arbitrária, depende de condições bastante específicas (ainda para usar o exemplo da orquestra, só alguém com uma formação bastante específica pode ser considerado maestro, mesmo que sua função seja apenas a de coordenar o conjunto).

Neste momento, vemos Searle, de maneira surpreendente, apontar uma explicação neurofisiológica (como a visão de rãs funciona a nível neurológico) como exemplo por excelência de explicação sobre a visão, chegando ao ponto de afirmar: “nunca ouvi ninguém dizer que tudo isso é simplesmente a implementação de *hardware*, e que eles [os autores] deveriam ter decifrado qual programa a rã estava implementando.” (SEARLE, 2006, p. 311). Isto faz parecer que Searle representa mal as relações entre neurociência e ciência cognitiva. Ambas, a nosso ver, buscam explicações para fenômenos diferentes: enquanto a ciência cognitiva, por exemplo, busca explicar os processos e operações envolvidos na memória, a neurociência busca explicar o que ocorre no nosso cérebro quando lembramos de algo⁵⁸. Se, portanto, neurociência e ciência cognitiva têm objetivos explicativos distintos, nunca uma pode se constituir como carência explicativa da outra. A única maneira de fugir a esta conclusão é defender que tudo o que há a ser explicado a respeito de fenômenos cognitivos é esgotado pela neurociência, o que, nos termos que apresentamos, equivaleria a dizer, partindo do exemplo que apresentamos, que tudo o que há a ser explicado sobre a memória esgota-se em sua localização neuronal. Isto é, na melhor das hipóteses, bastante difícil de defender; na pior delas, simplesmente falso.

Uma estratégia para defender este ponto seria apelar para limitações na capacidade humana de apreensão de teorias. Isto é, tudo o que podemos teorizar sobre a memória é esgotado pela neurociência: tudo o que podemos conhecer sobre ela se resume ao funcionamento do cérebro, ou melhor, a como o cérebro funciona quando lembramos. Logo,

58 Certamente, ao menos, pode-se afirmar que o objeto de estudo da neurociência pode ser colocado desta forma.

conhecer a memória é conhecer o funcionamento neuronal que a viabiliza. Perceba que permanecemos num vocabulário favorável às teorias computacionais da mente por ser colocado em termos epistemológicos e não ontológicos; pois poderíamos muito bem, nesta mesma linha argumentativa, dizer, ontologicamente, “logo, a memória é um certo padrão de funcionamento neuronal”. Isto, porém, uma vez que é consequência do argumento de Searle apresentado anteriormente, traria a conclusão de torná-lo não apenas um fisicalista, mas um fisicalista em sentido forte, um defensor da teoria da identidade mente-cérebro. Daí nosso espanto com sua adesão desta forma explicativa, tão destoante do naturalismo biológico na forma defendida por ele em outros textos⁵⁹.

De qualquer forma, não é apenas deste modo que Searle representa mal as relações entre neurociência e ciência cognitiva. Atualmente, livros-texto bastante conhecidos de neurociência (por exemplo, LENT, 2010) utilizam-se amplamente, desde os capítulos iniciais, da noção de informação (ainda que não da noção de informação mental defendida aqui). Logo, o pressuposto assumido por Searle é simplesmente falso: mesmo explicações neurocientíficas canônicas de fenômenos tão básicos como sinapses já trazem embutidas em si noções como a de informação e a de troca informacional.

Por outro lado, ainda que supuséssemos a incapacidade de coexistência entre neurociência e ciência cognitiva, nos termos em que as apresentamos anteriormente fica claro que esta é explicativamente superior àquela⁶⁰. A teoria da mente da neurociência é apenas uma explicação dos fenômenos neurofisiológicos que ocorrem no cérebro quando pensamos; ao receber a réplica de que ainda há o que explicar, ela simplesmente responde o contrário, a saber, que isto é tudo o que há a ser explicado. Teóricos computacionais da mente partem da ideia de que não só há mais coisas a serem explicadas, como também que a ciência cognitiva fornece o caminho para que cheguemos a tais explicações.

A terceira e última crítica de Searle ao cognitivismo, e aquela que nos interessa mais de perto, é a de que estamos falando de coisas diferentes quando aplicamos a expressão ‘processamento de informação’ (em suma: de computação) ao cérebro e a computadores. Isto

59 Não há espaço aqui, naturalmente, para realizar uma defesa pormenorizada de teorias computacionais da mente frente a teorias da identidade mente-cérebro; mas, conforme nossa exposição mostra, sua defesa das críticas de Searle já é, ainda que de maneira parcial, uma defesa deste tipo.

60 Este argumento deve ser entendido na forma de um condicional contrafactual, ou seja, se ambas as teorias concorressem teoricamente, então a ciência cognitiva seria explicativamente superior. Uma vez que não acreditamos que isto seja o caso, ele não deve ser compreendido literalmente – mas, na possibilidade de que seja o caso, segue-se o que afirmamos.

porque, mais uma vez, para atribuímos esta operação a um computador necessitamos de um homúnculo exterior a ele para realizar a atribuição, o que não é o caso no cérebro. Mas não só isso: para Searle, também a informação no cérebro é “específica para uma ou outra modalidade” (2006, p. 320), ou seja, há informação visual, auditiva, olfativa etc., mas nenhuma delas possui o nível de abstração requerido da informação tal como compreendida no cognitivismo. Portanto, conclui Searle, o cérebro não processa informação (ou computa) tal como pensa a ciência cognitiva, pois não processa a informação tal como ela a compreende.

Pelos motivos que já apresentamos, parece ser falso pensar que a atribuição de computação ao cérebro é tão elástica e trivial quanto pensa Searle, então gostaríamos de comentar brevemente alguns dos demais elementos que o filósofo nos traz nesta crítica. Antes de tudo, devemos dizer que a ciência cognitiva não rejeita a ideia de modalidades de informação – muito pelo contrário, isto pode ser depreendido da ideia de módulos mentais (FODOR, 1983), que desempenha nela papel central. Podemos pensar que Searle tem em mente que módulos mentais não trocam informação entre si (pois é uma característica deles que sejam encapsulados, isto é, que os processos que realizam sejam isolados e, portanto, que não troquem informação), o que está correto, mas a ideia da modularidade massiva, isto é, a ideia de que todos os processos mentais são modulares e que, portanto, a informação mental é sempre específica a um ou outro módulo ainda fomenta ainda alguma controvérsia. Certamente há quem a defenda, como Pinker (1998), mas o próprio criador do conceito de modularidade mental se coloca contra esta hipótese (FODOR, 2000).

4.5 Um pouco de Chalmers sobre computação e cognição

Algumas das ideias e estratégias adotadas nas réplicas que fizemos à crítica de Searle ao cognitivismo compartilham noções e argumentos defendidos por Chalmers (2011), cuja exposição a título de esclarecimento do que temos em mente em nossa crítica se faz necessária. Por exemplo, assim como Fodor - e como nós -, Chalmers parece acreditar que a melhor sustentação para explicações computacionais da mente - o “papel fundacional da computação” - está em sua capacidade de explicar fenômenos mentais. Isso está no que ele nomeia de tese da explicação computacional, isto é, a ideia de que a “computação provê um

framework para a explicação dos processos cognitivos e do comportamento.” (p. 324). Portanto, reiteramos, a crítica de Searle é mal colocada: não é objetivo da ciência cognitiva sustentar que a mente/cérebro é intrinsecamente um computador. Talvez a crítica de Searle se aplique de forma mais direta à segunda tese usada por Chalmers para sustentar o papel fundacional da computação, a saber, a tese da suficiência computacional, que afirma que “o tipo correto de estrutura computacional é suficiente para a posse de uma mente” (idem). Se esta tese é correta, então, ontologicamente, tudo o que tem esta estrutura também tem uma mente; se o cérebro, origem causal de nossa mente, tem uma estrutura computacional, então é devido a esta estrutura que possui mentalidade. Porém, há algumas considerações a fazer aqui. Por um lado, pensamos que, apesar de Chalmers usar ambas as teses para fundamentar o estudo da cognição, acreditamos que se aplicam a áreas diferentes do conhecimento. A tese da suficiência computacional aplica-se mais propriamente à inteligência artificial, que é uma área da Teoria da Computação, enquanto a tese da explicação computacional aplica-se à ciência cognitiva. Certamente, ambas as áreas têm elementos em comum, mas também se diferenciam amplamente, como tentamos mostrar no capítulo 1 deste texto. Além disso (para esclarecimento do que, efetivamente, incorporamos de Chalmers), como afirmamos lá, nos comprometemos de modo mais forte com a tese da explicação computacional, não com a tese da suficiência. Por outro lado, no texto abordado por nós Searle dirige sua crítica ao cognitivismo e não à IA-forte, área na qual a tese da suficiência computacional se enquadra. Portanto, ou este argumento não se aplica à tese da suficiência computacional, ou, de fato, não está tão claro em Searle o que, em seus argumentos, se aplica à Inteligência Artificial e o que se aplica ao Cognitivismo.

Um outro exemplo de como Chalmers nos ajuda em nossa réplica a Searle está em nossa constatação de que este filósofo trivializa, indevidamente (a nosso ver), as relações causais necessárias à implementação de uma computação por parte de um sistema. Há aqui certas confusões a desfazer. Segundo Chalmers (2011), podemos dizer que um sistema computa (implementa uma computação) quando há uma especificação abstrata de sua organização causal com a qual ele tem uma relação de isomorfismo (p. 331). Dizer que um sistema computa, porém, não equivale a dizer que ele é um computador, pois algo é um computador somente quando é programável, ou seja, capaz de implementar várias computações (p. 335). Para Chalmers, uma trivialização da atribuição de computacionalidade

não é problemática, uma vez que ele aceita pacificamente que todo sistema pode ser visto como implementando algum tipo de computação. Nesse sentido, adota alguma versão do pancomputacionalismo, que afirma que sempre há uma descrição de sistemas físicos ou um programa em termos de uma linguagem abstrata implementada por sistemas físicos – uma vez que podemos ver o funcionamento das leis da natureza como tal programa e a matemática como tal linguagem abstrata (LLOYD, 2006; VEDRAL, 2010). Mas escapa à crítica de Searle por defender que nem todas estas computações podem ser entendidas como cognição, pois no caso de sistemas cognitivos é em virtude de implementarem uma computação que podemos considerá-los como tal. Por exemplo, tanto o sistema nervoso quanto o sistema digestivo, para Chalmers, podem ser considerados como implementando computações, mas a diferença entre ambos é que o fato de implementar uma computação é irrelevante ao sistema digestivo (é irrelevante para o mecanismo da digestão), enquanto não é irrelevante para o sistema nervoso (não é irrelevante, por exemplo, para a visão) (CHALMERS, 2011, p. 332-333). Sua trivialização, além disso, tem limites mais restritos que a de Searle: há limites físicos bastante precisos à implementação de computações (p. 331-332).

Outra maneira de pensarmos que atribuições computacionais não são tão triviais quanto pensa Searle vem de Ladyman (2009), que nos lembra algo já conhecido em ciência da computação pelo menos desde os anos 60 a partir do trabalho de Rolf Landauer: computações obedecem a restrições físicas, mais especificamente, restrições térmicas. Resumidamente, podemos dizer que a pesquisa de Landauer versou sobre a termodinâmica de computadores. Ele descobriu o princípio que leva seu nome, o princípio de Landauer, segundo o qual todo computador necessariamente deve gastar energia em suas operações, uma vez que, sendo sua memória finita, deve apagar informação dela, ação que leva ao aumento de sua entropia e/ou da entropia do ambiente (o que se traduz em aumento de temperatura) (LANDAUER, 1961, *apud* LADYMAN, 2009, p. 377). Em outras palavras, é impossível construir um computador 100% eficiente: não se pode construir um computador que realize operações indefinidamente sem gasto de energia porque não se pode construir um computador com memória infinita, ainda que estas operações sejam abstratas. A fita infinitamente longa da máquina de Turing, afinal, só existe enquanto abstração. Em suma, por mais banais ou inatingíveis que sejam os requisitos causais sobre a implementação de computações por parte de um sistema, eles existem, e elas talvez não sejam tão baratas a ponto de termos de aceitar situações como a do

exemplo de Searle, que chega a afirmar que “a parede às minhas costas está agora mesmo implementando o programa Wordstar” (SEARLE, 2006, p. 298).

Dissemos acima que, segundo Chalmers, um sistema pode ser dito cognitivo quando não é irrelevante a seu funcionamento que implemente uma computação. Isto ocorre quando a computação implementada pelo sistema é uma especificação abstrata de sua organização causal (CHALMERS, 2011, p. 336-337), ou, em outras palavras, quando ele é um exemplar das teses de suficiência e explicação computacional. A organização causal do sistema tomada abstratamente, isto é, “abstraída da constituição (*make-up*) das partes individuais e da maneira como as conexões causais são implementadas” (p. 337) é o que ele chama de topologia causal. Uma propriedade P que permanece através de todas as alterações aplicadas a um sistema que preservam sua topologia causal é o que Chalmers chama de invariante organizacional (*idem*)⁶¹ (pode-se também entendê-la, talvez, como uma propriedade que ele possui devido a esta estrutura). Propriedades mentais (ao menos a maioria delas)⁶², isto é, propriedades que caracterizam um sistema como cognitivo, são, para Chalmers, propriedades deste tipo, ainda que se dividam em propriedades psicológicas – que dizem respeito ao que a mente faz – e propriedades fenomênicas – que dizem respeito a como a mente se sente (*the way it feels*)⁶³ (p. 338-339) - e que a defesa deste caráter para ambos os tipos de propriedades mentais não tenha o mesmo peso (parece ser mais difícil defender a tese para propriedades fenomênicas).

Portanto, sistemas que compartilham a mesma topologia causal compartilham as mesmas invariantes organizacionais. Mas, se invariantes organizacionais são propriedades mentais, então estes sistemas também compartilham estas mesmas propriedades, ou seja, compartilham mentalidade. Estabelece-se assim, para Chalmers, a tese da suficiência

61 Chalmers escapa pela tangente de problemas como o colocado por Block (1978) em seu argumento da nação chinesa. Se qualia são variantes organizacionais, então um arranjo feito por todos os habitantes da China em que cada um deles simularia a atividade de um neurônio formaria um sistema capaz de ter experiências qualitativas, o que parece contraintuitivo. Provavelmente, ele admitiria que tal sistema, de fato, possui experiências qualitativas, mas não nos dá razão para concordarmos com ele.

62 Propriedades mentais parcialmente dependentes do ambiente, como conhecimento e crença, não são, para Chalmers, invariantes organizacionais (CHALMERS, 2011, p. 338).

63 Tanto esta divisão entre propriedades psicológicas e fenomênicas quanto a divisão entre definição por papel num sistema causal e ter um sentir particular (a particular feel) nos parece arbitrária por uma única razão: sentires podem ter papéis causais. O problema que qualia colocam para teorias psicológicas nunca foi o da ausência de seu papel causal, mas, como bem coloca Jackson (1982), o de que as relações causais nas quais qualia participam ocorrem apenas entre estados mentais. Ou seja, qualia assustam teorias psicológicas com o espectro do epifenomenalismo (se se aceita também a tese de que todo estado mental tem um caráter qualitativo).

computacional. Por outro lado, compartilham uma organização causal, da qual dependem as propriedades físicas das quais, por sua vez, dependem as invariantes organizacionais. Se explicar propriedades mentais é explicar as propriedades físicas nas quais eles se baseiam, elas podem ser explicadas em termos da organização causal do sistema; neste caso, estabelece-se a tese da explicação computacional (CHALMERS, 2011, p. 341-342)⁶⁴. Os requisitos que Chalmers coloca para que um sistema possua cognição, porém, nos parecem escassos, mas isto será objeto de nossa atenção apenas posteriormente.

4.6 O que é uma atribuição de computação não-trivial? A exigência representacional

Vale a pena lembrar aqui algo que dissemos no início do texto: apesar de seu grande valor heurístico no estudo da realidade, há uma clara diferença entre modelos (computacionais ou não) feitos a partir dela e a realidade propriamente dita. Um modelo do clima não é o clima, um modelo da economia não é o sistema econômico mesmo, e assim por diante. O próprio Searle (2006) parece negligenciar esta distinção ao longo de seu texto, tratando como análogas, por exemplo, aplicações de modelos tais como os anteriores e a aplicação de modelos computacionais ao cérebro. Trata-se, naturalmente, de uma falsa analogia: não é evidente que, assim como um modelo do clima não é o clima, um modelo computacional da mente não é a mente mesma. Afinal, como é admitido por alguns defensores de teorias computacionais da mente (BODEN, 2008, p. 742; SHAGRIR, 2006, p. 394), ao aplicar modelos computacionais à mente, o próprio sistema modelado é tomado como algo que computa, diferente dos casos anteriores⁶⁵. Qual é, porém, a diferença entre os dois casos, ou, ao menos, qual é a diferença entre casos de (outros) sistemas físicos e a mente? Responder esta questão envolve definir a própria noção de computação, mostrando os critérios pelos quais podemos afirmar que um sistema computa. Envolve, portanto, diferenciar um sistema físico de um sistema computacional.

Segundo Shagrir (2006, p. 395), há duas caracterizações da computação baseadas

64 Aqui o problema que se coloca para Chalmers é o de que, se ele está correto, então explicações computacionais seriam apenas explicações de como sistemas implementam computações (em sentido amplo, isto é, sobre os requisitos para um sistema implementar uma computação), o que nos parece insuficiente.

65 Poder-se-ia argumentar que negar isso é, precisamente, o ponto de Searle: um modelo computacional da mente não é a mente mesma. Mas fazer equivaler esta analogia com outras analogias entre modelos e fenômenos naturais não é suficiente para sustentar seu ponto; precisamos de um argumento que mostre que as analogias estão no mesmo patamar, o que Searle não oferece.

em condições para que um sistema a execute. A primeira condição foi expressa sinteticamente por Fodor (1981b, p. 122) e Pylyshyn (1984, p. 62) na famosa expressão: “nenhuma computação sem representação”. Crane (2003, p. 102-104) parece ter em mente esta condição para a computação ao tratar da questão a partir da distinção entre instanciar e computar. Para ele, em modelos climáticos ou econômicos, podemos dizer que estes sistemas instanciam funções. É isso que permite que eles sejam descritos por elas. Porém, seria absurdo pensar que os fenômenos mesmos (economia, clima, movimentos de planetas ou qualquer outro fenômeno que possa ser descrito matematicamente) realizam os cálculos que os descrevem para ocorrer, porque isso exige o emprego de representações (ao menos no nível básico de tomar equações matemáticas como representações de fenômenos). Por sermos capazes de empregar representações na descrição destes fenômenos, podemos dizer que computamos as funções que os descrevem.

Aliás, esta é também uma das dificuldades do funcionalismo, pois, segundo o mesmo autor (*idem*, p. 132-133), partindo da distinção acima entre instanciar e computar, o funcionalismo afirma que a mente instancia tabelas de Máquinas de Turing, mas não que ela as computa. Ora, se a mente humana é o próprio paradigma da diferença entre instanciar e computar, diferença esta que se baseia, por sua vez, no emprego de representações em determinadas atividades, o funcionalismo, ao ignorar este emprego, deixa de fora um importante aspecto do funcionamento da mente humana, além dos já apresentados acima.

Esta distinção, porém, só resolve a questão a custo de introduzir um problema: o problema do homúnculo, apresentado anteriormente. No exemplo acima, só podemos afirmar que sistemas físicos instanciam funções enquanto nós as computamos porque eles são sistemas diferentes de nós. A atribuição não funciona para nós, pois não há um sistema externo capaz de fazer a atribuição. Esta dificuldade, por sua vez, nos mostra uma forma de restringir o que deve ser considerado como representação na definição de computação acima: “representações... cujo conteúdo é independente-do-observador (observer-independent)” (SHAGRIR, 2006, p. 395). Aplicando esta restrição à distinção de Crane, podemos afirmar que a atribuição de instanciação de funções a sistemas físicos é dependente-do-observador (no caso, nós), enquanto a computação de funções (que realizamos) é independente-do-observador. Isto nos permitiria escapar ao problema do homúnculo na atribuição de computação à mente humana pois, de fato, não se exigiria um nível mais básico para as

computações que realizamos por conta da própria natureza dos itens sobre os quais a computação ocorre (representações). Entretanto, é possível que os poderes representacionais de sistemas físicos não dependam de nós em absoluto. Há um sentido em que podemos dizer que eles representam sem que os atribuamos a capacidade de representar: na medida em que eles carregam informação sobre outros eventos (idem).

Uma outra maneira de restringir a noção de representação é definindo-a como parte de um sistema simbólico, isto é, “um sistema de representações com sintaxe e semântica combinatoriais” (idem, p. 396). Portanto, só poderíamos descrever um processo computacionalmente se ele possuísse tal tipo de representações. O problema desta definição de representação é que ela deixa de fora alguns tipos de computadores, como computadores analógicos e redes neurais.

A segunda condição para que um sistema execute computações é a “condição de formalidade”: um processo é computacional se, e somente se, ele é formal (FODOR, 1981a). Há três condições para que um processo seja considerado formal, que podem comparecer individualmente ou em conjunto: mecanicidade, abstratidade (abstractness) e algoritmidade. Começamos por falar da mecanicidade, que pode ser entendida em dois sentidos. Em um deles, um processo mecânico é “um processo causal subjacente a um certo fenômeno ou comportamento” (SHAGRIR, 2006, p. 396). Se entendemos que um processo é computacional se, e somente se, ele é formal; e um processo é formal se, e somente se, é mecânico (no sentido acima), então fica claro que entender de tal forma processos formais não avança em nada nosso empreendimento de distinguir sistemas computacionais de sistemas físicos, pois ambos são aqui como que identificados entre si. Num segundo sentido, um processo mecânico é entendido em sentido lógico: um processo é mecânico se, e somente se, “é cego ao conteúdo específico dos símbolos sobre os quais opera” (idem). Poder-se-ia dizer que entender mecanicidade assim não ajuda a distinguir sistemas computacionais de sistemas físicos, pois estes podem também operar de maneira cega sobre símbolos (idem, p. 397), mas é apenas através de uma atribuição (que corre todo o risco de ser trivial e arbitrária) que podemos dizer que sistemas físicos o fazem. Normalmente, não diríamos que coisas como movimentos planetários ou a digestão envolvem operações com símbolos.

A segunda condição para que um processo seja considerado formal é a abstratidade: um processo é formal se, e somente se, há uma descrição dele que é formulada

em termos de uma linguagem abstrata (matemática, lógica ou sintática – em outras palavras, o programa) e implementada por um sistema físico, isto é, se “seus estados e operações ‘espelham’⁶⁶ os estados e operações da descrição abstrata [ou seja, do programa]” (idem). Ora, esta condição não nos ajuda a distinguir sistemas físicos de sistemas computacionais pois sempre podemos, com relativa facilidade, dizer que aqueles são, em certo sentido, abstratos. Isto é pressuposto na hipótese pancomputacionalista, mencionada acima (p. 65).

A terceira condição para que um processo seja considerado formal é a algoritmidade. Uma maneira de entender algoritmidade é como uma combinação de mecanicidade e abstratidade, isto é, um processo seria algorítmico se ele fosse mecânico e abstrato. Mas também é possível compreendê-la como uma condição sobre os tipos de sistemas abstratos implementados por um sistema físico – eles devem ser digitais, ou discretos. Assim compreendida, a algoritmidade é uma extensão sobre a condição de abstratidade. Foi sobre esta condição que incidiu a crítica de Searle (2006) de que, em um certo nível de descrição, qualquer processo pode ser entendido como implementando um algoritmo.

As condições apresentadas acima não são excludentes entre si. Pelo contrário, pretende-se, ao menos nesta apresentação, que elas se complementem. Não constatar isso tem levado diversos filósofos a discutir sobre qual seria o requisito correto para que um sistema execute computações e, portanto, a chegar a concepções diferentes do que é um computador. Particularmente, acreditamos que atribuições não-triviais de computação só podem ser realizadas, como já afirmamos, se os itens computados são representações. Talvez este não seja um bom requisito para caracterizar computações *simpliciter*, mas, neste caso, não precisamos nos preocupar com a trivialidade das atribuições computacionais. Eu não me pergunto, enquanto digito este texto, se o *laptop* à minha frente é um computador; eu me pergunto se os processos que me levam a escrevê-lo são computacionais. E dentro da classe de computações que operam sobre representações, acreditamos que mente deve estar entre aquelas cujas representações não são dependentes-do-observador. Desta forma, na atribuição

66 Uma das maneiras de compreender a relação de espelhamento é como uma forma de correspondência entre estados físicos e abstratos, da seguinte maneira: “se um estado físico S corresponde a um estado abstrato A e se um estado físico T corresponde a um estado abstrato B e se S sempre ocasiona (*brings about*) T (se o condicional suporta contrafactuais), então A sempre ocasiona B” (SHAGRIR, 2006, p. 397, n. 9). Esta definição, porém, só esclarece a noção de espelhamento na medida em que é clara a noção de correspondência utilizada nela.

computacional à mente humana pensamos que ainda vale a máxima de Fodor: “nenhuma computação sem representação” (1981b, p. 122).

4.7 Computação semântica, representação, informação

É importante diferenciarmos o requisito representacional sobre a computação do requisito semântico sobre ela, em especial porque o requisito semântico tem sido usado como ferramenta para refutar a acusação de trivialidade à atribuição computacional à mente. Este requisito foi defendido por uma série de autores (DIETRICH, 1989; LADYMAN, 2009; SPREVAK, 2010) e afirma que, para que a operação executada por um sistema possa propriamente ser caracterizada como uma computação, além de que os símbolos computados sejam representações é necessário que seus significados sejam interpretados. Ou seja, estes autores defendem requisitos semânticos para a computação. Ao menos dois dos autores que defendem esta concepção de computação (DIETRICH, 1989; LADYMAN, 2009) defendem também uma concepção funcional de computação.

Um exemplo comumente dado por defensores da concepção semântica de computação é a diferenciação na computação das funções de verdade AND e OR (LADYMAN, 2009, p. 382; SPREVAK, 2010, p. 268-269). Resumidamente, a função de verdade AND é dada pela tabela de verdade da conjunção: assim como a proposição $p \wedge q$ é verdadeira se, e somente se, ambas as proposições p e q são verdadeiras, a função AND dá como saída V (de ‘verdadeiro’) quando ambas as suas entradas são V. Analogamente, a função de verdade OR é dada pela tabela de verdade da disjunção: assim como $p \vee q$ é verdadeira somente quando ou p , ou q ou ambas são verdadeiras⁶⁷, a função OR retorna V somente quando uma ou ambas as suas entradas são V. Estes valores de verdade costumam ser implementados através do uso de álgebra booleana, onde convencionou-se que corresponde ao número ‘1’ o valor de verdade V e ao número ‘0’ o valor de verdade falso.

Funções de verdade como AND e OR são implementadas em computadores reais, comumente, através de portas lógicas. Para entendermos o que são portas lógicas, consideremos um sistema elétrico que fornece uma saída de 5V quando ambas as suas

67 Em outras palavras – e para facilitar a exposição posterior - ‘ p ou q ’ é falsa se, e somente se, tanto p quanto q são falsas.

entradas são de 5V e, do contrário, fornece uma saída de 0V. Há um claro isomorfismo entre o padrão de correntes elétricas deste sistema e as funções lógicas apresentadas por nós: se fizermos corresponder, por exemplo, à voltagem de 5V o número ‘1’, então podemos dizer que o sistema implementa a função de verdade AND e ele será, assim, um *and-gate*. Daí dizermos que o sistema é uma porta lógica: ele só “deixa passar” (implementa) a função lógica adequada quando cumpre os requisitos (físicos) de sua implementação.

O problema que os defensores da concepção semântica de computação veem neste exemplo e que o torna, para eles, um caso paradigmático de exigências semânticas sobre a computação é o de que, assim como podemos ver o sistema como computando a função AND, também podemos vê-lo como computando a função OR: basta que atribuamos à voltagem de 5V o valor ‘0’ em vez do valor ‘1’. Uma vez que convencionamos que ‘0’ corresponde ao valor de verdade F e que as condições de verdade da disjunção preveem que uma disjunção é falsa se, e somente se, ambos seus termos são falsos, podemos dizer sem embaraço que, dadas estas condições, é possível dizer que o sistema computa a função OR, não a função AND. Na verdade, o problema consiste nisso: se não especificarmos um conteúdo semântico às entradas e saídas do sistema, não conseguimos diferenciar entre possíveis computações executáveis pelo sistema. Não temos critérios para afirmar que o sistema implementa esta computação e não aquela. O ponto a ser defendido a partir deste exemplo é o de que uma definição adequada de computação deve dar conta de distinções como esta, mas este é precisamente um caso do qual a concepção sintática de computação não dá conta. Portanto, ela não é adequada⁶⁸.

Por mais satisfatórios que sejam estes argumentos para a teoria da computação (e há muita controvérsia sobre isso), eles não se aplicam (ou não devem, idealmente, se aplicar) à cognição por um simples motivo: se aceitamos a situação acima como razão para pensarmos

68 Pode-se replicar que, apesar de o exemplo dar razões para que aceitemos a concepção semântica como a concepção mais adequada de computação, a concepção semântica de computação não consegue estabelecer que seus critérios devam ser seguidos para que uma operação seja considerada uma computação. Neste exemplo, ainda que do ponto de vista da computação sintática, vemos que há uma computação envolvida: nossa dúvida é sobre qual computação, efetivamente, é realizada pelo sistema. A isto, eu responderia que permitir tal frouxidão nas atribuições computacionais a um sistema faria com que, efetivamente, deixássemos de considerá-lo como computacional. O apelo do exemplo de Searle, de que a parede atrás dele poderia estar a implementar um programa simplesmente porque este programa poderia ser identificado nos padrões moleculares que a constituem, vem exatamente daí: se eu posso interpretar um sistema físico como implementando mais de uma computação, então a atribuição é trivial (dependente-do-observador) e, efetivamente, ele não está a implementar nenhuma. É por isso que o elo entre a computação realizada e a operação física correspondente deve ser tão restrito. O que defensores da computação semântica defendem é que só podemos estabelecer este elo de tal maneira ao impor requisitos semânticos sobre a computação.

que toda e qualquer computação tem requisitos semânticos, então também aceitamos que estes requisitos são colocados por um agente externo. Assim como na atribuição da função computada pelo sistema (ao menos para os casos em que os defensores de requisitos semânticos para a computação abraçam uma concepção funcional de computação), também somos nós que atribuímos semântica ao computador para que ele seja capaz de diferenciar entre a computação de AND e de OR. Isso nos faz cair, novamente, na crítica de Searle de que a atribuição computacional é sempre dependente de um agente externo. Não podemos partir do exemplo acima e dizer que o computador “sabe” que está computando a função AND e não a função OR se ele só é capaz de tal operação porque foi configurado, por um agente externo, para tal. Computações mentais, que são nosso objeto de estudo e que constituem o que se pode propriamente chamar de cognição, não devem depender de atribuições de agentes externos para serem consideradas como tal, pois não concebemos⁶⁹ nenhum agente externo que pudesse realizar tal atribuição. Portanto, das duas uma: ou aceitamos a concepção semântica de computação e abandonamos o paradigma computacional para a mente, na esteira das críticas de Searle, ou rejeitamos a concepção semântica de computação. É por este motivo que pensamos que esta concepção de computação trivializa sua aplicação, ou, ao menos, torna a aplicação da noção à mente trivial.

69 Dentro de um paradigma naturalista.

5 CONCLUSÃO

A título de conclusão, gostaríamos de salientar as respostas fornecidas às três perguntas que nortearam o núcleo de nossa discussão (o que é a informação mental? porque a informação mental é semântica? como ocorre seu processamento?) e de preencher algumas lacunas teóricas remanescentes no texto.

O conceito de informação adotado por nós (o “quê” da informação mental) é o mesmo que pode ser apreendido da Teoria Matemática da Comunicação: redução de possibilidades. A informação mental é fruto de operações de composição (escolhas) entre representações para formar sentenças do mentalês, que envolvem redução de possibilidades dentro de um conjunto finito de representações primitivas. Uma vez, porém, que não só as representações resultantes destes processos, mas também seus componentes básicos (que também chamamos de representações) possuem referência, podemos dizer que esta informação tem caráter semântico. Como já afirmamos, o processamento desta informação se dá através de operações de redução de possibilidades na composição de sentenças na linguagem do pensamento, em cujo processamento escolhas anteriores influenciam as probabilidades envolvidas em escolhas posteriores.

O parágrafo anterior encaminhou um resumo de nossas respostas às principais perguntas que nortearam esta dissertação. Passamos agora às respostas as questões que deixamos em aberto no seu decorrer.

5.1 O problema do homúnculo outra vez: descarregando os *homunculi*

Na seção 4.3, falamos um pouco sobre do problema do homúnculo no contexto da crítica de Searle ao caráter intrínseco de atribuições computacionais, mas o leitor atento provavelmente notou que o problema também se aplica à nossa própria proposta do processamento mental. Nesta seção, tentaremos lidar com este problema nela também.

Nosso tratamento do problema é encaminhado, ainda que de maneira breve, já em nossa resposta (no capítulo 3, seção 3.3) à pergunta sobre os correspondentes de termos (por exemplo, transmissor e receptor) presentes em nossa aplicação do esquema do sistema comunicacional à mente: trata-se de uma questão empírica. Utilizar-se de racionalidade para explicar a mente não é, como Dennett aponta, exatamente um problema explicativo, desde

que paguemos os “empréstimos de inteligência” realizados. Para o filósofo, fazemos um empréstimo de inteligência sempre que utilizamos termos como ‘sinal’, ‘mensagem’ ou ‘transmissor’, por exemplo (ou seja, o que ele denomina vocabulário intencional), na construção de teorias (DENNETT, 1997, p. 12). Aliás, ele mesmo recomenda a estratégia:

Ao buscar conhecimento do design interno [de um sistema], nossa tática mais promissora é tomar empréstimos de inteligência, dotar eventos internos e periféricos de conteúdo e então procurar por mecanismos que funcionarão apropriadamente com tais ‘mensagens’ de tal forma que possamos pagar de volta os empréstimos. (idem, p. 15).

Mas, naturalmente, a construção de teorias pode ser vista como um empreendimento coletivo. Do nosso ponto de vista, há, pelo menos, dois lados envolvidos na construção – e na economia – de uma teoria da mente: o filosófico e o científico. O lado filosófico, no qual nos posicionamos, possui a considerável vantagem de permitir ao filósofo tomar tanto quanto desejar de empréstimos de inteligência. Desta forma, ele construirá uma teoria com maior ou menor quantidade de vocabulário intencional e ficará mais ou menos no “vermelho” quanto aos empréstimos de inteligência. Sua desvantagem é, porém, o fato de que uma certa falta de autocontrole nestas atribuições pode fazer com que jamais seja possível (ou cientificamente desejável) descobrir o *design* das estruturas correspondentes aos termos intencionais de sua teoria, tamanho é o débito intencional que ela adquire – neste caso, costumeiramente vemos “gastos desnecessários”, que podem ser identificados com o uso de termos intencionais cujos correspondentes só existem na cabeça de seu criador.

O lado científico, que cuidará de levar adiante a estratégia dennettiana de procurar pelos mecanismos correspondentes aos estados, eventos, estruturas etc. dotados de intencionalidade, serve de fiador da teoria filosófica. Ele é o responsável por julgar se a teoria é suficientemente boa a ponto de garantir o empreendimento de tentar pagar esta dívida – e, naturalmente, se não fez uma dívida excessivamente alta. Dadas estas condições, parte à busca dos mecanismos correspondentes aos termos intencionais.

Desta forma, justificamos o porquê de não arriscarmos atribuir referentes a estruturas da nossa aplicação do esquema do sistema comunicacional à mente: deixamos tal tarefa para cientistas que acreditarem em nossa proposta e que confiam em nossas “finanças intencionais”. A teoria é incompleta; mas, acreditamos, por uma boa razão. Ela é completa na medida do filosoficamente possível a nós.

5.2 Em que sentido computações requerem representações?

Apesar das contribuições anteriores fornecerem maneiras de entender a noção de computação, a esta discussão também está atrelada à discussão sobre se operações computacionais devem requerer semântica (interpretação dos símbolos manipulados como entidades que se referem a algo além delas, ou, em síntese, símbolos dotados de conteúdo) ou se podem ser concebidas sintaticamente como manipulação de símbolos de acordo com regras. Nossa investigação, na medida em que toma como objeto principal a noção de informação e não a computação, consegue passar ao largo desse debate, ao menos parcialmente.

Partimos, como afirmamos antes, de uma análise do conceito de computação que a define como processamento de informação. Uma vez que tratamos aqui da informação no contexto da mente, podemos adaptar a definição acima para o cérebro afirmando que ele é um computador que processa informação mental. Já estabelecemos (p. 37) que a informação mental é algo carregado por representações mentais, ou seja, o processamento de informação mental é processamento sobre representações mentais, e que ‘processamento’ é um outro nome para designar as operações realizadas por um computador, ou, caso se prefira, um nome que designa o conjunto destas operações, cujo modelo adotado aqui é o da Máquina de Turing. Aqui devemos distinguir da discussão anterior a investigação que efetuaremos no restante desta seção: o requerimento semântico sobre computações não é o mesmo que o requerimento semântico sobre a informação mental, ou mesmo sobre a possibilidade de acesso a caracteres semânticos dos objetos operados por parte de um sistema cognitivo. O requerimento semântico sobre computações diz que, se um computador não tem acesso a caracteres semânticos dos objetos processados, ele não é capaz de executar a computação, ou que não somos capazes de diferenciar entre as possíveis computações que executa. O requerimento sobre a informação mental colocado aqui é o de que, para que ela tenha um caráter próprio, ela deve ser semântica, e para isso deve estar atrelada a representações. Naturalmente, em diversas instâncias (talvez não em todas), a manipulação desta informação exigirá acesso, por parte de um sistema (ou subsistema, se considerarmos como sistema a cognição como um todo) a caracteres semânticos dela. O máximo que se pode extrair de nossas posições é que, em tais sistemas, a computação não ocorreria caso o acesso a tais caracteres semânticos não fosse o caso, ou mesmo se fosse impedido, mas não pelos motivos

acima e sim porque estes sistemas operam sobre determinada *aboutness* e se constituíram em função dela. É por levá-la em conta que desenvolveram o acesso a caracteres semânticos das representações; conceber que não possuem tal acesso implica conceber uma mente humana originada em um mundo com diversos fatos diferentes dos fatos do mundo atual.

Há qualquer coisa de intuitivo na ideia de que a mente, ou ao menos a mente humana, é representacional. É por isto que estamos longe de pensar que o caráter representacional da mente é estabelecido por uma concepção semântica de computação, ou que a posição contrária poderia endossar o antirrepresentacionalismo. Pelo contrário, é este caráter representacional que nos permite pensar na mente como um computador, mas nem por isso devemos ser levados a afirmar que esta computação é semântica. Pelo contrário, não há nada de contraditório na ideia de que a computação operada pela mente sobre representações seja sintática (ou seja, não requeira semântica – acesso a caracteres semânticos de representações), desde que possuamos outro tipo de explicação para este acesso. Por exemplo: nosso modelo prevê que tal acesso, longe de ser um requisito, ocorre nas etapas finais do processamento da informação, quando a ela é dada um destino ou propósito (comunicar, executar uma ação etc.), ou para fins de inspeção. Deste modo, acredito que nosso acesso a representações (ou seja, a consciência) pode ser o caso (e estamos longe de negar que seja)⁷⁰ ainda que não seja um requisito das operações da mente/cérebro, até porque muitas delas ocorrem sem ele.

Em síntese: quaisquer que sejam os requisitos de execução de computações *simpliciter*, acreditamos que o cérebro os cumpre e, por outro lado, que nada de peculiar pode ser extraído dele para esse debate. Se o cérebro não se encaixa nos requisitos de computação estabelecidos nesta discussão, tanto pior para ele (ou melhor, para a concepção que o trata como um computador).

De qualquer forma, resta uma dificuldade aqui: se a interpretação das representações não é requerida para a execução da computação, que outros motivos teríamos para postular entidades como representações mentais? Ora, esta objeção parte do pressuposto de que o único papel das representações mentais é fazer a computação seguir adiante, o que não é o caso. O papel das representações mentais para a mente, enquanto portadoras da informação, é informar; a informação, por sua vez, tem o papel de colocar em ação os

⁷⁰ Isto, porém, não nos leva a endossar posições como a de Searle (2006), que coloca a consciência como requisito do caráter mental de um evento cerebral.

sistemas mentais que operam sobre determinada *aboutness*. É por isso que produz seus efeitos independentemente de interpretação.

5.3 Computação é processamento de informação?

A definição de computação como processamento de informação nos coloca um problema, ignorado até aqui: a computação é também classicamente compreendida como as operações de uma Máquina de Turing (MT). O desafio é partir da teoria computacional de Turing e chegar a um conceito de informação a partir de uma conciliação entre as duas noções de computação apresentadas – ou juntar suficientes razões para abandonar uma delas. Como fazê-lo?

A saída é, ao menos inicialmente, simples: mostrar o que, no funcionamento de uma MT, corresponde à informação e o que corresponde a seu processamento. De fato, o próprio funcionamento da MT pode ser entendido como processamento. Isto nos leva a relacionar a informação com os símbolos da MT (na LoTH, as representações em “mentalês”), mas esta relação não é tão simples. Isso porque, se dizemos que o objeto das operações (do processamento) da MT são símbolos, por mais que eles sejam dotados de sentido, a informação gerada por este processo pode não ser semântica, o oposto do que pretendemos defender aqui.

Entretanto, o formalismo da máquina de Turing, que é o que deve propriamente ser chamado de computação, é útil para compreendermos como nossa mente codifica estados-de-coisas em representações, que são os elementos de nossa mente que carregam a informação, ou como as representações são codificadas, por sua vez, em instruções. Porém, não deve ser utilizado como descrição de como nossa mente opera com a informação propriamente dita, pois há um domínio destas operações que diz respeito a sua transferência entre diversos subsistemas; para tanto, acreditamos que o sistema comunicacional de Shannon se sai melhor⁷¹.

Devemos eliminar aqui uma ambiguidade na compreensão da computação como processamento de informação. ‘Processamento’ se refere à totalidade das operações de codificação, decodificação e transferência de informação dentro do sistema que compõe a

71 Concordam conosco nesta separação entre computação e processamento de informação (ao menos em parte e não nos mesmos termos) Piccinini e Scarantino (2010).

mente. Processamento é aquilo que faz com que a mente, a partir de suas percepções (e das informações que elas fornecem), produza um comportamento em um organismo. Logo, temos de diferenciar computação de processamento de informação, uma vez que computações são apenas parte destas operações, a saber, as operações que ocorrem nos subsistemas da mente: na conversão, por parte dos sistemas perceptivos, de um estado-de-coisas em uma mensagem do sistema, na codificação desta mensagem em um sinal, na decodificação deste sinal por parte do sistema receptor, etc. Esta informação deve ser dita semântica por ser gerada a partir de entidades com caráter semântico, a saber, representações.

REFERÊNCIAS

- BARROSO, C. A. C. **Uma análise internalista dos nomes próprios**. 2010. 240 f. Tese (Doutorado em Filosofia) – Departamento de Filosofia, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2010.
- BARROSO, C. A. C. Porque representações são um problema para representacionalistas. **Problemata: R. Intern. Fil.**, Paraíba, v. 6, n. 2, p. 240-259, out. 2015.
- BLOCK, N. Troubles with functionalism. **Minnesota Studies in the Philosophy of Science**, Minneapolis, v. 9, n. 1, 1978.
- BODEN, M. A. **The philosophy of artificial intelligence**. Oxford: Oxford University Press, 1990.
- BODEN, M. A. Information, Computation, and Cognitive Science. *In*: ADRIAANS, P.; VAN BENTHEM, J.; GABBAY, D. M.; THAGARD, P.; WOODS, J. (eds.). **Handbook of the Philosophy of Science: Philosophy of Information**. Amsterdam: Elsevier, 2008. p.741-761.
- BOGDAN, R. J. Information and Semantic Cognition: An Ontological Account. **Mind & Language**, Oxford, v. 3, n. 2, p.81-122, jun. 1988.
- CARNIELLI, W. A.; EPSTEIN, R. L. **Computabilidade, funções computáveis, lógica e os fundamentos da matemática**. 2.ed. São Paulo: Editora UNESP, 2009.
- CHALMERS, D. J. On Implementing a Computation. **Minds and Machines**, Holanda, v. 4, n. 4, p.391–402, nov. 1995.
- CHALMERS, D. J. Does a Rock Implement Every Finite State Automaton?, **Synthese**, Holanda, v. 108, n. 3, p. 309–333, set. 1996.
- CHALMERS, D. J. A Computational Foundation for the Study of Cognition. **Journal Of Cognitive Science**, Seul, v. 12, n. 4, p. 325-359, dez. 2011.
- CHALMERS, D. J. The Varieties of Computation: A Reply, **Journal of Cognitive Science**, Seul, v. 3, n. 3, p. 213–248, 2012.
- CRANE, T. **The mechanical mind**: a philosophical introduction to minds, machines and mental representation. 2. ed. London: Routledge, 2003.
- CUMMINS, R. **Meaning and mental representation**. 1. ed. Cambridge, MA: MIT Press, 1989.
- COMPANY calls epilogue. Intérprete: Benjamin Gibbard. Compositor: Benjamin Gibbard. *In*: WE HAVE the facts and we're voting yes. Seattle: Barsuk Records, 2000. 1 CD, faixa 8.

DENNETT, D. **Brainstorms**: philosophical essays on mind and psychology. London: Penguin Books, 1997.

DIETRICH, E. Semantics and the computational paradigm in cognitive psychology. **Synthese**, Holanda, v. 79, n. 1, p.119-141, abr. 1989.

DODIG-CRNKOVIC, G. **Investigations into information semantics and ethics of computing**. 2006. 133 f. Tese (Doutorado em Ciência da Computação e Eletrônica) - Departamento de Ciência da Computação e Eletrônica, Universidade de Mälardalen, Arkittektkopia, 2006. Disponível em: <http://mdh.diva-portal.org/smash/get/diva2:120541/FULLTEXT01.pdf>. Acesso em: 20 ago. 2020.

DOYLE, A. C. A banda malhada. *In*: DOYLE, A. C. **As aventuras de Sherlock Holmes**. Rio de Janeiro: Zahar, 2011, p. 236-274.

DRETSKE, F. **Knowledge and the Flow of Information**. 1. ed. Oxford: Blackwell, 1981.

DRETSKE, F. Misrepresentation. *In*: BOGDAN, Radu (ed.). **Belief**: form, content and function. Oxford: Oxford University Press, 1985.

EDELMAN, S. **Computing the Mind**. Oxford: Oxford University Press, 2008.

EGAN, F. Must psychology be individualistic?. **Philosophical Review**, Nova Iorque, v. 100, n. 2, p. 179-203, abr. 1991.

EGAN, F. Individualism, computation, and perceptual content. **Mind**, Reino Unido, v. 101, n. 403, p. 443-459, jul. 1992.

EGAN, F. Computation and content. **Philosophical Review**, Nova Iorque, v. 104, n. 2, p. 181-204, abr. 1995.

EGAN, F. In defence of narrow mindedness. **Mind and Language**, Oxford, v. 14, n. 2, p. 177-194, jun. 1999.

FODOR, J. A. **The language of thought**. Nova Iorque, NY: Thomas Y. Crowell, 1975.

FODOR, J. A. **Representations**: philosophical essays on the foundations of cognitive science. Cambridge, MA: MIT Press, 1981a.

FODOR, J. A. The mind-body problem. **Scientific American**, Nova Iorque, v. 244, n. 1, 114–123, jan. 1981b.

FODOR, J. A. **The modularity of mind**. Cambridge, MA: MIT Press, 1983.

FODOR, J. A. **A Theory of Content and Other Essays**. 1. ed. Cambridge, MA: MIT Press, 1990.

FODOR, J. A. **The mind doesn't work that way**. Cambridge, MA: MIT Press, 2000.

GLEICK, J. **A informação: uma história, uma teoria, uma enxurrada**. 1.ed. São Paulo: Companhia das Letras, 2013.

GRICE, H. P. Meaning. *In*: GRICE, H. P. **Studies in the way of words**. 1. ed. Cambridge, MA: Harvard University Press, 1991.

JACKSON, F. Epiphenomenal qualia. **The philosophical quarterly**, Oxford, v. 32, n.127, p. 127-136, abr. 1982.

JOYCE, J. **Ulisses**. trad. Bernardina da Silva Pinheiro. Rio de Janeiro: Objetiva, 2005.

LADYMAN, J. What does it mean to say that a physical system implements a computation? **Theoretical Computer Science**, Holanda, v. 410, n. 4-5, p.376-383, fev. 2009.

LANDAUER, R. Irreversibility and heat generation in the computing process. **IBM Journal of Research and Development**, Estados Unidos, v. 5, n. 3, p. 183-191, 1961.

LECLERC, A. **Uma introdução à filosofia da mente**. 1. ed. Curitiba: Appris, 2018.

LENT, R. **Cem bilhões de neurônios?: conceitos fundamentais de neurociência**. 2.ed. São Paulo: Atheneu, 2010.

LLOYD, S. **Programming the universe: a quantum computer scientist takes on the cosmos**. 1. ed. New York, NY: Vintage Books, 2006.

LUCAS, J. R. Minds, machines, and Gödel: a retrospect. *In*: MILLICAN, P. J. R. & CLARK, A. (eds.). **Machines and thought: the legacy of Alan Turing**. Oxford: Oxford University Press, 1996. p. 103–124.

PAPINEAU, D. Naturalism. *In*: ZALTA, E. N. **The Stanford Encyclopedia of Philosophy**. [S. l.], 2016. Disponível em: <https://plato.stanford.edu/archives/win2016/entries/naturalism/>. Acesso em: 20 ago. 2020.

PEIRCE, C. S. O que é um signo? *In*: MARCONDES, D. **Textos básicos de linguagem: de Platão a Foucault**. Rio de Janeiro: Zahar, 2010.

PENROSE, R. **A mente nova do rei: computadores, mentes e as leis da Física**. Rio de Janeiro: Campus, 1993.

PICCININI, G. Computation without Representation. **Philosophical Studies**, Alemanha, v. 137, n. 2, p. 205-241, set. 2006.

PICCININI, G; SCARANTINO, A. Computation vs. information processing: why their difference matters to cognitive science. **Studies In History And Philosophy Of Science**,

Reino Unido, v. 41, n. 3, p.237-246, set. 2010.

PINKER, S. **Como a mente funciona**. 2. ed. São Paulo: Companhia das Letras, 1998.

PITT, D. Mental representation. *In*: ZALTA, E. N. **The Stanford Encyclopedia of Philosophy**. [S. l.], 2018. Disponível em: <https://plato.stanford.edu/archives/win2018/entries/mental-representation/>. Acesso em: 20 ago. 2020.

PUTNAM, H. The nature of mental states. *In*: PUTNAM, H. **Mind, language, and reality: philosophical papers**. 1. ed. Cambridge: Cambridge University Press, 1975. p. 429-440. vol. 2.

RESCORLA, M. The Computational Theory of Mind. *In*: ZALTA, E. N. **The Stanford Encyclopedia of Philosophy**. [S. l.], 2017. Disponível em: <https://plato.stanford.edu/archives/spr2017/entries/computational-mind/>. Acesso em: 20 ago. 2020.

SEARLE, J. R. **Mente, Cérebro e Ciência**. Lisboa: Edições 70, 1984.

SEARLE, J. R. **A redescoberta da mente**. 2. ed. São Paulo: Martins Fontes, 2006.

SHAGRIR, O. Why we view the brain as a computer. **Synthese**, Holanda, v. 153, n. 3, p.393-416, out. 2006.

SHANNON, C. E. A Mathematical Theory of Communication. **Bell System Technical Journal**, Estados Unidos, v. 27, n. 3, p.379-423, jul.1948.

SPREVAK, M. Computation, individuation, and the received view on representation. **Studies In History And Philosophy Of Science**, Reino Unido, v. 41, n. 3, p.260-270, set. 2010.

STICH, S. **From folk psychology to cognitive science: the case against belief**. Cambridge, MA: Bradford/MIT, 1983.

VEDRAL, V. **Decoding reality: the universe as quantum information**. New York, NY: Oxford University Press, 2010.

WEAVER, W. Recent contributions to the mathematical theory of communication. *In*: SHANNON, C. E. & WEAVER, W. **The mathematical theory of communication**. Urbana, IL: University of Illinois Press, 1972.