

3. Origens históricas dos diferentes problemas da causação mental

3.1. O problema de Elizabeth

A questão da causação mental tem não uma, mas duas fontes históricas: o dualismo de René Descartes e a filosofia da mente de Donald Davidson. Esta seção traz uma breve descrição de ambas, para que fique claro o pano de fundo das ideias a serem desenvolvidas.

Por que a questão da interação entre mente e corpo é tão gritante na filosofia cartesiana? A radical separação entre ambos criou uma armadilha da qual parece ser impossível escapar. Descartes acreditava que a extensão é a essência das substâncias corpóreas. Todas as outras noções relativas à matéria apontam para esta propriedade essencial, pois é através da mesma que podemos inteligir os importantes conceitos de dimensão, forma e movimento. Não é possível, portanto, apreender a ideia de substância corpórea separadamente dessa característica crucial. A compreensão da matéria se dá, portanto, através da ideia de extensão.⁶³ Podemos conceber corpos sem atributos como cor, calor (curiosamente, em Descartes, estes pressupõem percepções, e por isso são características da mente) ou dureza. Mas não é possível uma operação mental onde a extensão dos corpos é removida sem destruí-los. Descartes também argumenta que, se podemos formular provas a respeito das características geométricas dos corpos, trazendo à tona fatos outrora desconhecidos, é correto concluir que tais verdades têm uma contrapartida na realidade objetiva. Portanto, os corpos têm necessariamente as características geométricas a eles atribuídas.⁶⁴ Em verdade, os corpos são “os objetos geométricos tornados reais”.⁶⁵

Nada precisamos acrescentar aqui sobre a ontologia dos corpos na mecânica cartesiana. Falta esclarecer a perspectiva cartesiana sobre o movimento dos mesmos. Descartes, da mesma forma que os demais teóricos da mecânica pré-

⁶³ *Princípios*, Parte I, art. 53.

⁶⁴ *Meditações*, Parte 5, AT VII 65.

⁶⁵ Garber, D. *Descartes' Physics*. In *The Cambridge Companion to Descartes*.

newtonianos, dava como certo que impactos respondem por todas as ações físicas. Toda mudança na dinâmica dos corpos se deve a contatos entre corpos. Todas as forças são características internas a objetos extensos, sendo transferidas a outros corpos mediante impacto.⁶⁶

A mente, por sua vez, tem como atributo essencial o pensamento. O pensamento consciente se identifica com a mente como um todo. Descartes chega a essa conclusão após estabelecer a certeza de sua própria existência com o célebre *cogito* da *Segunda Meditação*. Tendo obtido a prova de que seu próprio pensamento é real, e, portanto, ele mesmo tem que sê-lo, Descartes medita sobre a veracidade de sua autoimagem. Seu corpo, suas ações e percepções podem todas ser mentirosas, meros produtos de um gênio maligno, mas seu pensamento necessariamente tem que existir. Se o pensamento cessasse permanentemente, o sujeito que pensa também seria obliterado. Sendo assim, Descartes conclui que ele mesmo e seu pensamento são inseparáveis. Daí se segue que a mente difere radicalmente da matéria, podendo ser concebida sem qualquer relação com coisas extensas. O que pode ser concebido sem contradição pode ser feito por um Deus todo-poderoso; conclui-se que a mente é uma substância, ou seja, algo capaz de existir independentemente de todo o resto. O espírito, portanto é uma realidade distinta do mundo material. Além disso, as operações do espírito são sempre perfeitamente evidentes para o próprio espírito; segue-se que, não sendo capaz de encontrar divisões na mente (seria absurdo falar em “um terço da alma”), elas não existem. A matéria, ao contrário, é sempre divisível. Se isto é verdade, a conclusão acima é duplamente justificada. Evidentemente, os argumentos são problemáticos, mas sua consistência não nos interessa aqui. A questão a ser tratada aparece na suposta interação entre estas realidades, que examinamos a seguir.

Descartes sustenta que as realidades absolutamente distintas da extensão e do espírito interagem na pessoa humana. Um ser humano, de fato, é uma conjunção de mente e corpo. A alma é ligada ao corpo como um todo, mas existe uma conexão especial entre a mesma e o cérebro.⁶⁷ O fundamento desta asserção

⁶⁶ Papineau, D. *The Rise of Physicalism*, p. 14-17.

⁶⁷ *As Paixões da Alma*, Artigo 31.

vem das “inúmeras observações”⁶⁸ de correlações entre alterações no cérebro e alterações em estados mentais. O cérebro, mais especificamente a glândula pineal, canaliza as determinações da alma para o corpo e vice-versa. A glândula pineal foi eleita porque

todas as demais partes do cérebro são duplas, como o são todos os órgãos de nossos sentidos externos – olhos, mãos, ouvidos etc”. Mas na medida em que temos apenas um pensamento simples sobre um dado objeto a cada momento, deve necessariamente haver um lugar onde as duas imagens advindas dos dois olhos, ou as duas impressões de um único objeto transmitidas pelos órgãos duplos de qualquer sentido, podem unir-se em uma única imagem ou impressão antes de atingirem a alma, de forma a apresentar a ela um objeto, e não dois.⁶⁹

A alma acessa o corpo pela mesma porta, “reverberando pelo resto do corpo por meio dos espíritos animais, dos nervos e mesmo do sangue, que pode receber as impressões dos espíritos e carregá-las através das artérias para todos os membros”.⁷⁰ Neste intercâmbio entre matéria e pensamento, muito do que comumente se atribui o segundo é na verdade realizado pela primeira. O armazenamento de impressões na memória, os apetites e paixões provém do corpo, assim como reações instintivas como a fuga diante de um inimigo. Só assim é possível para Descartes sustentar que animais com comportamento complexo são um tipo de autômato, ou seja, um objeto sem alma.

Independentemente das curiosidades do dualismo cartesiano, é possível que todo o edifício desabe na ausência de uma explicação consistente para a interação entre substâncias. A fragilidade da filosofia da mente cartesiana é evidenciada na correspondência entre o filósofo e a princesa Elizabeth da Boêmia. Lisa Shapiro diz que os dois provavelmente se conheceram na corte boêmia exilada em Haia, antes da publicação, em 1641, das *Meditações*. Eles se corresponderam entre 1643 e 1649, e os escritos resultantes abarcam uma grande variedade de temas além do problema mente-corpo: física, medicina, ética e filosofia política. Seja como for, a primeira carta de Elizabeth, escrita em 6 de maio de 1643, aborda sua incredulidade diante da doutrina dualista das *Meditações*:

⁶⁸ *Meditações*, Parte 6.

⁶⁹ *As Paixões da Alma*, Artigo 32.

⁷⁰ *Idem*, Artigo 34

Eu lhe peço por obséquio para dizer como a alma de um ser humano (sendo ela apenas uma substância pensante) pode afetar os espíritos corpóreos, de forma a gerar ações voluntárias. Me parece que toda determinação de movimento acontece pela impulsão da coisa movida, pela maneira como ela é empurrada por aquilo que a move, ou pelas qualidades particulares e forma da superfície desta última. As duas primeiras condições requerem contato físico, e a terceira requer extensão. O senhor exclui por completo uma [extensão] da sua noção de alma, e a outra [contato físico] me parece incompatível com um ente imaterial⁷¹.

A perplexidade de Elizabeth tem origem em seu comprometimento com uma visão mecanicista do mundo físico, onde relações causais requerem atributos possuídos unicamente por corpos – extensão ou superfície. O uso da abordagem mecanicista contra a filosofia da mente cartesiana é um caso agudo de feitiço que se volta contra o feiticeiro, já que o próprio Descartes foi um proponente crucial do mecanicismo contra a física escolástica. Duplamente irônica, então, é a resposta de Descartes, que usa o conceito escolástico de peso, que havia rejeitado em sua física. A física escolástica dizia que corpos pesados tendem a cair devido à propriedade essencial de peso, ao contrário de coisas como o ar, que tendem a subir graças à propriedade essencial de leveza. Descartes considerava tal física inadequada, mas observou em sua réplica que podemos aproveitar a aparente consistência do conceito de peso na física para formar uma analogia com as operações da mente. Quando se diz que um corpo é movido em direção ao solo pelo seu peso, não estamos supondo que o peso é um objeto distinto em contato com o primeiro. Descartes afirma, então, que o corpo é movido pela alma de forma análoga aos corpos movidos pelo peso ou leveza. Conclui-se, então, que Elizabeth teria aplicado indevidamente as leis dos corpos no reino do espírito⁷².

Em sua tréplica, Elizabeth escreve que sua concepção de ente imaterial é uma “negação da matéria, que não pode ter qualquer contato com ela”.⁷³ Ela diz também que lhe seria mais fácil admitir uma alma material do que uma união entre corpo extenso e alma inextensa.⁷⁴ Ela acrescenta que é implausível supor que a alma é ao mesmo tempo perfeitamente capaz de existir independentemente do corpo e tão suscetível às mudanças do mesmo:

⁷¹ AT 3, p. 661. Minha tradução.

⁷² AT 3, p. 667-668

⁷³ AT 3, p. 684. Minha tradução.

⁷⁴ AT 3: 685.

é muito difícil entender que uma alma, da forma como o senhor a descreve, tendo o poder e o costume de raciocinar corretamente, possa perder tudo isto por alguns vapores, e que, sendo capaz de subsistir sem o corpo, e nada tendo em comum com o mesmo, a alma ainda é governada por ele.

A suspeita, então, é que o dualismo cartesiano é incoerente: temos uma alma que independe por completo do corpo mas vulnerável aos caprichos da carne; existe um espírito sem extensão que move um corpo extenso e é por ele movido. Muito embora não seja necessário um conhecimento extenso de física para enxergar o problema, é razoável supor que o conhecimento da mecânica da época tornou o problema mais agudo para a princesa. Conforme descrita acima, a mecânica da época admitia como forças somente forças internas aos corpos, transmitidas a outros corpos por contato. Imersa nessas ideias durante sua formação intelectual, Elizabeth não enxergava uma entrada para a influência causal da alma, que não pode colidir com corpos físicos. Se todas as forças são imanentes aos objetos da geometria materializados, temos uma forma de fechamento causal do mundo físico, ainda que bastante diferente do que hoje considerariamos como tal (a próxima seção consistirá em mais detalhes a respeito). Quanto à analogia com as propriedades escolásticas, Elizabeth diz não ver qualquer relevância ali para a discussão, e pede mais esclarecimentos a Descartes.⁷⁵

Em sua carta de 28 de junho de 1643, Descartes argumenta que, tendo refletido sobre as diferentes naturezas do corpo e da alma, um filósofo pode perder de vista o contexto mais amplo da união dos dois. Ter-se-ia então o famoso erro de focar as árvores, mas deixando de ver a floresta. A jornada filosófica de Elizabeth fez com que ela perdesse de vista um fato trivial. A união entre mente e corpo, afinal, é um dado inegável da experiência cotidiana. Trata-se de uma noção primitiva, uma ideia fundamental que não pode ser sujeita a uma análise continuada em termos de noções mais básicas. De qualquer forma, seria correto atribuir extensão à unidade alma-corpo; o que é inaceitável é pensar a mente em si como coisa extensa. A extensão que figura nas reflexões de Elizabeth pertence à união e não ao espírito considerado isoladamente⁷⁶. Elizabeth retruca a estas asserções insistindo na razoabilidade de uma alma extensa, e observa que a

⁷⁵ AT 3: 684.

⁷⁶ AT 3: 693-694.

resposta de Descartes é insuficiente.⁷⁷ Sua perplexidade permanece intacta, e por bons motivos. Infelizmente, Descartes interrompe abruptamente o debate em sua carta seguinte, que trata de problemas matemáticos.⁷⁸

Não obstante o *insight* de Elizabeth, chamada por Shapiro de “a primeira pessoa a propor o problema mente-corpo”⁷⁹, a alma milagrosa do cartesianismo manteve fôlego. As intuições que rechaçam uma mente material chegam a nossos dias e continuam a gerar controvérsia. No entanto, a repugnância à concepção material de mente toma uma forma muito diferente na atual filosofia analítica da mente. Existem ali poucos defensores do dualismo cartesiano. Resiste-se ao reducionismo por meio de dualismos de propriedade, interpretações antirreducionistas da tese funcionalista de Putnam ou ainda a filosofia do monismo anômalo de Davidson. Isto significa que a reaparição do problema da causalção mental no século XX deriva de compromissos filosóficos bastante distintos daqueles debatidos por Elizabeth e Descartes no século XVII. Ainda assim, é necessário destacar que o insight de Elizabeth acerca da ininteligibilidade da perspectiva cartesiana ainda assombra o dualismo e as filosofias antirreducionistas em geral. A volta do dualismo como alternativa metafísica plausível, no entanto, trouxe um irônico retorno do dualismo cartesiano como teoria respeitável da mente. Entender essa virada de mesa é um passo importante rumo à compreensão da questão do fechamento causal como compreendido em nossos dias. O atual debate sobre causalção mental, por sua vez, é alimentado pela mesma ideia, que a seguir será examinada.

3.2. Como o fechamento causal foi abandonado e ressuscitado

Resumamos o que acaba de ser exposto: assumimos que o acima exposto ceticismo de Elizabeth foi alimentado pelo seu conhecimento da física de seus contemporâneos, que, recordemos, considerava como forças apenas as propriedades internas a corpos. Toda aceleração, portanto, era imprimida por impacto. Conseqüentemente, a manipulação do corpo por uma alma imaterial

⁷⁷ AT 4: 2.

⁷⁸ AT 4: 37-43.

⁷⁹ Introdução à *Correspondência entre a Princesa Elizabeth da Boêmia e René Descartes*, p. 23. Minha tradução.

incapaz de contato algum com corpos torna-se impossível.⁸⁰ Mesmo a física de Leibniz não oferecia qualquer consolo para um cartesiano de mente aberta, pois as forças evocadas pelo alemão eram igualmente imanentes aos corpos. A única causa possível para uma mudança no movimento de um corpo é o impacto de outro corpo. Por isso, são sombrias as perspectivas de uma interação mente-corpo coerente, ainda que as ideias cartesianas sobre conservação fossem exploradas. Um cartesiano poderia explorar estas como uma saída para o problema (um recurso que Leibniz atribui a Descartes)⁸¹, mas a intractabilidade do impacto que isto implica permaneceria. Seja como for, Leibniz chegou ao fechamento causal do físico por meio de suas leis de conservação. A física cartesiana supunha a conservação de uma quantidade de movimento, definida como massa vezes velocidade. Este movimento é não direcional; isto significa que a direção do movimento de um corpo pode ser alterada sem mudança na quantidade de movimento (o que contrasta com a física de Newton, como veremos mais adiante). Leibniz, por sua vez, adotou a conservação da energia cinética e do momento linear. Quando somadas à premissa de que não existe ação à distância (forças só são transmitidas via impacto), estas leis exigem o fechamento causal do mundo físico. Isto acontece porque o momento linear é uma quantidade vetorial; tais quantidades fazem referências a direções. Se este é o caso, não é possível alterar a direção do movimento de um corpo sem afetar a quantidade de movimento do mesmo. Somada à conservação de energia cinética, a conservação do momento determina totalmente a velocidade e direção dos objetos que interagem por impacto. Melhor dizendo, os estados físicos de um sistema de corpos determinam por completo seus estados posteriores. Se a velocidade e a direção do movimento é totalmente determinada por fatores físicos, não sobra espaço para a ação da mente sobre o corpo. Esta é a constatação que levou Leibniz a adotar sua doutrina da harmonia preestabelecida, onde a causação mental não passa de uma ilusão. Seja qual for a plausibilidade desta doutrina peculiar, uma física conservadora no sentido leibniziano faz das causas mentais um elemento redundante no mundo físico.⁸²

⁸⁰ Papineau, *The Rise of Physicalism*, p. 13-16.

⁸¹ *Monadologia*, p. 327 (referência em Papineau).

⁸² Papineau, *The Rise of Physicalism*, p. 15-16.

Como é possível, então, que o dualismo interacionista tenha sobrevivido tanto tempo após os escritos de Leibniz? A influência da mecânica de Newton, com suas forças impressas, é a resposta. A força paradigmática na física newtoniana é a gravidade; evidentemente, a gravidade de Newton é uma *força incorpórea* que perpassa o espaço e age sobre os corpos “do lado de fora”. A aceitação desta perspectiva trouxe consigo a possibilidade de ação à distância, ainda que esta tipicamente implique a presença de corpos com massa a gerar a força necessária.⁸³

Se a gravidade é uma força incorpórea, as comportas estão abertas para mais forças “igualmente incorpóreas”. O abandono do impacto como causa paradigmática de movimento significa uma física compatível com a existência de fatores exóticos a contribuir com a dança dos corpos no universo. Essa asserção é reforçada por mais uma característica da mecânica newtoniana, a saber, a ausência de uma lei sobre a causa das forças *em geral*. A lei do inverso do quadrado governa a gravidade, determinando forças gravitacionais como uma função dos corpos com massa em um sistema e suas localizações. Infelizmente, não existe na física de Newton uma lei que faça algo similar pelas forças em geral. Conclui-se que não estão descartadas forças mentais exóticas, transmissíveis sem impacto (ação à distância) e manipuladoras de corpos.⁸⁴

Neste contexto, poder-se-ia objetar que as leis de conservação da física clássica restringem o conjunto de forças admissíveis, mas tais leis só foram incorporadas à física newtoniana bem mais tarde. Nem Newton nem seus seguidores imediatos davam grande ênfase a leis de conservação de energia (embora a Terceira Lei implique a conservação do momento). A conservação de energia cinética e potencial em todos os processos físicos só foi absorvida pela tradição newtoniana em meados do século XIX, em parte porque os primeiros newtonianos não possuíam um conceito de energia potencial (definida por Papineau como “a energia armazenada após a compressão ou extensão de uma mola, ou quando dois corpos em gravitação se afastam.”).⁸⁵ Um sistema composto por dois corpos em gravitação não parecia conservar energia, portanto; a soma de

⁸³ Idem, p.16-17

⁸⁴ *The Rise of Physicalism*, p. 16-20.

⁸⁵ Idem, p.34. Minha tradução.

suas respectivas energias cinéticas decresce enquanto eles se afastam.⁸⁶ Mesmo em se tratando de impacto, não havia um compromisso com a preservação da energia cinética. Pensava-se que a perda de energia cinética após colisões de corpos não tinha que ser compensada por um ganho equivalente entre as partes internas dos corpos, ao contrário do que propunha Leibniz.⁸⁷ Voltaremos a discutir esta mudança um pouco mais adiante.

Outro ponto interessante diz respeito ao suposto determinismo da física clássica. Parece plausível que qualquer determinismo implica o fechamento causal da física. No entanto, este não é o caso. Para tornar a física completa (ou seja, todos os eventos do domínio físico determinados por eventos dentro do mesmo) é necessário um determinismo *físico* (condições anteriores *físicas* bastam para fixar os estados físicos posteriores). Se este é o caso, a física newtoniana, antes de incorporar a conservação de energia, era compatível com um determinismo onde forças mentais contribuem para fixar a evolução dos estados dos corpos.⁸⁸ Eis a porta de entrada para as forças mentais postuladas no século XVIII pelo teórico da medicina Robert Whytt. O escocês Whytt escreveu sobre uma alma que age diretamente sobre o corpo como um todo. As forças mentais atribuídas a este ente regem os movimentos voluntários e involuntários do corpo e, naturalmente, as operações racionais, assim como todas as demais dinâmicas fisiológicas. Whytt acreditava que as forças imprimidas pela alma agiam de forma determinística, e deixou claro que as via como análogas à força da gravidade de Newton. Longe de ser uma excentricidade em sua época, a utilização de tais conceitos era então ideia popular.⁸⁹ O fisiologista alemão Albrecht Von Haller teorizou em 1751 sobre uma força de sensibilidade controlada pela alma e transmitida através dos nervos para o resto do corpo. Esta força, porém, parecia para Haller agir indeterministicamente, submetida, por assim dizer, aos desígnios espontâneos da alma. Complementar à força de sensibilidade, Haller postulou uma força de irritabilidade determinada por estímulos prévios.⁹⁰

⁸⁶ Ibidem.

⁸⁷ Ibidem.

⁸⁸ Idem, p. 19.

⁸⁹ Idem, p. 20.

⁹⁰ Ibidem.

Ideias como as de Haller e Whytt foram incorporadas no século seguinte ao movimento alemão dos *Naturphilosophen*. Segundo estes filósofos, corpos vivos são possuidores de “força vital” (*Lebenskraft*), que os sustenta e controla. Esta força é um fenômeno emergente, sustentado pela organização complexa da matéria orgânica. Nesse contexto, a *Lebenskraft* perde a roupagem científica do século anterior e adquire características místicas e idealistas. O químico alemão Justus Von Leibig trouxe então as forças vitais de volta ao domínio da ciência, voltando a sustentar que tais poderes são forças análogas à gravidade como entendida na época. O caráter emergente das mesmas (as forças vitais só aparecem em matéria organizada “organicamente”) e a sua irredutibilidade a forças mais básicas permaneciam, porém. Portanto, Leibig deixou a questão como os *Naturphilosophen* a encontraram.⁹¹

O *zeitgeist* científico na Alemanha do século XVIII, portanto, era desfavorável à concepção do domínio físico como um sistema fechado a influências externas. Avanços muito significativos foram necessários para que a conclusão de Leibniz sobre a possibilidade de interações entre domínios fosse retomada. Foi preciso que, no século seguinte, o físico e fisiologista alemão Hermann Von Helmholtz unisse indícios até então isolados para que uma nova perspectiva de conservação prevalecesse. O físico inglês James Joule havia demonstrado a equivalência da energia mecânica e do calor: uma determinada quantidade de calor sempre é produzida pelo gasto de uma determinada quantidade de energia mecânica (quando um gás é comprimido, por exemplo) e vice-versa (quando um gás quente empurra um pistão). Conclui-se que uma mesma quantidade permanece em várias dinâmicas físicas. Daí foi possível elaborar uma física onde forças não conservadoras como a fricção (quando um corpo é desacelerado por fricção, nenhuma força capaz de acelerar o corpo novamente parece ser acumulada) só o são aparentemente, já que a energia cinética que parece perdida pelo sistema se encontra “preservada” pelo calor adicionado ao objeto que apresenta resistência. Uma única quantidade manifesta em diferentes roupagens em processos naturais diversos: tal era a ideia apta a ser generalizada pelos estudos de Helmholtz. Somada às contribuições de Hamilton, D’Alembert, Lagrange e Laplace, que desenvolveram as ferramentas matemáticas

⁹¹ Idem, p.24-25.

para descrição de sistemas de partículas, ela deu a Helmholtz uma base para um princípio universal de conservação de energia.⁹²

Helmholtz elaborou então um trabalho intitulado *Über die Erhaltung der Kraft (Sobre a Conservação de Força)*, em que, entre outros tópicos, demonstra que a energia mecânica (energia cinética – a energia que um objeto possui devido a seu movimento, ou o trabalho necessário para acelerar um corpo de determinada massa da inércia a sua velocidade atual – somada à energia potencial, isto é, a energia armazenada em um sistema e que pode ser convertida em outras formas de energia) total de um sistema de partículas em interação permanece a mesma quando todas as forças são as conhecidas forças “centrais”, que independem de tempo e velocidade. Helmholtz mostrou também que existem quantidades preservadas quando agem forças elétricas e magnéticas, análogas ao calor e à energia mecânica. Helmholtz então estendeu essa perspectiva de conservação de energia aos seres vivos em geral. Segundo o novo ponto de vista, os organismos vivos respeitariam a todo tempo a conservação da soma de energia cinética e potencial, estando unidos a todo o resto da natureza pelos mesmos princípios mecânicos. Todas as forças em ação nos seres vivos podem ser reduzidas às forças mais básicas que preservam energia no nível microscópico. Infelizmente, Helmholtz não utilizou nenhum argumento cientificamente decisivo para estabelecer a continuidade dos organismos vivos com o resto do universo físico. As evidências apresentadas são de caráter indutivo. Ao apresentar um conjunto de forças redutíveis a forças mais básicas e conservadoras de energia, Helmholtz esperava tornar verossímil a hipótese de seres vivos despossuídos de qualquer característica fisicamente excepcional ou “emergente”.⁹³ Ainda que inconclusivo, o argumento é reforçado por considerações adicionais que David Papineau enxerga de relance em Helmholtz. Ei-las: assume-se em primeiro lugar a conservação de energia. Se existem forças vitais/mentais fundamentais, e estas conservam energia, é preciso que haja evidências de que elas o fazem como as demais. Desafortunadamente, tais dados não existem. Não existe nenhuma motivação independente para crer que tais forças se comportam como as demais. Portanto, postular forças emergentes fundamentais significa enfraquecer os

⁹² Idem, p. 20-23.

⁹³ Ide, p. 22-30.

fundamentos indutivos da conservação de energia.⁹⁴ Infelizmente, um emergentista poderia argumentar que o comportamento conservador das forças investigadas por Helmholtz constitui evidência indutiva *a favor* da ideia de que as forças mentais também são conservadoras. De fato, uma vez que a tese da conservação de energia não nos diz quais são as forças mais básicas no universo, a existência de forças mentais fundamentais fica em aberto. Tudo o que a tese exige é que, sejam quais forem as forças, elas têm de operar de forma conservadora. O que isto de fato exclui é a existência de um fator “espontâneo” como o de Haller. Tais forças indeterminísticas são de fato incompatíveis com as ideias de Helmholtz; este é o caso por que não faz sentido afirmar que perdas de energia cinética são compensadas por acumulações de energia potencial e vice-versa, se não existe uma lei física a regular os ditos acúmulos ou perdas. Portanto, resta ao emergentista afirmar que suas leis *sui generis* são conservadoras e determinísticas, como fez Whytt.⁹⁵

Neste estágio, seria injusto retratar o emergentismo como uma excentricidade, um capricho ou uma superstição. O emergentismo tampouco era uma peculiaridade dos *naturphilosophen* alemães. Somente no século XX deram-se os passos decisivos rumo a uma união da química, e conseqüentemente, da biologia, com a física de partículas. As leis que regem as ligações químicas eram desconhecidas, o que tornou plausível uma perspectiva emergentista também no mundo anglo-saxônico. Trinta anos após a morte de Mill, em 1909, podia-se encontrar asserções como as que se seguem, feitas pelo químico Walter Nernst:

Somos forçados a admitir que no período em questão [1865-1905] não apareceu uma resposta para a questão [sobre a natureza da ligação química] que realmente nos diga mais do que podemos ver com nossos próprios olhos. Parece muito provável que devemos admitir a existência não somente de forças elétricas e, portanto, polares, mas também de forças não polares um tanto assemelhadas à gravidade newtoniana.⁹⁶

Seguindo-se ciência de então, formou-se o chamado Emergentismo Britânico.⁹⁷ Trata-se de uma filosofia da natureza como um sistema de “camadas”. Estas camadas se distinguem pelo grau de complexidade na

⁹⁴ Idem, p. 29.

⁹⁵ Idem, p. 25.

⁹⁶ Citado em McLaughlin, B. *The Rise and Fall of British Emergentism*, p. 55. Minha tradução.

⁹⁷ O'Connor e Wong, *Emergent Properties*, p. 1-2

organização da matéria. A física fundamental está no nível mais “básico”, vindo em seguida química, biologia e ciências sociais.⁹⁸ Cada nível é estudado por meio de uma ciência específica, que esclarece as leis e/ou propriedades exclusivas do domínio. Omovimento tomou forma com a filosofia de John Stuart Mill, que definiu propriedades emergentes da seguinte maneira:

Todos os corpos organizados são compostos por partes, similares àquelas a compor a natureza inorgânica, e que existiram mesmo em um estado inorgânico; mas os fenômenos da vida, que resultam da justaposição das ditas partes de uma determinada maneira, são totalmente diversas [*bear no analogy*] dos efeitos que seriam produzidos pela ação das substâncias constituintes, quando consideradas como meros agentes físicos. Não importando o quanto imaginemos a extensão e aperfeiçoamento do nosso conhecimento das propriedades dos ingredientes vários de um corpo vivo, é certo que nenhum relato das ações isoladas destes elementos poderá jamais valer como a ação do corpo vivo em si.⁹⁹

Mill defendia, então, que os princípios constitutivos da química e da fisiologia eram irredutíveis à física fundamental. Esta foi a oportunidade para a introdução, pelo escocês, dos conceitos de *lei heteropática* e *lei homopática*. Leis mecânicas onde o efeito da combinação de várias causas é idêntico à soma das ações de todas as causas atuando isoladamente são leis homopáticas. Mill considerava leis de adição vetorial de forças como bons exemplos: o efeito total de duas forças F e G agindo simultaneamente sob uma partícula *p* nada mais é do que o efeito de F sobre a dita partícula seguido pela ação de G sobre a mesma.¹⁰⁰

A conjunção de causas na química, por sua vez, é um fenômeno onde a ação conjunta de múltiplas causas não é a soma dos efeitos das causas consideradas isoladamente. Temos na química, por exemplo, que:



O resultado da reação química não é neste caso a soma dos efeitos dos reagentes, um ácido e uma base. Mill chegou à conclusão de que estes efeitos são o produto das mencionadas leis heteropáticas.¹⁰¹ Como se pode deduzir, estas seriam as leis chamadas mais tarde de “leis emergentes” (o termo “emergente” foi

⁹⁸ Idem, p. 9

⁹⁹ Mill, J.S. *A System of Logic*, Livro III, capítulo 6, 1. Minha tradução.

¹⁰⁰ O'Connor e Wong, *Emergent Properties*, p.2.

¹⁰¹ Idem, p.3.

introduzido pelo filósofo George Henry Lewes em 1875).¹⁰² Os diferentes tipos de leis são compatíveis porque as leis físicas são de caráter *ceteribus paribus* ou porque não excluem a ação de forças não físicas, como vimos em nossa discussão do panorama newtoniano.¹⁰³

Mill escreveu seu *A System of Logic* em 1843; várias décadas mais tarde, em 1920, o filósofo britânico-australiano Samuel Alexander aplicou a perspectiva emergentista ao estudo da mente. Alexander, deixando de lado a ideia de forças emergentes fundamentais, chama atenção para “qualidades” fundamentais a emergir de processos neurais. Estas qualidades caracterizam a atividade da mente, não podendo ser reduzidas às propriedades das camadas de maior simplicidade. Temos aqui, portanto, fatos tão “brutos” como as leis elementares da física.¹⁰⁴ As características irreduzíveis da matéria viva são ao mesmo tempo causalmente eficazes e ininteligíveis em termos de propriedades das camadas inferiores; ainda assim, se comportam segundo as exigências da física mais básica.¹⁰⁵

O emergentismo britânico atingiu seu apogeu com *The Mind and its Place in Nature*, de C.D. Broad. Ali, voltam as leis fundamentais emergentes, que conectam as propriedades irreduzíveis dos níveis diversos de uma realidade estratificada. Broad introduz a ideia de leis “transordinais” (*trans-ordinal laws*) – leis que conectam as propriedades características das diferentes camadas de sua ontologia. Como seria de se esperar, o caráter destas leis é definido como o oposto daquelas utilizadas na descrição de dinâmicas internas a uma mesma camada (leis intraordinais [*intra-ordinal laws*]). Mais tarde, ao discutirmos as ideias de Fodor, veremos que as leis trans-ordinais de Broad se parecem bastante com as mais modernas *bridge laws* empregadas em uma influente teoria do reducionismo.

O emergentismo permaneceu entrincheirado na academia ainda no início do século XX, mas em meados daquele século esta doutrina perdeu grande parte de seu apelo. Este fato se deve aos avanços científicos na compreensão de processos bioquímicos e neurofisiológicos. Não é coincidência que *Mind and its Place in Nature*, de 1925, tenha sido o último grande trabalho nesta tradição. Neste

¹⁰² Idem, p.1.

¹⁰³ Idem, p.3.

¹⁰⁴ Alexander, *Space, time, and deity*, vol.2, p. 46-7

¹⁰⁵ O'Connor e Wong notam a contradição. Ver *Emergent properties*, p. 7-8.

período, os sucessivos triunfos da mecânica quântica tornaram possível a união entre as ciências que parecia improvável para os emergentistas.¹⁰⁶ A mecânica quântica tornou possível explicar ligações químicas por meio da força eletromagnética. As propriedades dos elétrons esclarecem o que era até então inexplicável para os químicos. Daí advêm avanços fundamentais na biologia molecular. Uma vez identificados os ciclos bioquímicos básicos, a constituição e papel das enzimas e a estrutura das proteínas, tornou-se possível verificar a presença de forças emergentes dirigindo tais processos. O mesmo se deu com as investigações neurofisiológicas do sistema nervoso e dos mecanismos elétricos da atividade neuronal. Eis então o fato fundamental: Após décadas de pesquisa, nenhum indício de forças peculiares foi revelado. Assim, se o emergentismo pode ser considerado uma hipótese científica séria, como quer Brian McLaughlin, ele tem que ser abandonado, pois seu conteúdo empírico positivo implicaria a saliência destas forças.¹⁰⁷ Pode-se concluir, portanto, que organismos vivos obedecem as mesmas leis que todo o resto do universo, e apenas estas leis. Os obscuros fatores exclusivamente mentais tornaram-se então uma relíquia.¹⁰⁸

3.3. O reaparecimento da questão no século XX

Mesmo com o abandono do cartesianismo na época de surgimento da filosofia analítica, a questão da causalção mental não poderia ter simplesmente ressurgido como um obstáculo para o intercâmbio entre coisas pensantes e entes extensos. São dois os caminhos paralelos que conduzem à volta do debate: a reação à teoria da identidade mente-cérebro (*Identity Theory of Mind*)¹⁰⁹ e as respostas à teoria do monismo anômalo de Donald Davidson;¹¹⁰ este, por sua vez, também foi proposto, posteriormente, como uma resposta à primeira teoria.¹¹¹

¹⁰⁶ *The Rise and Fall of British Emergentism*, p. 88-89.

¹⁰⁷ *Idem*, p. 89-90.

¹⁰⁸ *Idem*, p. 30-32.

¹⁰⁹ Kim, J. *Philosophy of Mind*, p. 81-114 (Kim, 2006).

¹¹⁰ Marras, A. *Methodological and Ontological Aspects of the Mental Causation Problem*, (Marras, 2003) p. 243-244.; Davidson, D. *Actions, Reasons and Causes* (Davidson, 1963).

¹¹¹ Davidson, D. *Mental Events*, p.116-117 (Davidson, 1970).

O progresso avassalador da ciência natural no século XIX e início do século XX inspirou em positivistas como Carl Hempel otimismo em relação a uma possível “Ciência da Natureza” a abarcar toda a realidade conhecida, inclusive a condição humana. Hempel argumentava que o modelo científico dedutivo-nomológico deveria ser aplicado a disciplinas como a História, Sociologia e todas as demais ditas “ciências humanas”. A réplica de historicistas como William Dray, por outro lado, girava em torno da autonomia das ciências do homem; fatos históricos e ações racionais estão fora do alcance das ciências exatas e só podem ser abordadas através de uma interpretação hermenêutica.¹¹² Vale observar que o que interessava a ambas as correntes era examinar a possibilidade de leis empíricas capazes de ligar o pensamento à ação. As relações entre eventos mentais e eventos cerebrais não estavam em pauta. Logo, a agenda em questão era atual em um sentido e datada em outro: discutia-se a possibilidade de sistematizar as ideias da *folk psychology* de modo a elaborar uma psicologia rigorosa (um projeto atual), sem maiores preocupações com a questão mente-corpo.¹¹³

Um estranho no debate, o filósofo inglês Gilbert Ryle, procurou, em seu livro *The Concept of Mind*, demonstrar que ambos os lados na discussão compartilhavam um pressuposto que inviabilizava uma solução satisfatória. Esta premissa nociva era, segundo Ryle, um resquício de cartesianismo no imaginário dos filósofos da mente. Insistir em debater o problema com esta perspectiva seria sempre inútil, pois o erro não era só um mero detalhe, e sim um “erro de categoria”: a subsunção de um conceito em uma categoria à qual ele não pertence, de tal maneira que as suas verdadeiras propriedades se tornam inacessíveis ao intelecto. Ryle escreveu: “Eu espero provar que [o cartesianismo residual] é inteiramente falso, e falso não em detalhe mas em princípio”.¹¹⁴

A investida de Ryle inclui uma explicação mais detalhada do que é um erro de categoria, e podemos resumi-la brevemente aqui. Ryle pede ao leitor que imagine um estrangeiro ingênuo em uma visita à Universidade de Oxford, onde ele próprio lecionava. O guia lhe mostra diversas facetas do campus, como as bibliotecas, os laboratórios, as salas de aula etc. No fim da visita, o visitante,

¹¹² Marras, 2003, p. 243-244.

¹¹³ Paipineau, *The Rise of Physicalism*, p. 4.

¹¹⁴ Ryle, 1949, p. 34. Minha tradução

decepcionado, pergunta ao guia onde está a Universidade de Oxford. Evidentemente, não existe tal instituição pertencente à mesma categoria das instalações visitadas. O estrangeiro se engana ao se referir à Universidade de Oxford como um membro de um determinado tipo lógico, que em verdade lhe é estranho. Eis um exemplo de “erro de categoria”: O turista procurava “pelo tipo errado de coisa”.¹¹⁵

Qual seria, então, o erro de categoria presente no debate entre positivistas e historicistas? É ilegítimo, dizia Ryle, abordar a mente, seja ela física ou não, como um conjunto de eventos discretos, distintos entre si e do corpo com que interagem. Estes eventos parecem pertencer a processos que ocorrem segundo leis deterministas. No entanto, isto não é possível, porque um corpo físico, onde tais eventos de fato ocorrem, não pode ter seu comportamento determinado por razões. Isto é demonstrado pelo *insight* de Hume sobre a conexão conceitual de razões e comportamento. A razão para um determinado ato, como ler um livro, não pode ser compreendida de forma totalmente independente do comportamento para o qual aponta – neste caso, a ação de ler o livro. Causas têm que ser, ao contrário, concebíveis como distintas do efeito que produzem. O princípio ativo de uma determinada droga, por exemplo, é algo totalmente diverso da cura por ele proporcionada. Sendo assim, não pode haver relações de causa e efeito entre mente e comportamento e entre diferentes estados mentais. A crença em tal processo é o que Gilbert Ryle jocosamente chamou de “Mito do Fantasma na Máquina”¹¹⁶ ou “a concepção paramecânica da mente”.¹¹⁷ A mente certamente existe, mas não pertence absolutamente à categoria dos processos causais, onde encontramos o funcionamento do corpo. Processos mentais simplesmente não se encaixam no molde dos processos corporais e físicos como um todo. O uso deste modelo obsoleto leva à concepção do corpo como uma máquina governada por um tipo de máquina espectral.¹¹⁸ Eis o cartesianismo supracitado. Seu abandono traz a interessante consequência, diz Ryle, de tornar quimérico o problema do

¹¹⁵ Idem, p. 34-35.

¹¹⁶ Idem, p.34.

¹¹⁷ Ibidem.

¹¹⁸ Idem, p. 36.

livre-arbítrio, onde se aponta a dificuldade de conciliar a concepção paramecânica da mente com a crença de que a mente não é uma máquina determinística.¹¹⁹

A reação de Davidson à crítica dos Ryleanos à concepção tradicional da mente também acabou em alívio para os que intuem o livre-arbítrio, mas passa por um caminho bastante diverso, rechaçando os ataques de Ryle e pondo novamente em evidência a “concepção paramecânica”. Davidson, em resumo, procura, em *Actions, Reasons and Causes*, artigo de 1963,¹²⁰ demonstrar que razões podem, ao contrário do que pensava Ryle, ser causas. Este artigo colocou o pensamento de Davidson no mapa da filosofia da ação e gerou uma concepção abrangente a lidar em mais detalhe com o próprio problema do livre-arbítrio e com questões como a *akrasia*.

No que consiste a refutação levada a cabo por Davidson? Já vimos que Ryle argumenta que não pode haver relações causais de razões para atos porque a conexão entre os fatores em questão é lógica, e não empírica. Conexões causais são sempre contingentes e a concepção paramecânica exige mais. Davidson procura mostrar no artigo supracitado que a conclusão é prematura. Ele argumenta que relações causais unem eventos, não importando a forma como os descrevemos. As relações lógicas apontadas pelos ryleanos, por sua vez, aparecem apenas em *determinadas descrições* de eventos. O filósofo Alfred Mele exemplifica isto da seguinte forma: se x , o impacto sobre um sino, causa y , o bater do mesmo, nossa descrição de x como “a causa do bater do sino” gera o enunciado “a causa do bater do sino causou o bater do sino” – um enunciado analítico.¹²¹ Ainda assim, isto não muda o fato de x ter causado y . O famoso exemplo de Davidson evoca um cenário onde alguém entra em um quarto escuro e liga a luz para, suponhamos, procurar suas chaves.¹²² A ação de acender a luz é também a ação de alertar um ladrão que estava dentro da casa. A conexão lógica entre a intenção de acender a luz e acender a luz de fato existe, mas desaparece quando substituímos a descrição “acender a luz” por “alertar o ladrão”. Existem ainda incontáveis descrições do ato que não envolvem conexão lógica com a razão

¹¹⁹ Idem, p. 37.

¹²⁰ Davidson, 1963.

¹²¹ Mele, 2003, p. 68.

¹²² Idem, p. 1.

alegada pelo agente – acordar o gato, espantar a mariposa na parede, aumentar ligeiramente a temperatura do recinto etc. A conexão conceitual só aparece quando queremos racionalizar o que o agente fez. Mesmo neste caso Davidson duvida da tese de que relações causais não podem ser lógicas, somente empíricas. Se supusermos que “x causou y” é verdadeiro, temos que a causa de y = x; o enunciado obtido, “a causa de y causou y”, é analítico. Como no exemplo do sino, temos descrições logicamente dependentes uma da outra, mas sem excluir relações causais.

Davidson também intenciona refutar a alegação de Hart e Honoré de que razões não são causas por não exemplificarem leis. A objeção à concepção usual de ação nesse caso vem da crença de que “enunciados causais singulares implicam generalizações”,¹²³ ou seja, leis que subsumem os eventos analisados; “leis estão essencialmente envolvidas em explicações causais comuns, mas não em racionalizações.”¹²⁴ Em sua refutação, Davidson admite que sempre existe uma lei a reger eventos ligados causalmente. As leis em questão, no entanto, não são necessariamente descritas da mesma maneira como expressamos os eventos em questão. É possível explicar ações através de razões sem que tenhamos a menor ideia de qual lei determina o resultado. É seguro afirmar que o impacto de uma pedra quebrou a janela, e que existe alguma descrição em termos físicos do evento. Esta descrição mostra que os eventos tinham que ocorrer, mas a maior parte das pessoas não tem uma ideia clara de como formulá-la. Com ações em geral ocorre o mesmo. Davidson tinha certeza de que existe uma lei que conecta a intenção (apropriadamente descrita) de acender a luz com os movimentos corporais que de fato acendem a luz. Não sabemos que lei é esta, mas isto não significa que a relação em questão não é causal.

A refutação dos argumentos anticausalistas é complementada com uma concepção positiva da relação entre razões e ações, estas últimas definidas como movimentos corporais produzidos por complexos de desejos e crenças. Razões explicam atos por meio do que Davidson denomina “racionalização”. Este tipo de explicação causal “revela o objetivo ou propósito dos *explananda*.”¹²⁵ Razões são

¹²³ Idem, p. 6. Minha tradução.

¹²⁴ Ibidem. Minha tradução.

¹²⁵ Mele, 2003, p. 70. Minha tradução.

causas no sentido em que podemos entender por que determinada mudança ocorreu, mas são também um tipo especial de causa, já que nos levam a compreender a ação como uma atitude racional. Davidson alega que não se pode explicar uma atitude com meras menções da atratividade da mesma para o agente. É preciso mostrar que aspecto da atitude era atraente para o agente. Dito de outra forma, compreender uma ação é perceber a propriedade em virtude da qual a ação era razoável para quem a perpetra. Quando isto é feito, temos uma explicação causal em forma de *racionalização*.¹²⁶ Quando existe uma ação racional, há em primeiro lugar um agente possuidor de um tipo de “pro atitude” (*pro attitude*)¹²⁷ relativa a ações dentro de uma determinada categoria. O termo “pro atitude” engloba tanto aspectos permanentes do caráter do agente, como “amar crianças” ou “preferência por companhia barulhenta”¹²⁸ quanto caprichos passageiros quaisquer que o agente satisfaz conscientemente (um desejo súbito de tomar sorvete, por exemplo). O que se tem aqui é uma vasta gama de

desejos, vontades, anseios (*urges*), instigações, e uma grande variedade de perspectivas morais, princípios estéticos, preconceitos econômicos, convenções sociais, e objetivos públicos e privados que possam ser interpretados como atitudes de um agente relativas a ações de um certo tipo.¹²⁹

Evidentemente, estas atitudes não são necessariamente convicções de que todas as ações do tipo são razoáveis. Davidson exemplifica isto citando um homem que possui ao longo de toda a sua vida vontade de beber uma lata de tinta. Mesmo no momento em que finalmente cede ao desejo, ele não precisa pensar que fazê-lo é uma boa ideia. Isto significa que racionalizações mostram a razoabilidade de atos de uma maneira altamente subjetiva.

À pro atitude do agente relativa a uma classe de atos combina-se uma crença de que uma determinada ação pertence àquela classe. A conjunção de pro atitude e crença que encaixa uma ação em uma categoria de alguma forma desejável é chamada em *Actions, Reasons and Causes* de “razão primária”.¹³⁰ Uma ação é causada por sua razão primária, e leva à sua compreensão de uma forma peculiar, a já mencionada racionalização. Racionalizar uma ação significa localizar uma

¹²⁶ Davidson, 1963, p.1.

¹²⁷ Ibidem

¹²⁸ Ibidem

¹²⁹ *Actions, Reasons and Causes*, p. 1.

¹³⁰ Ibidem

razão primária *R* que explica por que um agente tomou uma atitude *A* sob a descrição *d* (é preciso mencionar uma descrição específica porque, citando novamente o exemplo de Davidson, a vontade de acender a luz racionaliza porque o dono da casa acendeu a luz mas não o seu alerta ao ladrão). *R* explica *A* por consistir em uma atitude proposicional (desejo ou crença) em relação a ações que satisfazem determinada propriedade, juntamente com a crença de que *A*, sob a descrição *d*, satisfaz tal propriedade.¹³¹ A noção davidsoniana de racionalização, é portanto, uma noção instrumental.¹³²

A refutação do argumento humeano dos seguidores de Ryle é considerada bem-sucedida pela maior parte dos atuais filósofos da mente. No entanto, ela foi apenas a fase inicial do projeto de Davidson de esclarecer o lugar da mente no universo. Tendo refutado a alegação de que mentes não podem ser causas, Davidson procura encaixar sua teoria do agir em um contexto metafísico, a saber, a estrutura causal da realidade. Esta tarefa é o foco do artigo *Mental Events*.¹³³ Neste escrito, o filósofo argumenta que, embora todo evento mental seja um evento físico, não é possível elaborar leis deterministas (leis estritas, *strict laws*) conectando eventos neurais a atitudes proposicionais. A tese resultante, a do “monismo anômalo”, é o núcleo de toda a filosofia da mente de Davidson. É a partir desta tese que Davidson desenvolve sua concepção de seres humanos como agentes livres e ao mesmo tempo integrantes do mundo natural. Isto se deve a uma característica única da mente humana, a conexão férrea entre razão e pensamento. De forma muito resumida, o argumento de Davidson é que existe um indelével grau de racionalidade e coerência em qualquer agente humano, mesmo que estúpido ou insensato. O mundo da física, no entanto, não apresenta o mesmo tipo de exigência. Da incompatibilidade de certas propriedades dos dois domínios se deduz a impossibilidade de conectá-los por leis científicas. Evidentemente, não basta listar propriedades contrastantes de dois campos de conhecimento para demonstrar que eles são incomensuráveis neste sentido. É preciso mostrar boas razões para a adoção de uma tese do gênero. Doravante, teremos como chaves para a discussão de *Mental Events* os comentários de Jaegwon Kim, juntamente, é claro, com o próprio texto de Davidson.

¹³¹ Idem, p. 2.

¹³² Mele, 2003, p. 72.

¹³³ Davidson, 1970.

A razão pela qual os terrenos do físico e do mental são incomensuráveis é análoga àquela que torna inferências de valor econômico a partir de características físicas uma falácia. A falácia conhecida pelos economistas como “falácia física” é a asserção de que existe um valor fixo para cada produto, valor este determinado pelas suas características materiais, independentemente da sua dinâmica de oferta e demanda. É com base nesta falácia que se crê que o preço de determinado produto é abusivo porque “ ele não passa de um pedaço de couro”, etc.¹³⁴ Enunciados análogos são falaciosos porque 1) o ambiente econômico que determina os preços está necessariamente conectado a considerações de disponibilidade, demanda e especulação que são totalmente estranhas ao domínio físico e 2) não é (pelo menos em trocas comerciais normais) possível traduzir “justo” em valores utilizáveis em trocas econômicas. Resumidamente: Não é possível traduzir a linguagem de um campo para a do outro. De forma similar, sistemas intencionais têm características essenciais que seriam violadas caso houvesse leis a conectá-los com o domínio físico, e vice-versa. Conectar disciplinas de forma nomológica é, de certa maneira, transmitir certas características de um campo para o outro. Leis psicofísicas não são possíveis porque características incompatíveis com as disciplinas relacionadas teriam que ser “exportadas” para as mesmas.¹³⁵

Vejamos agora, adaptando exemplos de Kim,¹³⁶ como funciona a demonstração de Davidson. Pensemos em um agente humano qualquer que crê que o Rio de Janeiro fica a menos de 70 km de Niterói. Se o agente em questão está ciente deste fato, fará sentido lhe atribuir a crença de que as duas cidades distam menos de 80 km, 90 km, 1200 km e daí em diante. Atribuir-lhe ou mesmo considerar as crenças opostas, porém, seria sempre absurdo se supormos que o agente entende o enunciado e conhece aritmética elementar. Nas palavras de Kim, o conceito de crença “requer que o sistema de crenças seja coerente como um todo – ele deve em certo sentido ser bem-comportado (*hang together*), sem lacunas inexplicadas”.¹³⁷ O intérprete sensato culpa sua própria interpretação quando não consegue achar um mínimo de coerência nas crenças de seu interlocutor. É

¹³⁴ Ver Pinker, S. *Tábula rasa. A Negação Contemporânea da Natureza Humana*, p. 322.

¹³⁵ Kim, *Psychophysical laws*, p. 201.

¹³⁶ *Philosophy of Mind*, p. 184-187

¹³⁷ Idem. P. 185, minha tradução.

inaceitável afirmar que alguém possui crenças que muito obviamente se contradizem. O segundo exemplo de Kim tem a ver com uma situação mais comum: pergunta-se a um interlocutor qualquer se ele aprova o governo, digamos, de Barack Obama, obtendo-se a resposta “Sim e não”. Diante da mesma, ninguém se sente tentado a assinalar ao respondente uma crença contraditória, onde Obama é e não é um bom presidente no mesmíssimo sentido. O único caminho racional aqui é interpretar a resposta como significando “gosto de certas ações de Obama e não de outras”. Kim observa que, se a pessoa em questão insistisse em afirmar que realmente “gosta e não gosta de Obama, e ponto final”, o intérprete procuraria uma explicação que elimina a contradição aparente, ao invés de incorporá-la. O aparente crítico/admirador de Obama talvez não saiba se expressar na língua que tentava usar, ou talvez não sabe o uso da negação no mesmo idioma etc.¹³⁸ Faz mais sentido localizar o absurdo na interpretação dos enunciados do que nos pensamentos do agente. A teoria empregada na interpretação deve ser ajustada de forma a tornar consistente aquilo que é dito.¹³⁹ É exatamente assim que podemos inteligimos os quase-paradoxos da literatura. Ivan Turguêniev escreveu em *Pais e Filhos*: “...”todos sentiam-se um pouco constrangidos, um pouco entristecidos mas, no fundo, sentiam-se muito bem”.¹⁴⁰ Igualmente Philip Roth e seu jovem açougueiro *kosher* em *Indignação*:

Lembro-me desses sete meses como um tempo maravilhoso – maravilhoso exceto pelas horas em que tinha que eviscerar as galinhas. E até isso era de certo modo maravilhoso, por ser alguma coisa que eu fazia, e fazia bem, mesmo não gostando de fazer.¹⁴¹

Está claro que o açougueiro obtém satisfação com um aspecto do seu trabalho que é diferente daquele que lhe causa repulsa; de outra forma teríamos de concluir que a passagem descreve uma pessoa louca. Esta exigência de racionalidade pode ser generalizada para outras atitudes proposicionais, como desejos; afinal, seria ridículo afirmar que alguém *quer* e no mesmo sentido *não quer* comer risoto no jantar (por exemplo, o sabor do risoto em si seria um deleite e também excruciante), e daí em diante. Daniel Dennett procurou reforçar este

¹³⁸ Ibidem.

¹³⁹ Davidson, *Mental Events*, p.123.

¹⁴⁰ Turguêniev, 2004 [1862], p. 293.

¹⁴¹ Roth, 2009, p.15.

ponto, chamando atenção para a convergência das exigências biológicas e racionais:

Uma espécie poderia “experimental” por mutação qualquer número de crenças ineficazes, mas nenhum desses sistemas mereceria ser chamado de sistema de crenças justamente por causa de seus defeitos, sua falta de racionalidade e, logo, um sistema de crenças falsas seria uma impossibilidade conceitual. Para tomarmos emprestado um exemplo de um conto de MacDonald Harris, um peixe solúvel seria uma impossibilidade evolutiva, mas um sistema de crenças falsas não pode nem mesmo ter uma descrição coerente.¹⁴²

É oportuno ecoar esta passagem com o texto do próprio Davidson:

Atribuir a pessoas um alto grau de consistência não é mera boa vontade: é inevitável fazê-lo se quisermos estar em posição de fazer acusações consistentes de erro e de certo grau de irracionalidade. Confusão global, e erros universais são coisas impensáveis, não por um recuo de imaginação, mas porque confusão demais não deixa nada para ser confundido e erros sistemáticos apagam o sustentáculo de crenças verdadeiras contra o qual falhas podem ser erigidas.¹⁴³

Estados intencionais, portanto, obedecem a princípios de racionalidade que garantem um conjunto de disposições superficialmente consistentes. Esta coerência mínima é uma característica essencial ou constitutiva daquilo que chamamos de “mente”. Dito de outra forma, estados mentais intencionais têm que ter alguma interpretação que os tornem inteligíveis para aqueles que o assim os denominam. Como diz Jaegwon Kim,

Crenças, intenções e coisas afins [*the rest*] são possíveis somente enquanto elementos de um sistema integrado e “holístico” e o que dá ao sistema uma estrutura inteligível são os princípios de racionalidade, consistência e coerência. Para Davidson, as normas de racionalidade e coerência que fundamentam o holismo mental, são os “princípios constitutivos” da mentalidade; eles dão à mentalidade intencional sua identidade distinta de domínio autônomo.¹⁴⁴

A relevância da generalização acima está na inexistência de delimitação semelhante no campo da física. Suponhamos que haja um mapeamento perfeito do funcionamento do cérebro humano, e também leis psicofísicas estritas¹⁴⁵ conectando eventos neurais a atitudes proposicionais. Com a ajuda de um

¹⁴² *Sistemas Intencionais*, p. 50.

¹⁴³ *Mental Events*, p.122-123. Minha tradução.

¹⁴⁴ Kim, 2003, p. 119. Minha tradução.

¹⁴⁵ Leis estritas são aquelas que: 1) valem sempre, ao contrário de leis *ceteribus paribus* (generalizações que não são necessariamente falsificadas por exceções), e 2) são elementos de um conjunto de leis que forma uma teoria completa de um determinado campo de conhecimento. As leis mais fundamentais da física, sejam eles quais forem, são leis estritas (Idem, p. 116)

escaneamento cerebral em tempo real, obter-se-ia conhecimento exato do que creem agentes em um dado momento. Infelizmente, não existe neste cenário nada que impeça que o dito agente tenha crenças ou desejos contraditórios no sentido acima discutido. Para que isto acontecesse, afinal, bastaria que seu cérebro seditasse um evento neurológico E1, ligado nomologicamente (por uma lei científica) a uma crença *p* (Necessariamente, E1 aparece em um sujeito *S* em *t* se e somente se *p* aparece em *S* em *t*), e um segundo evento E2, ligado nomologicamente a *nãop* (Necessariamente, E2 aparece em um sujeito *S* em *t* se e somente se *nãop* aparece em *S* em *t*). Se não existe nada que impeça semelhante absurdo em teorias do mundo físico, estas não bastam para explicar a mente; não pode haver redução do mental ao físico, nem uma correlação nomológica estrita entre ambos. Nas palavras de Davidson, “Não existem leis psicofísicas estritas devido aos compromissos contrastantes dos esquemas físico e mental”. Outra forma de compreender o que Davidson quer dizer é imaginar que temos as seguintes leis: Necessariamente, RN aparece em um sujeito *S* em *t* se e somente se *N* aparece em *S* em *t* e , RN2 aparece em um sujeito *S* em *t* se e somente se N2 aparece em *S* em *t* , onde RN é a crença de que o Rio de Janeiro fica a menos de 70 km de Niterói e RN2 é a crença de que o Rio de Janeiro fica a menos de 80km de Niterói, e *N* e N2 são eventos neurais distintos. Já vimos que esperamos que todo agente racional que crê em RN deve crer em RN2. Consequentemente, se houvesse leis psicofísicas estritas, teríamos duas leis adicionais: (1) Necessariamente, se RN ocorre em *t* , RN2 ocorre em *t*; e (2) Necessariamente, se *N* ocorre em *t* , N2 ocorre em *t* . A consequência inaceitável destas leis é a possibilidade de prevermos uma lei física a partir de leis psicológicas justificadas pelas exigências da racionalidade. Isto não é cabível porque significa impor normas ao domínio físico normas que lhes são estranhas – princípios de coerência em interpretação. Como observa Kim, “tais normas não têm aplicação [na ciência física]”.¹⁴⁶O conjunto de estados intencionais que sustentam a deliberação e a ação tem um irremovível caráter normativo e holístico (não se pode atribuir atitudes proposicionais uma por vez) que são totalmente estranhas ao domínio físico. Ligar os dois terrenos é impossível porque a transferência de características resultaria na convivência de propriedades incompatíveis.

¹⁴⁶ Kim, 2003, p. 120. Minha tradução.

Contrastando com as muitas discussões sobre os *qualia* como obstáculos ao materialismo radical na filosofia da mente, é a racionalidade que vem ao socorro dos antirreducionistas. De fato, existe pouco interesse em sensações e percepções na obra de Davidson. Ainda assim, ele enfatiza que sua filosofia da mente é um monismo e não um dualismo. Em sua perspectiva, todo evento mental é também um evento físico; os critérios para ambos os tipos de predicado são satisfeitos simultaneamente quando ocorre o que quer que seja que podemos chamar de “mental”. A diferença está nas *propriedades* e não nos eventos. Eventos mentais são eventos físicos em virtude de serem *eventos cerebrais* (ou computacionais, se existir IA “forte” etc.), e são mentais em virtude de possuírem, *além das primeiras*, propriedades mentais. Davidson propõe, então, uma teoria do monismo anômalo, “monismo” porque só existem eventos físicos, e “anômalo” porque “não existem leis estritas com base nas quais podemos explicar e prever fenômenos mentais”.¹⁴⁷ O mental e o físico habitam o mesmo mundo, mas contribuem para o mesmo de maneiras díspares – o primeiro não está escravizado pelas leis que governam o segundo. Evidentemente, a tese de Davidson tem que ser compatível com o truísmo de que existem correlações entre estados cerebrais e estados mentais. Não seria a realidade destas coincidências um indício de que há algo errado com os argumentos anteriores? O que se deve observar é que, mesmo que de fato todos os possuidores de uma determinada propriedade mental estejam simultaneamente em um dado estado neural, isto não bastará para estabelecer uma lei conectando as duas ocorrências. Afinal, leis se distinguem de uma mera generalização verdadeira por duas características: 1) leis são generalizações que se estendem a situações contrafactuais e condições subjuntivas, e 2) são confirmadas pela observação de exemplos.¹⁴⁸ Por conseguinte, a verdade de generalizações como “todos os possuidores de uma determinada propriedade mental M estão simultaneamente no estado neural N” não tem efeito sobre a tese de Davidson. Uma generalização deste tipo só poderia ser provada por uma observação de cada caso; isto torna a generalização uma verdade empírica altamente contingente, sem nenhum interesse filosófico.¹⁴⁹ A relação que sustenta as correlações entre propriedades físicas e mentais é uma relação assimétrica de dependência,

¹⁴⁷ Davidson, *Mental Events*, p. 124. Minha tradução.

¹⁴⁸ Idem, p. 121. Veremos mais adiante que existem condições adicionais e mais complexas a caracterizar a modalidade das leis.

¹⁴⁹ Idem, p. 199-200.

chamado por Davidson de *superveniência*. Há muito a dizer sobre o conceito de superveniência, o que será na segunda parte da tese; o que se deve dizer no momento é que a superveniência em questão é basicamente dependência ontológica do mental em relação ao físico, sem que o contrário ocorra. Mais especificamente, isto significa que não pode haver dois eventos fisicamente idênticos que diferem em um aspecto mental, ou que um objeto não pode ter um aspecto mental modificado sem sofrer alguma alteração física.¹⁵⁰

Na filosofia de Davidson, soma-se à impossibilidade de leis psicofísicas estritas a impossibilidade de leis preditivas puramente psicológicas, já que “acontecem demasiadas coisas a afetar o mental que não são uma parte sistemática do mental”.¹⁵¹ Na ausência de leis psicológicas preditivas, temos a possibilidade de *leis normativas* (regras de inferência, princípios de tomada de decisão, etc).¹⁵² Seja como for, a relativa liberdade do psicológico em relação ao físico leva a uma ontologia peculiar. O monismo anômalo de Davidson é um tipo de *token physicalism* – a classe de teorias onde todo evento mental é também um evento físico porque apresenta os dois tipos distintos de propriedades. No campo do fisicalismo, esta categoria contrasta com o *type physicalism*, em que todo evento mental é um evento físico porque propriedades mentais *são* propriedades físicas, não obstante o fato de que enunciados sobre mentes têm significado bastante diferente de enunciados neurológicos. O mundo, para Davidson, é composto exclusivamente por objetos e eventos físicos, mas alguns destes portam as propriedades não-físicas que denominamos “mentais”. Esta ontologia difere do dualismo de substância de Descartes porque eventos físicos e eventos mentais não são realidades separadas que independem uma da outra. Os eventos mentais no monismo anômalo são claramente dependentes do mundo físico, e não constituem uma classe de eventos à parte do mundo físico. A mente é uma anomalia ontológica, a combinar “dependência causal e independência nomológica”.¹⁵³ Desta forma, justifica-se o propósito de Davidson de demonstrar a eficácia causal do mental. Conclui-se que tal coisa é possível com a constatação de que todo evento mental é também um evento físico, e que não existe, portanto, um

¹⁵⁰ *Mental events*, p. 119

¹⁵¹ Davidson, *Mental events*, p. 124. Minha tradução.

¹⁵² Kim, *Psychophysical laws*, p.211.

¹⁵³ Kim, *Psychophysical Laws*, p. 199-200. Minha tradução.

obstáculo para a intervenção do pensamento no mundo. O livre-arbítrio intuído pelos teóricos da moral também mostra-se justificado, já que a deliberação não está sujeita a leis determinísticas que tornam toda ação um fato consumado. Os argumentos dominadores que assombram os defensores do livre-arbítrio são refutados com o desaparecimento da concepção “paramecânica” da mente. Nas palavras de Davidson,

Explicamos as ações livres de um homem, por exemplo, com um recurso aos seus desejos, hábitos, conhecimento e percepções. Tais explicações do comportamento intencional operam em uma estrutura conceitual fora do alcance direto de leis físicas, através da discrição tanto de causa como de efeito, razão e ação, como aspectos da caracterização de um agente humano. O anomalismo do mental é portanto uma condição necessária para que ações sejam vistas como autônomas”.¹⁵⁴

Afirmamos anteriormente que a teoria do monismo anômalo é um dos caminhos que levam a atual filosofia da mente ao problema da causação mental. Estamos agora aptos a mostrar como isto se ocorre. Davidson utiliza sua teoria para indicar como eventos mentais têm influência sobre o mundo físico. Como já vimos, eventos mentais têm este poder por serem eles mesmos eventos físicos; mais especificamente, existem eventos que compartilham propriedades físicas e propriedades mentais, e todo evento chamado “mental” pertence a esta categoria. Os efeitos gerados por estes eventos são eles mesmos ocorrências físicas adicionais. Na filosofia de Davidson, onde existe causalidade existem leis causais estritas; e conforme descrito anteriormente, toda lei deste tipo é uma lei física, nunca psicofísica ou puramente mental. Daí resulta que afirmar que um evento mental *m* causa um evento qualquer *p* significa afirmar que *m* tem uma propriedade física *Q* (ou que satisfaz um tipo físico *Q*) tal que uma lei liga *Q* (ou eventos que satisfazem o predicado *Q*) a alguma propriedade física *P* de *p*.¹⁵⁵ Se, como observa Davidson, não existem leis regendo a interação de propriedades mentais e físicas, são as leis físicas que, como diz Kim, “fazem todo o trabalho causal”. Deve-se concluir daí que eventos mentais só influenciam o mundo físico por intermédio de propriedades físicas.¹⁵⁶ Somente propriedades físicas ilustram as leis estritas que sustentam relações causais. Se alguém se apaixona por uma linda mulher e seu desejo *d* o faz discar o número dela no telefone (*t*), deve haver uma

¹⁵⁴ Ibidem. Minha tradução.

¹⁵⁵ Kim, 2006, p. 188.

¹⁵⁶ Ibidem. Minha tradução.

lei física conectando os dois eventos. Esta lei física, no entanto, exige uma re-descrição física de *d* que põe em evidência uma propriedade física ligada nomologicamente a *t*. Por meio desta re-descrição, ter-se-á uma ideia clara de como *d* pôde influenciar o mundo físico. Infelizmente, daí se conclui que *d* só causou *t* em virtude de uma de suas propriedades físicas. Se não existem leis psicofísicas estritas, isto não poderia ter sido de outra forma. Eventos mentais dependem única e exclusivamente de suas propriedades físicas para causar o que quer que seja no mundo. Enquanto desejo inebriante, *d* é totalmente inerte; a dita paixão só faz diferença no mundo por andar de carona em um evento possuidor de certa propriedade física, a saber, uma propriedade neural. Conclui-se então que a tentativa de Davidson de demonstrar a realidade da causação mental acabou por torná-la ainda mais problemática. Se propriedades mentais são causalmente inertes, como o argumento acima parece mostrar, o monismo anômalo requer um tipo de mente epifenomenal: eventos mentais são inertes *enquanto eventos mentais*. Dito de outra forma, as propriedades que justificam chamar certos eventos de “mentais” em nada contribuem para tornar os ditos eventos causas ou efeitos. Kim observa que isto significa que uma redistribuição radical das propriedades mentais no mundo físico não faria diferença alguma no mesmo. Nenhuma relação causal seria alterada.¹⁵⁷ Longe de apresentar uma teoria bem-comportada da mente como um sistema causal, Davidson dá um passo em falso e ameaça arruinar o quadro da racionalidade humana construído a partir de *Actions, Reasons and Causes*. O problema da causação mental será um obstáculo para qualquer teórico que afirme a eficácia do mental sem dar um papel causal aos eventos mentais *enquanto eventos mentais*, ou seja, mostrar como propriedades mentais influenciam a rede causal do mundo. Se não for demonstrado que eventos têm certos efeitos por possuírem esta ou aquela propriedade mental, ter-se-á, pairando, a ameaça do epifenomenalismo, ou talvez uma metafísica ainda mais estranha. Como veremos mais adiante, existem filósofos que não veem estes cenários como “ameaças”, mas procuraremos demonstrar que é exatamente isso o que são.

O segundo caminho que conduz ao problema da causação mental na filosofia contemporânea surge com a problemática em torno da *identity theory of*

¹⁵⁷ Idem, p. 189.

mind (doravante ITM, também conhecida como “teoria da identidade mente-cérebro”) dos anos 1950. Como veremos adiante, a ITM em si torna a questão da interação irrelevante, mas as críticas por ela suscitadas geraram um panorama teórico onde o reducionismo volta a ser questionado. Com isso, o problema da interação ressurgiu em nova roupagem.

O proponente mais célebre da ITM é o filósofo anglo-australiano John Jamieson Carswell Smart (1920). Em seu artigo *Sensations and Brain Processes*¹⁵⁸, Smart apresenta um esboço de teoria onde a mente como um todo é identificada com processos neurais. Neste contexto, o pensamento de Smart estava alinhado com aqueles dos filósofos Herbert Feigl e Ullin Place.¹⁵⁹ Vejamos então como Smart se coloca em relação à metafísica da mente. Por séculos tem-se observado correlações entre estados mentais e cerebrais. Como vimos, este dado já era conhecido pelo próprio Descartes, que verificou a existência de uma relação mais íntima da mente com o cérebro do que com as demais partes de nossa anatomia. A união entre a coisa pensante e o cérebro é evidenciada por “incontáveis observações”.¹⁶⁰ Smart se propõe, então, a explicar estas correlações, com a solução encontrada terminando por fazê-lo por eliminação. O caráter materialista do escrito de Smart vem da constatação de que uma mente não física teria uma posição desconfortável em uma teoria atual que abrangesse o universo como um todo. A metafísica resultante de uma conjunção de universo físico e uma “coisa fantasmagórica” (*ghost stuff*)¹⁶¹ seria inaceitavelmente desajeitada. Smart enxerga na coisa fantasmagórica um ar inverossímil de estranho no ninho. Vamos aos argumentos: o panorama de rápidos avanços científicos, mencionado na introdução deste trabalho, já parecia irresistível na época de Smart. O filósofo anglo-australiano vê a “perspectiva segundo a qual organismos podem ser vistos como mecanismos físico-químicos” como prometendo teorias capazes de explicar “o comportamento do homem em termos mecânicos.”¹⁶² As intuições e argumentos contrários à redução materialista do comportamento humano constituem, no entanto, um obstáculo a esta tese. Deixe a mente de fora e terá

¹⁵⁸ Smart 2003 [1959].

¹⁵⁹ Consultar Feigl, 2003 [1958] e Place 2003 [1956].

¹⁶⁰ Curiosamente, isso não foi sempre assim. Jaegwon Kim chama atenção para a crença grega no coração como sede das emoções.

¹⁶¹ *Sensations and Brain Processes*, p. 66.

¹⁶² *Idem*, p. 61. Minha tradução.

deixado o homem de fora. Uma mente imaterial, portanto, nos leva novamente a uma metafísica dual onde os elementos da consciência existem paralelamente a estados neurais, estando, evidentemente, correlacionados com os mesmos. O homem torna-se ele mesmo um ser dual sem paralelo no universo. Mais do que um ser especial, ele ganha *status* de milagre, um *skyhook*, como memoravelmente escreveu Daniel Dennett. Smart vê uma inverossimilhança inaceitável neste cenário. A ontologia resultante consiste em um mundo totalmente descritível em termos físicos, de um lado, e sensações pairando fora do tecido material, de outro. As correlações observadas entre estados mentais e conteúdos conscientes são explicadas por leis psicofísicas conectando as duas realidades. Estamos nos referindo, naturalmente às leis eloquentemente rejeitadas por Davidson. Eis o quadro visto por Smart como “francamente inacreditável”.¹⁶³ É importante observar que Smart visava alvejar o *dualismo* mente-corpo, e não as teorias mais sutis de *token physicalism* como aquela proposta por Davidson. Seja como for, a ontologia repudiada por Smart baseia-se em um universo físico existindo paralelamente ao que ele chama entusiasmadamente de “penduricalhos nomológicos” (*nomological danglers*).¹⁶⁴ A expressão vem de Herbert Feigl e rotula, como deve estar claro, corpos estranhos em uma metafísica de outra forma “bem-comportada”, e ligados a esta por leis suspeitamente *ad hoc*. Presume-se a existência destas últimas na metafísica atacada por Smart, porque as correlações mente-cérebro são tomadas como exemplos de interação causal entre as duas realidades, que por sua vez requerem leis causais como fundamento. Infelizmente, diz Smart, tais leis causais teriam um “cheiro bizarro”.¹⁶⁵ Smart espera que toda lei fundamental a ser encontrada com o avanço da física diga respeito a elementos simples, como partículas fundamentais. Sendo assim, é implausível supor que podemos também formular leis que governam as relações entre tais elementos simples e configurações consistindo em um número vastíssimo de neurônios. As formulações resultantes seriam aberrações nomológicas; nada comparável é conhecido nas ciências naturais como as conhecemos. Com semelhantes elementos, desfigurar-se-ia uma metafísica elegante. Smart observa então que argumentos que apontam para a existência destas leis devem sempre conter

¹⁶³ Ibidem. Minha tradução.

¹⁶⁴ Ibidem. Minha tradução.

¹⁶⁵ Ibidem. Minha tradução.

alguma “trapaça” (*catch*). Eis a “confissão de fé”¹⁶⁶ de Smart: temos que evitar uma ontologia de lés a lés física exceto por uma bizarra classe de fenômenos. A doutrina da parcimônia de Smart encontra eco em Feigl, que defende a elegância teórica como motivação para a identificação do mental com o físico.¹⁶⁷ É preciso procurar alternativas ao dualismo. Pela mesma razão, pode-se questionar a integridade do emergentismo britânico como ciência, já que as leis ali requeridas são leis inacomodáveis com o restante das leis físicas microscópicas fundamentais (vimos na nossa definição de física que esta é uma ciência que busca conceitos ubíquos e microscópicos); por conseguinte, têm que ser penduricalhos nomológicos da mesma forma que leis psicofísicas dualistas.

O caminho defendido por Smart vem de sua rejeição de uma abordagem behaviorista da consciência, compartilhada por Place.¹⁶⁸ Smart não crê que sensações podem ser eliminadas por explicação através de uma análise behaviorista que traduz a linguagem das sensações em linguagem comportamental. Quando relatamos ver uma imagem qualquer, diz Smart, tudo indica que estamos de fato fazendo um relato legítimo, ou seja, apontando para um evento real.¹⁶⁹ Esta intuição leva Smart a considerar as virtudes de uma filosofia que identifica eventos neurais com eventos cerebrais. A formulação resultante é um esboço de ITM.

Smart afirma que a identidade por ele defendida é “estrita” (*strict*). Este é o tipo de identidade afirmada em enunciados como “7 é idêntico ao menor número primo maior que 5”, ou relâmpagos são um tipo x de descarga elétrica”. Esta categoria de identidade contrasta com casos de entes contíguos no espaço ou no tempo. Consideremos o exemplo de Napoleão para ilustrar este segundo caso. É este último que temos ao dizer que o comandante do *Grande Armée* era a mesma pessoa que o garoto que cresceu em uma família empobrecida da Córsega. Este tipo de afirmação, diz Smart, envolve “fatias temporais” (*time slices*) do mesmo objeto quadridimensional.¹⁷⁰ A noção de identidade estrita, por sua vez, aparece quando identificamos o objeto quadridimensional que engloba o general e é

¹⁶⁶ Ibidem. Minha tradução.

¹⁶⁷ *The “Mental” and the “Physical”* (Feigl 2002 [1958]), p. 68-71.

¹⁶⁸ *Is Consciousness a Brain Process?* p.55

¹⁶⁹ Ibidem. Smart menciona retroimagens (*afterimages*).

¹⁷⁰ Idem, p. 62.

idêntico àquele que inclui o garoto corso. O ato de identificação é análogo às asserções científicas “água = H₂O” ou “calor = energia transferida de um corpo ou sistema para outro como resultado de temperaturas diferentes”.¹⁷¹ Estas observações apontam para um fato crucial e várias vezes enfatizado por Smart: a ITM *não* envolve afirmar que “percepção visual” ou qualquer outro evento mental *significa* o mesmo que “tipo *x* de processo cerebral”. O que de fato ocorre é a asserção de relatos perceptuais quaisquer e demais conteúdos conscientes *vêm a ser* (*hapen to be*) processos cerebrais. Quando falamos de nossas sensações, estamos falando de algo que é um processo cerebral devido a *fatos empíricos*. Feigl e Smart também procuram deixar isto claro em suas respectivas versões da ITM:

Dizer que enunciados sobre consciência são enunciados sobre processos cerebrais é manifestamente falso. Isso é mostrado (...) pela possibilidade de descrever sensações e imagens mentais sem conhecimento algum de processos cerebrais ou sem saber sequer que tais coisas existem.¹⁷²

A identificação [do mental com o físico] é justificada empiricamente, portanto não pode haver equivalência lógica entre os conceitos (ou enunciados) nas duas linguagens.¹⁷³

Procurando deixar isto claro, Smart emprega mais uma analogia, desta vez com nações. Ele observa que nações não são algo adicional a cidadãos, mas ainda assim os dois têm significados diferentes. Dito de outra forma, enunciados de identidade estrita não significam que os termos têm a mesma lógica. É por isso que podemos utilizar coerentemente o termo “calor” sem saber coisa alguma sobre transferência de energia resultante de diferenças de temperatura (exemplo: “Hoje o calor está de matar”). De forma análoga, esperamos que um ignorante sobre processos neurológicos possa discursar sobre todo tipo de sensação. Sendo assim, a ITM defendida por Smart manifesta a expectativa de uma redução completa dos fenômenos experiências, a ser conduzida por pesquisas empíricas. A teoria resultante revela a “verdadeira natureza”¹⁷⁴ do fenômeno, da mesma maneira que a ciência moderna explica relâmpagos com disgressões sobre descargas elétricas. Evidentemente, o que é aqui chamado de “fenômeno” é o objeto em si, e não a

¹⁷² *Is Consciousness a Brain Process?* p. 56. Minha tradução

¹⁷³ The “Mental” and the “Physical”, p. 69. Minha tradução.

¹⁷⁴ *Idem*, p. 63. Minha tradução.

aparência do mesmo. Analogamente, Smart pretende reduzir a experiência em si, e não o conteúdo fenomenal da mesma. Este é um ponto importante. Smart discute sobretudo experiências visuais em sua análise. A ITM exige a identidade entre a experiência que temos de uma imagem, por exemplo, e certo processo cerebral, e não desta com a imagem. Não faria sentido dizer que um coelho cinza presente em uma alucinação, por exemplo, é um evento cerebral. É por isso que alegar que cirurgiões nunca encontram animais cinzas ou nada parecido em nossos cérebros é uma objeção irrelevante.¹⁷⁵ Temos, então, uma explicação para a lacuna entre as propriedades apresentadas pelos objetos intencionais da consciência e os fatos do cérebro. Experiências são processos perfeitamente reais, mas a experiência acima descrita não é um objeto de cor cinza. Em outras palavras, a realidade das experiências não é questionada (processos cerebrais são, afinal de contas, parte da realidade). O que inexistente é um ente psíquico com as propriedades atribuídas ao seu objeto. Estas últimas são características *do objeto*. “Árvores e papel de parede podem ser verdes, mas não a experiência de ver ou imaginar uma árvore ou papel de parede”, diz Smart.¹⁷⁶ Place chama atenção para o mesmo fato, alertando contra a tentação da “falácia fenomenológica”.¹⁷⁷ A mesma coisa faz David Armstrong em um artigo a ser discutido mais detalhadamente aqui.¹⁷⁸ Estas colocações foram retomadas e articuladas muitos anos depois, por Daniel Dennett e Gilbert Harman,¹⁷⁹ nas controvérsias em torno do conceito de *qualia*. É evidente que Smart pretendia refutar objeções que evocassem qualquer coisa parecida com “propriedades essenciais da experiência”, presentes nos *qualia* contemporâneos e estranhos a eventos cerebrais. Fica clara, portanto, a atualidade do ensaio de Smart, não obstante sua aparente obsolescência pós-funcionalismo. O suposto “verde” da experiência é a “tinta mental” ridicularizada por Dennett. O filósofo do Modelo de Esboços Múltiplos ensina seu leitor a descartar a coisificação dos objetos intencionais. Dennett procura mostrar que um sistema computacional pode armazenar informações sobre objetos visuais sem guardar na memória qualquer imagem colorida. De fato, não há imagem alguma, pois o computador em questão só precisa de um bit-map para representar as cores. Cores não têm que ser

¹⁷⁵ Idem, p.65.

¹⁷⁶ Ibidem. Minha tradução.

¹⁷⁷ *Is Consciousness a Brain Process?* (Place 2002 [1956]), p. 56, 58-60.

¹⁷⁸ *The Causal Theory of Mind*, p. 85.

¹⁷⁹ Harman, G. 1990. *The Intrinsic Quality of Experience. Philosophical Perspectives* 4:31-52.

representadas por cores.¹⁸⁰ De forma similar, Harman procura mostrar que a coisificação dos objetos intencionais é um erro de categoria. Atribui-se à experiência o que está no mundo. Tal é a origem das muitas discussões sobre “o vermelho que Mary viu” etc. Este argumento, claro, é um parente muito próximo do ataque de Edmund Husserl à concepção de representação na filosofia moderna.¹⁸¹ Filósofos como Descartes, Locke e Leibniz acreditavam que a mente tem consciência de *ideias*; estas últimas, portanto, são os objetos dos pensamentos. Estes objetos fazem uma espécie de mediação entre o sujeito da percepção e o mundo à sua volta. Husserl demole esta perspectiva em suas *Investigações Lógicas*, argumentando que temos consciências de objetos externos e não das ideias ou representações. A atividade representacional faz parte do processo de percepção, não sendo ela mesma o objeto representado. Está claro agora que as considerações husserlianas de Smart, quer ele soubesse ou não de sua origem, deram aos filósofos contemporâneos uma grande tela para pintar.

Neste estágio da discussão, torna-se oportuno abordar uma característica da ITM de Smart que justifica seu dar de costas ao behaviorismo e à suposta impossibilidade de relatar estados mentais. A proposta de Smart é analisar relatos de experiências da seguinte maneira: quando relatamos sensações, dizemos que algo acontece similarmente ao que nos ocorre quando estamos aptos a perceber o mundo, temos uma sensação e o objeto da sensação realmente está presente. Quando afirmamos experimentar a sensação de ver uma laranja, por exemplo, agimos como quem diz “algo ocorre similarmente ao que me ocorre quando estou de olhos abertos, existe uma laranja adequadamente iluminada diante de meus olhos, e eu realmente vejo a laranja”.¹⁸² Mencionar dores é fazer algo análogo: se digo que sinto uma dor aguda na mão esquerda, posso dizer que “algo semelhante ao que se passou quando a saúva mordeu minha mão esquerda está ocorrendo agora”. Os relatos entre aspas são descritos por Smart como apresentando palavras “semilógicas” (*quasi-logical*) ou “tópico-neutras” (*topic-neutral*).¹⁸³ É por isso que um completo ignorante sobre processos cerebrais é capaz de fazer relatos corretos sobre suas percepções. As palavras utilizadas no relato têm uma espécie

¹⁸⁰ *Consciousness Explained*, p. 344-350.

¹⁸¹ *Investigações Lógicas* (Husserl, 1970), p. 90-193.

¹⁸² *Idem*, p. 64.

¹⁸³ *Ibidem*. Minha tradução.

de vagueza que as torna compatível com ontologias diversas. Enunciados sobre sensações têm uma neutralidade similar àqueles de enunciados como “alguém entrou no quarto”. O “alguém” daqui é tão vago quanto o “algo” dali, e o enunciado sobre o que ocorre no quarto continua adequado se aquele “alguém” vem a ser um médico, e o agente que o proferiu nunca ouviu falar sobre médicos. Só uma pessoa entrou no quarto, não duas (alguém mais o médico). Smart vê nestas asserções uma possibilidade de explicar a tão explorada inefabilidade da experiência das teorizações antirreducionistas. Os *qualia* (chamados por Smart de *raw feels*) parecem indizíveis porque sua verbalização é vaga como uma verbalização sobre “algo”. Quando falamos de “algo” lhe atribuímos praticamente nenhuma propriedade. Eis a chave para entendermos o caráter “escorregadio” da experiência e a compatibilidade de seu discurso com ontologias díspares. Smart oferece, com base nesta discussão da linguagem da experiência, uma definição de “cor”. Cores são poderes de acionar capacidades sensoriais discriminatórias. Dada a ITM, as sensações ativadas pelas cores vêm a ser idênticas a processos neurológicos.¹⁸⁴

Neste momento, deve estar claro que a ITM de Smart não é um exemplo de *token physicalism*, ao contrário do monismo anômalo de Davidson. *Token physicalism*, vale lembrar, é uma classe de teorias onde propriedades mentais e físicas são instanciadas pelos mesmos entes. Sendo assim, qualquer evento mental apresenta esta ou aquela propriedade física. É por isso que, nesta classe de teorias, todo evento mental é um evento físico. O mental e físico são expressos por propriedades diferentes, mas que co-existem nos mesmos eventos. É importante observar que o *token physicalism* por si só não oferece nenhuma explicação para as correlações entre experiência e cérebro de que falamos no início da seção. Conforme observado na discussão do monismo anômalo, o *token physicalism* é compatível com a inexistência de qualquer relação sistemática entre propriedades físicas e mentais. Conforme discutimos anteriormente, uma redistribuição total de pensamentos pelo mundo não implicaria em qualquer mudança no universo físico. De fato, a asserção relativa à coincidência de ambos os tipos de propriedade no mesmo evento diz muito menos do que aparenta em um primeiro momento. Kim mostra isso com um paralelo envolvendo cores e formas: todo objeto que tem uma

¹⁸⁴ Ibidem.

cor tem também uma forma. Logo, as propriedades de ter cor e ter forma são instanciadas pelos mesmos objetos. No entanto, não há relação sistemática entre tais propriedades; não existe uma forma que exige esta ou aquela cor, e vice-versa. Analogamente, a simples aparição conjunta de propriedades no *token physicalism* é insuficiente que uma relação interessante entre elas seja verificada. Kim observa que por esta razão, o *token physicalism* só é um fisicalismo em um sentido muito pouco rigoroso. Não exigindo leis psicofísicas, a simples adoção da dita posição nada oferece para explicar as correlações que se deseja entender. Nas palavras de Kim, “uma relação propriedade-propriedade sistemática entre mentalidade e nossa natureza corporal é de importância fundamental para uma posição fisicalista robusta”.¹⁸⁵ Sem a adição de leis psicofísicas, o *token physicalism* parece um caminho pouco promissor para a metafísica da mente.

O que a ITM teria a oferecer para preencher a lacuna? Já vimos que Smart defende uma identidade estrita entre tipos mentais e tipos físicos revelados pela neurociência. Isto significa que propriedades mentais são instanciadas pelos mesmos eventos que certas propriedades físicas porque ambas *são as mesmas coisas*. Temos uma explicação, portanto para as correlações observadas entre sensações e processos cerebrais. É necessário observar, porém, que esta é uma explicação peculiar; as correlações não sobrevivem à explicação; elas são eliminadas por esta última (*explained away*). Este é o caso por que não se pode correlacionar uma coisa com ela mesma. Como diz Smart, é possível “correlacionar pegadas com ladrões, mas não Bill Sykes o ladrão com Bill Sykes o ladrão”.¹⁸⁶

Ainda que a ITM de Smart, Feigl e Place parecesse plausível para aqueles com simpatias materialistas, um desastre logo se abateu sobre ela. O clássico artigo de Hilary Putnam, *The Nature of Mental States*¹⁸⁷ convenceu muitos estudiosos da implausibilidade da ITM. O dito artigo também estabeleceu o funcionalismo como uma nova ortodoxia na filosofia da mente (sofrendo depois, é

¹⁸⁵ *Philosophy of Mind*, p. 105. Minha tradução.

¹⁸⁶ *Sensations and Brain Processes*, p. 61. Minha tradução.

¹⁸⁷ Putnam, 2002 [1967], p. 73-79.

evidente, muitos ataques)¹⁸⁸, introduzindo a noção de “realização múltipla”. O funcionalismo passa então a ser proposto como uma filosofia anti-reducionista a justificar a posição de “ciência autônoma” da psicologia. Trata-se de uma posição, afinal, de elementos que não são em si mesmos nem físicos nem espirituais, nem mentais. Kim observa que o fisicalismo reducionista havia sido abandonado pela maior parte dos filósofos em meados da década de 70, ampliando o impacto do monismo anômalo de Davidson.¹⁸⁹ Discutiremos o funcionalismo em detalhes mais adiante, mas antes será oportuno descrever a refutação a que Putnam submete a ITM.¹⁹⁰

Para o bem ou para o mal, a maioria de nós está bem familiarizada com sensações dolorosas. Daniel Dennett observa em seu livro *Tipos de Mentes*¹⁹¹ que o comportamento notavelmente inteligente dos polvos sugere que eles também são capazes de sentir dor. Putnam observa que fatos como estes exercem tensão insuportável sobre a ITM. O leitor deve lembrar-se da identificação de dores e de mais sensações com tipos cerebrais. Para cada tipo mental, temos uma identidade com um tipo neurobiológico. Mas suponhamos que os polvos são, de fato, capazes de sentir dores. Não é extremamente repugnante à intuição afirmar que polvos e homens, sempre que sentem dores, compartilham eventos cerebrais idênticos? A repugnância da ideia vem da constatação de que homens e polvos são organismos muito diferentes, com cérebros idem. De fato, Dennett observa que polvos são mais parecidos fisiologicamente com ostras do que com mamíferos.¹⁹² Poder-se-ia retrucar negando realidade à supostas dores dos espertos moluscos, mas tampouco é razoável supor que as sensações dolorosas dos macacos ou cães são, em última análise, eventos cerebrais iguais àqueles do *Homo Sapiens*. O problema fica ainda mais claro quando pensamos em uma possível espécie extraterrestre dotada de consciência análoga à nossa. Tendo passado por processos evolutivos muito diversos (se é que a Teoria da Evolução se aplicaria aqui), não seria de se esperar

¹⁸⁸ Uma crítica abrangente e interessante por não evocar o ubíquo conceito de *qualia* está em Paul Churchland, *Eliminative Materialism and the Propositional Attitudes* (Churchland, 2002 [1996], p. 574-576.

¹⁸⁹ *Philosophy of Mind*, p. 115, 149n.

¹⁹⁰ *The Nature of Mental States*, p. 76-77.

¹⁹¹ Dennett, 1997.

¹⁹² *Tipos de Mentes*, p. 94-95.

que os ETs em questão possuíssem os mesmíssimos processos cerebrais que temos. Kim procura deixar isto claro quando afirma:

Durante o processo de lidar com os perigos presentes em seu hábitat, espécies animais têm que ter desenvolvido mecanismos de dor, que podemos chamar “detectores de danos em tecidos”. É plausível que espécies diferentes, interagindo com condições ambientais diferentes e evoluindo de forma independente, tenham desenvolvido mecanismos dissimilares com o mesmo propósito.¹⁹³

É possível alcançar um fim psicológico através de uma ampla variedade de meios neurológicos. Se este é o caso, a tentativa de igualar estruturas neurológicas a funções psicológicas está condenada de antemão (*for doomed*).¹⁹⁴

Antecipando uma possível resposta dos teóricos da identidade, Putnam afirma ainda que a utilização de tipos disjuntivos (dor = evento cerebral C1 (relativo à espécie E1) v C2 (relativo à espécie E2) etc.) seria uma sobrevida da ITM inaceitavelmente *ad hoc*.¹⁹⁵ Eis a nêmesis sessentista do *type physicalism*.

A noção fundamental da proposta funcionalista é enxergar fenômenos mentais como nodos em uma estrutura causal abstrata, apta a ser implementada em variados substratos físicos. Se os argumentos expostos logo acima estão corretos, estados de consciência podem ser reproduzidos em organismos díspares; basta que os mesmos reproduzam à sua própria maneira um determinado conjunto de relações causais. O funcionalismo está ligado, então, a uma teoria da realização múltipla. Kim observa, porém, que existe um compromisso implícito da filosofia funcionalista com o fisicalismo: as entidades a reproduzir as referidas relações causais são entidades físicas – ninguém neste debate leva a sério a possibilidade de anjos ou outros entes inteiramente não físicos, como as coisas pensantes cartesianas. Nas palavras de Kim, “qualquer coisa a exibir mentalidade deve ser um sistema físico – um organismo biológico, por exemplo”.¹⁹⁶

A discussão das sensações de dor em humanos e polvos ilustra o ponto fundamental visado por Putnam: devemos entender as propriedades mentais como propriedades abstratas a gerar uma descrição funcional. Kim explica esta concepção evocando o conceito de *artefato*. Qualquer ente capaz de realizar uma

¹⁹³ *Philosophy of Mind*, p. 117. Minha tradução.

¹⁹⁴ *Special sciences (or: The Disunity of Science as a Working Hypothesis)*, p. 130. Minha tradução.

¹⁹⁵ *The nature of mental states*, p.77.

¹⁹⁶ *Philosophy of Mind*, p. 116. Minha tradução.

tarefa específica satisfaz um dado conceito de artefato. Um motor, por exemplo, é um engenho utilizado para converter energia em movimento. A tarefa pode ser realizada por vários equipamentos diferentes, movidos a gasolina, a diesel ou a querosene; projetados como turbinas, motores a pistão, reatores nucleares, etc. Enquanto a tarefa puder ser realizada, teremos um motor, pouco importando a sua composição física. O mesmo vale para conceitos biológicos; recordemos as variadas soluções que a natureza encontra para o mesmo problema: seres humanos e insetos dispõem de equipamentos ópticos díspares para captar informações proporcionadas pela abundância de luz em seus ambientes. Ambas as categorias, porém, satisfazem o conceito de *olho*, pois *funcionam* de forma análoga, explorando a luz de forma a elaborar representações dos objetos à sua volta. Kim dá prosseguimento a sua exposição mostrando como sensações podem ser entendidas da mesma maneira. É possível elaborar uma descrição da tarefa à qual a natureza incumbiu as sensações de dor (sua classe de exemplos favorita). Conforme exposto acima, animais os mais diversos necessitam de detectores de danos a seus tecidos para lidar com as ameaças que abundam em seu hábitat. Espécies diferentes teriam “encontrado” soluções diferentes para o problema – diferentes “designs” de detectores de danos sensíveis a agressões diversas que, uma vez acionados, geram as reações adequadas – fuga, revide, etc. Como argumentou Putnam, estes detectores podem ser mecanismos com propriedades físicas bastante diferentes. Conclui-se que propriedades mentais são abstratas ou formais,¹⁹⁷ aparecendo no mundo por meio de substratos diversos, mas reproduzindo o mesmo padrão causal. A categoria de sensações em discussão – dor – ilustra bem a ideia por ser facilmente compreendida como algo gerado por certo tipo de *input*, gerando por sua vez determinados tipos de *output* (uma série de comportamentos observáveis e também outros estados mentais, tais como apreensão, irritação, medo, etc). Um exemplo curioso: argumenta-se que é permissível a vegetarianos comer ostras, pois tudo indica que tais moluscos bivalves não são capazes de sentir dor. Chega-se a esta conclusão a partir da constatação de que ostras não podem se esquivar de danos, porque não podem se movimentar à vontade. Traduzindo em linguagem funcionalista: ostras não apresentam um tipo de *output* presente na definição de “dor” e, portanto, não

¹⁹⁷ Idem, p. 118.

sentem dor.¹⁹⁸ Pela mesma lógica, lagostas devem sentir dor, já que exibem o output esperado diante de danos em seus tecidos: quando cozidas vivas, elas se mexem freneticamente e tentam escalar a panela para escapar. Chega-se a esta conclusão independentemente do fato de que lagostas não têm um sistema nervoso central. Outra evidência anedótica da zoologia em favor do funcionalismo é a constatação óbvia de que lagostas têm percepções visuais mesmo sem possuir um córtex visual.¹⁹⁹

Nesta discussão adotamos o exemplo usual da dor como um estado funcional. Uma teoria promissora da mente, no entanto, tem que ser capaz de gerar análises de muitos outros fenômenos. Felizmente para os funcionalistas, este parece ser o caso. Mesmo sentimentos e o que o filósofo Christopher Hill chama de “a dimensão hedônica da emoção”²⁰⁰ podem ser analisados funcionalmente. Hill escreve que não existe um *componente qualitativo* comum a todas as experiências que consideramos agradáveis, como contemplar um Vermeer ou beber bom vinho. O que estas experiências têm em comum, diz Hill, é uma propriedade funcional aparentada à categoria dos desejos.²⁰¹ Apreciar uma experiência consiste, entre outras coisas, em

1. Dirigir a atenção para o objeto da experiência;
2. Exibir uma tendência a lembrar do objeto;
3. Tender a se lembrar das circunstâncias que puseram o sujeito em contato com o dito objeto;
4. Tender a exibir comportamentos que intensificam e prolongam a experiência;
5. Tender a exibir comportamentos que prometem trazer experiências futuras similares;
6. Tender a uma redução de severidade da expressão facial;
7. Tender a uma redução de tensão.²⁰²

¹⁹⁸ Ver o artigo de Christopher Cox, *Consider the Oyster*, disponível em <http://www.slate.com/id/2248998/>.

¹⁹⁹ Ver o artigo de Megan Jungwi, *Do Lobsters feel Pain?* Disponível em <http://marine-life.suite101.com/article.cfm/do-lobsters-feel-pain>

²⁰⁰ Hill, C. *Consciousness*, p.209.

²⁰¹ Idem, p. 210.

²⁰² Ibidem.

Kim procura tornar clara a posição funcionalista contrastando-a com o behaviorismo.²⁰³ Nesta última doutrina, tipos mentais são identificados com tipos comportamentais, o que significa que diferentes organismos apresentam dores se e somente se exibem certo padrão de comportamento ou disposição ao mesmo, a saber, gemidos, lamentos, contorções faciais etc. Questões sobre a realidade dos processos a intermediar entre as agressões sofridas e as respostas comportamentais são descartadas para todos os propósitos práticos. Neste sentido, o behaviorismo é uma abordagem instrumentalista dos fenômenos psicológicos. O funcionalismo, ao contrário, exige uma postura realista no que diz respeito aos estados internos do organismo que causam o comportamento. Os fenômenos mentais de fato “estão lá”, influenciando o curso dos acontecimentos. Tal coisa não faria sentido para o behaviorista, cujas referências a estados mentais apontam somente para comportamento possível ou factual.²⁰⁴ Sendo assim, uma diferença fundamental entre as duas posições é o quão sério elas levam a noção de *estado interno* de um sistema.²⁰⁵ É verdade que se ambas as posições usam os conceitos de *input* e *output* como peças-chave, sendo, portanto, aparentadas. Tal relação de parentesco é esclarecida se considerarmos o funcionalismo como uma variante mais sofisticada de behaviorismo.²⁰⁶ Kim prossegue em sua exposição com uma menção de outra diferença importante entre os dois “parentes”: no funcionalismo é permitido encarar estados mentais como *inputs* e/ou *outputs* de um outro estado mental. O behaviorista, ao contrário, tenta eliminar por explicação qualquer referência a tais estados.

Em resumo, a filosofia introduzida por Putnam é uma perspectiva que identifica eventos mentais com posições em uma rede causal. Diz-se que determinada sensação é uma dor, deleite ou percepção com base na sua posição em um conjunto de fenômenos também mentais e “condições *input-output*”.²⁰⁷ Conclui-se que a identidade dos componentes da consciência é determinada por sua dependência mútua. Se insistirmos na dor como exemplo, um evento físico

²⁰³ Idem, p. 119.

²⁰⁴ Idem, p. 122.

²⁰⁵ Ibidem.

²⁰⁶ Ibidem. De fato, Daniel Dennett, que apresenta uma teoria funcionalista da mente em *Consciousness Explained*, é acusado pelo cientista de Inteligência Artificial Yorick Wilks de “behaviorista radical”, e Bo Dahlbom nos diz que os trabalhos de Dennett estão repletos de “ideologia behaviorista”, o que Dennett nega.

²⁰⁷ Idem, p.124. Minha tradução.

qualquer só é uma dor por exibir certa relação com ansiedade, irritação, e medo, entre outras reações, e essas, por sua vez, também devem sua identidade a todas as demais. Lembrando que se tratam sem exceção de fenômenos físicos, temos uma concepção holística do mental,²⁰⁸ onde é impossível definir um evento independentemente de todo o resto. Fica claro então o quanto o funcionalismo difere da ITM; nesta última as identificações teóricas se restringem aos pares de supostos correlatos. Infelizmente, o caráter holístico do funcionalismo lhe dá uma aparência de circularidade. Se a definição de “fúria” depende de definição de, digamos, “sensação de ultraje”, e vice-versa (supõe-se que a segunda causa a primeira) parece não ser possível utilizar o funcionalismo para avançar rumo a uma teoria das duas. Logicamente, isto parece valer para todo o resto. Putnam não aborda o problema em seu artigo; felizmente, o filósofo australiano David Malet Armstrong vem a seu resgate. Armstrong rechaça o incômodo evocando conceitos “correlativos” (*correlative concepts*) que poucos considerariam problemáticos: “...os conceitos de marido e esposa ou os conceitos de soldado e exército. Não existem maridos sem esposas ou esposas sem maridos. Nem soldados sem um exército, ou um exército sem soldados”.²⁰⁹ O que ocorre nestes exemplos é que “os conceitos correspondentes têm que ser introduzidos juntos ou estar de todo ausentes”.²¹⁰

O esquema funcionalista de realização múltipla é um cenário convidativo para analogias com a ciência da computação. Uma mesma máquina virtual, ou máquina de Turing, afinal, pode ser instalada em uma variedade de substratos físicos. Não somente relações input-output podem ser reproduzidas por mecanismos computacionais que chegam ao mesmo resultado por meios diferentes; o mesmo processo computacional pode ser implementado por máquinas de composição totalmente diversa. Kim pede aos seus leitores que imaginem um projeto de aula onde os alunos devem construir uma máquina capaz de realizar as computações de uma certa Tabela de Estados. Kim prossegue supondo que cada grupo de alunos apresentaria uma máquina física completamente diferente: umas poderiam ter fitas de papel onde outras utilizariam sinais magnéticos; esta máquina implementaria a máquina de Turing com discos

²⁰⁸ Ibidem.

²⁰⁹ *The Causal Theory of Mind*, p. 83. Minha tradução.

²¹⁰ Ibidem. Minha tradução.

de computador, aquela o faria com mecanismos hidráulicos. O número de possibilidades, enfim, é vastíssimo.²¹¹

Com o funcionalismo, a relação entre mentes e computadores na história das ideias terminou de percorrer um círculo completo. Filosofias contemporâneas propõem a utilização de computadores como modelos para mentes, mas os primeiros engenhos computacionais foram criados para simular processos de raciocínio humanos. Em 1790, Napoleão Bonaparte decidiu criar um novo sistema de taxaço para seu país, e o escritório encarregado assinalou algumas dezenas de funcionários para preencher as tabelas necessárias. Esta tarefa envolvia a elaboraço de novos mapas, que introduziriam o sistema métrico de mediço no lugar do antigo sistema imperial. As operaço de adiço e subtraço realizadas pelos ditos funcionários eram em si mesmas triviais, mas a sua enorme quantidade justificava o número de “computadores humanos”. Foi este contexto que inspirou o matemático inglês Charles Babbage a tentar reproduzir o trabalho dos funcionários franceses com uma nova máquina. As elucbraço de Babbage o levaram até a ideia de um “Engenho Analítico” (*Analytical Engine*) apto a realizar todo tipo de cálculo matemático. Infelizmente, intrigas entre Babbage e o governo britânico impediram a construçõ desta máquina, mas um *Difference Engine* de capacidades mais modestas (a elaboraço de tabelas como as de Napoleão) foi de fato construído, catapultando Babbage para a notoriedade.²¹²

Continuando a trajetória do círculo: a afinidade entre a realizaço múltipla da teoria da computaçõ e aquela que Putnam supôs existir na psicologia o levou a uma variante do funcionalismo onde a mente humana é a “encarnaço” de uma máquina de Turing.²¹³ É possível ter esta mesma máquina “rodando” em corpos marcianos e em todo tipo de besta, o que levaria a uma mentalidade idêntica. A reproduço de estados funcionais em organismos díspares reproduziria as mesmas sensações. A máquina de Turing em questõ, no entanto, é o que se chama de Autômato Probabilístico. A diferença desta para a máquina de Turing original é que nesta as transiço entre estados do sistema são determinísticas, ao passo que um Autômato Probabilístico passa de um estado para outro de forma

²¹¹ *Philosophy of Mind*, p.130.

²¹² Campbell - Kelly 2009.

²¹³ *The Nature of Mental States*, p. 75-76.

probabilística. Putnam presumivelmente fez uso do conceito para acomodar a intuição da mente como um sistema não-determinístico. A Tabela de Estados do sistema incorpora as noções de *input* sensorial e *output* motor; desta forma, a Tabela de Estados especifica uma instrução para cada combinação de inputs e estado do sistema. As instruções em questão determinam as probabilidades de novos estados e de respostas corporais adequadas. Os estados interconectados probabilisticamente são os estados mentais do organismo. A organização funcional do organismo (ou mesmo de um sistema consciente, mas inorgânico, já que nada impede que robôs realizem a mesma máquina virtual de seres humanos) exige configurações capazes de certos tipos de comportamento, ou seja, de ter certos efeitos no mundo físico; sendo assim, é óbvio que não é possível ter um ser consciente feito de manteiga ou gelo seco. Ainda assim, as configurações capazes de dar vida à máquina probabilística são legião. Para que um sistema possua uma determinada sensação, basta que ele: 1) Seja um Autômato Probabilístico; 2) Tenha certa Organização Funcional relativa a uma dada Descrição (isto é, existe ao menos um enunciado verdadeiro afirmando a existência, no sistema, de estados S1, S2, S3... , SN interconectados por uma Tabela de Estados a especificar inputs sensoriais e respostas motoras); 3) Seja incapaz de decomposição em partes possuidoras de Descrições como aquela mencionada em (2), o que evita o absurdo de “exames de abelhas como sujeitos de dores”²¹⁴ e 4) Tenha, para cada descrição mencionada em (2), um subconjunto de inputs sensoriais cuja presença é necessária e suficiente para provocar a sensação em questão. Putnam observa que, da mesma maneira que a ITM, sua teoria possui a vantagem de ter conteúdo empírico. A favor de seu funcionalismo estão os fatos de uma menor vagueza e, conseqüentemente, condições de verificação mais claras.²¹⁵

Como seria de se esperar, a mensagem de Putnam ecoou rapidamente pelos departamentos de filosofia americanos, e Jerry Fodor logo juntou-se a seu ataque funcionalista à ITM. Fodor, também defensor notório da teoria computacional e autor de uma versão bem mais detalhada, explora o *insight* da realização múltipla em termos nomológicos. Jerry Fodor procura expor o *type physicalism* como desnecessário para uma metafísica e uma epistemologia amigáveis ao progresso

²¹⁴ Idem, p.76. Minha tradução

²¹⁵ Ibidem.

científico e, portanto, bem-comportadas. O leitor certamente se lembra do ideal, defendido por Smart e Feigl, de uma metafísica elegante, a acomodar confortavelmente os progressos da ciência e livre de penduricalhos nomológicos. Fodor argumenta que a unidade da ciência não requer a descoberta de tipos naturais físicos cuja extensão coincide com cada tipo natural justificadamente presente nas ciências “menos básicas”. É enganoso propor este tipo de redução mesmo se a ideia de realização múltipla for incorporada à ITM (o que Putnam, recordemos, acha inaceitável). O resultado seria identificar a propriedade a ser reduzida com um conceito disjuntivo. Fodor alega que este objetivo é quimérico. O que é razoável esperar é uma teoria sólida de como os eventos físicos podem estar organizados de forma a obedecer leis das chamadas “ciências especiais”: as disciplinas “especializadas”, tais como a economia, e é claro, a psicologia. Fodor espera demonstrar que o êxito da segunda tarefa, uma empreitada auspiciosa, felizmente não requer o êxito da primeira, um projeto equivocado. O *token physicalism* é tudo o que uma metafísica consistente requer.

Fodor observa que o *insight* da realização múltipla não vale somente para a psicologia; o escopo da ideia engloba as demais ciências especiais. Já vimos que a economia é uma delas. Fica claro que aplicar a ideia de realização múltipla à economia é legítimo quando consideramos, por exemplo, conceitos como o de “troca monetária”. Toda troca monetária deve envolver algum tipo de evento físico, mas que tipo? Evidentemente, os candidatos existem em enorme número, não só de objetos individuais, mas de tipos físicos. O valor dos objetos, afinal, depende mais de propriedades que lhes são conferidas por agentes do que de suas características intrínsecas. Seja como for, espera-se dos teóricos da economia a elaboração de generalizações científicas sobre trocas monetárias. O que os interessados nestas generalizações procuram não são coincidências com tipos físicos. A forma física dos eventos em questão, em outras palavras, não os interessa. Os objetos que realizam trocas monetárias seguem padrões interessantes, mas estes padrões são salientes quando consideramos os objetos *enquanto entes econômicos*, e não *enquanto entes físicos*. A disjunção que abarca todos os objetos envolvidos em trocas econômicas pouco ou nada apresenta de sistemático quando consideramos as propriedades físicas dos mesmos. Dito de outra maneira, se o tipo econômico “troca monetária” compartilha sua extensão

com algum tipo físico, isto é *meramente accidental*. Fodor observa ainda que o mesmo vale para as demais ciências especiais, o que significa que não devemos esperar uma redução de tipos psicológicos a tipos físicos como predicados neurológicos.²¹⁶ Um cenário evocado por M.J Cain pode ajudar a tornar mais plausíveis as observações de Fodor. A dinâmica populacional de sistemas ecológicos obedece a uma lei que nos capacita a prever flutuações populacionais dentro destes sistemas. Se a população de leões na savana decresce, a população dos animais dos quais os leões se alimentam cresce (caso todo o resto se mantenha constante). A mesma lógica vale para todos os elementos do ecossistema envolvidos em relações predatórias. Um aumento no número de predadores reduz a população de presas e uma redução da primeira implica o crescimento da segunda. Evidentemente, o que temos aqui é uma lei que não se reduz a nenhuma lei física, já que as características físicas de diferentes populações e ecossistemas são muito dissimilares.²¹⁷ Segundo Cain,

Conclui-se que existe uma lei de poder preditivo e explanatório em relação a qual a física é totalmente cega. Consequentemente, a física, ainda que capaz de explicar e prever uma explosão da população de trutas em um lago enquanto evento físico, não é capaz de prever e explicar tal evento enquanto explosão populacional de uma colônia de animais ocupando determinado papel em uma cadeia alimentar.²¹⁸

Fodor vai mais longe que Putnam em sua condenação da redução de tipos psicológicos a tipos físicos. Já vimos que Putnam descarta a possibilidade de uma redução de conceitos mentais a conceitos físicos disjuntivos como inaceitavelmente *ad hoc*. Pode-se duvidar, porém de um argumento tão pouco elaborado. Percebendo esta lacuna, Fodor raciocina de forma a demonstrar que o bisturi do *type physicalism* não é afiado o bastante para cortar todas as verrugas nomológicas. O resultado da operação, ao contrário, cria um novo monstro de Frankenstein. O argumento baseia-se na ideia de redução via “leis de ligação” (*bridge laws*) que funcionam da seguinte maneira: suponhamos a existência de uma lei, psicológica ou não, presente em uma ciência especial qualquer, que diz que toda situação S_1 leva a uma situação S_2 :

²¹⁶ *Special Sciences*, p.128-130.

²¹⁷ Cain, 2002, p. 48.

²¹⁸ *Ibidem*. Minha tradução.

$$(1) S_1x \rightarrow S_2x$$

Se S_1 e S_2 são redutíveis a tipos físicos, temos as seguintes de ligação:

$$(2) S_1x \text{ se e somente se } P_1x$$

$$(3) S_2x \text{ se e somente se } P_2x$$

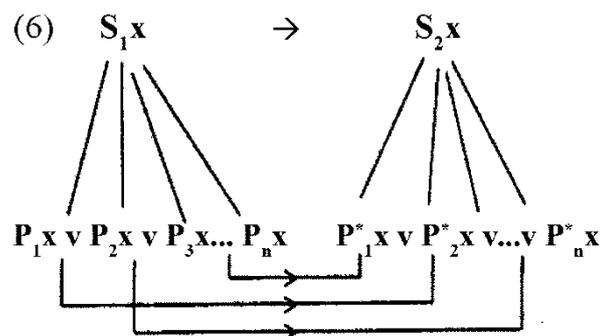
E portanto:

$$(4) P_1x \rightarrow P_2x$$

Se a ciência especial reduzida tiver suas propriedades identificadas com um único tipo físico, tudo estará bem. Obter-se-á enunciados aptos a ilustrar leis científicas. Se é o caso que as propriedades S da ciência especial são realizadas de múltiplas maneiras, porém, surge um problema. Dada uma disjunção de tipos físicos da ciência “reduzora”, examinemos uma lei de ligação expressa por

$$(5) S_x \text{ se e somente se } P_1x \vee P_2x \vee \dots \vee P_nx$$

Considerando $P_1 \vee P_2 \vee \dots \vee P_n$ como ausentes na lista de tipos naturais da ciência a ser reduzida, o quadro nomológico resultante é aberrante. Este é o caso porque, supondo-se que as leis da física são estritas, haverá leis físicas ligando a satisfação de cada membro da disjunção à qual se reduziu o antecedente com a satisfação de um membro da disjunção do consequente. Sendo assim, dada a lei (1), é verdadeiro que



Fodor nota então que os disjuntos estão eles mesmos ligados nomologicamente; segue-se que eles mesmos são tipos naturais. Infelizmente para o reducionista radical, isto parece absurdo. O que torna o cenário como um todo inaceitável é que temos, como o leitor irá se recordar, disjuntos absolutamente diversos. Os objetos aptos a aparecer em trocas econômicas variam enormemente, como variam aqueles aptos a sustentar propriedades mentais. Conclui-se que teríamos uma lei relacionando tipos naturais de exotismo ímpar. Fodor pede ao leitor que assuma como leis:

- a) A luz do sol causa síntese de nutrientes em plantas verdes
- b) Fricção gera calor

Se o cenário nomológico proposto pelos reducionistas estiver correto, os mesmos terão que admitir que é uma lei estrita que (a luz do sol incidindo sobre plantas verdes v fricção) causa (síntese de nutrientes v calor). Isto é absurdo, como é absurdo supor que (síntese de nutrientes v calor) e (a luz do sol incidindo sobre plantas verdes v fricção) são tipos naturais. Se propriedades mentais são passíveis de realização múltipla e a redução das ciências especiais se dá via leis de ligação, a ITM é inviável. Fodor reconhece que o conceito de lei científica não é claro a ponto de tornar o argumento indubitável, mas crê que as considerações acima são suficientes para solapar a credibilidade da ITM e fortalecer a alternativa funcionalista.

Segundo o ponto de vista de Fodor, estas dificuldades com o materialismo reducionista reforçam a plausibilidade de sua própria teoria, conhecida como *Computational Theory of Mind* (doravante CTM). Com esta formulação, Fodor espera fazer justiça aos fatos da múltipla realização e integrar a filosofia da mente às ciências cognitivas. De fato, Fodor foi um divulgador da ciência cognitiva entre seus colegas filósofos nos anos 60, e esteve envolvido em experimentação psicológica no início de sua carreira. A ascensão do cognitivismo nesta época foi o fio condutor de suas investigações filosóficas. Como seria de se esperar, tal programa de pesquisa consiste em esclarecer a natureza da cognição humana. As ideias desenvolvidas abarcam fenômenos como raciocínio lógico, competência

linguística e percepção.²¹⁹ Previsivelmente, uma dos fatores a inspirar este movimento foi o desenvolvimento de computadores digitais.²²⁰ O caráter mecânico dos modelos resultantes é exatamente o recurso teórico necessário para refutar a ideia behaviorista de que qualquer forma de realismo em relação envolve postular homúnculos dentro do cérebro.²²¹ A ideia básica é enxergar as capacidades cognitivas humanas como análogas às aptidões de um computador para resolver problemas de processamento de informações. Partindo de dados sobre os poderes da mente humana, pode-se construir uma teoria destes poderes por meio de reflexões sobre os problemas computacionais envolvidos. As hipóteses resultantes podem ser testadas por meio de comparações entre o desempenho esperado por um sistema operando segundo determinados modelos e o desempenho de sujeitos humanos. Se este é o caso, objeções behavioristas que apontam para o suposto caráter não científico de fenômenos mentais perdem toda sua força.²²²

A CTM não é exatamente uma teoria funcionalista, mas discuti-la nesta seção é justificado, por ter a mesma forte afinidade com a dita classe de teorias, a saber, a exploração do território aberto pela múltipla realização e a analogia com engenhos computacionais. A razão pela qual se classifica a CTM como não funcionalista é a preferência de Fodor por teorias não funcionalistas do conteúdo mental.²²³ Não sendo necessária uma discussão detalhada de tais ideias sobre conteúdo mental, podemos examinar a LOT no que se segue. Isto tornará claras as implicações metafísicas por ela compartilhadas com o funcionalismo de Putnam e as ideias de Armstrong, e de fato com todas as teorias classificáveis como *token physicalism*.

A ideia de representação mental pode ser usada como chave para o que intencionamos expor. Estados intencionais são para Fodor relações envolvendo representações mentais, que, por sua vez, podem ser definidas como símbolos inseridos na mente.²²⁴ É intuitivo imaginar representações mentais como entes

²¹⁹ Cain, p. 28.

²²⁰ Idem, p.30

²²¹ Ibidem.

²²² Ibidem.

²²³ Idem, p. 63

²²⁴ Idem, p. 52.

imagísticos; como vimos, esta intuição converte-se facilmente na falácia fenomenológica mencionada por Smart, Harman, Dennett e vários outros. Na CTM, porém, representações mentais têm muito mais em comum com símbolos linguísticos do que com imagens. De fato, as representações que usamos em nossos pensamentos pertencem a uma linguagem não natural (mais detalhes a seguir) chamada “Linguagem do pensamento” (*Language of Thought*, doravante LOT). O construto teórico resultante explica atitudes proposicionais em geral da seguinte maneira: desejar comer *steak tartar* para o jantar envolve nada mais do que possuir a relação “desejar” relativa à representação mental *steak tartar*. A mente rastreia eventos no mundo por meio de sentenças da LOT. Representações consistem na posse de uma sentença da LOT “sobre” o objeto rastreado. Uma vez que toda representação mental é a ativação na mente de uma sentença da LOT, o caso em exame envolve a ativação de uma sentença da LOT equivalente a “steak tartar”. Esta ativação é classificada pelo sistema como pertencendo à categoria de desejos. Metaforicamente falando, existem “caixas” de desejos, “caixas” de crenças e compartimentos semelhantes para todo tipo de atitude proposicional. Logo, o desejo por *steak tartar* é a ativação de uma sentença na LOT classificada como pertencente a uma determinada “caixa”, e recebendo o processamento característico desta última.²²⁵

Evidentemente, a noção de “sentença” na LOT é inútil se não soubermos o que exatamente isto envolve. Para que possamos enxergar representações mentais como elementos de um tipo de linguagem, é preciso considerar certas características de línguas naturais como a nossa. A língua portuguesa, como tantas outras línguas naturais, abarca um número infinito de sentenças, que por sua vez são estruturas simbólicas compostas por símbolos simples (palavras em português) combinados. A língua portuguesa tem um número finito de palavras e regras sintáticas ou gramaticais. Estas regras são empregadas nas combinações de palavras utilizadas para formar frases e sentenças. De posse das regras sintáticas, é possível classificar combinações de palavras como sentenças bem-formadas ou não. As regras da língua dizem que palavras pertencentes a tais e tais classes gramaticais têm este ou aquele papel na formação de estruturas complexas.

²²⁵ Idem, p. 60.

Ao entendermos o caráter recursivo das regras sintáticas da língua portuguesa e demais línguas naturais, fica claro que se pode obter infinitas sentenças a partir de conjuntos finitos de vocábulos e regras gramaticais. Por recursão é possível aumentar indefinidamente a complexidade de sentenças bem-formadas. Para tanto, basta embutir uma sentença em outra. Assim, falantes do português podem rotineiramente produzir e compreender sentenças que até então lhes eram desconhecidas. O significado das sentenças, por sua vez, é determinado unicamente pelo significado das palavras que a compõem e pela estrutura sintática. Evidentemente, estas estruturas e seus componentes têm incontáveis realizações físicas possíveis. Sentenças podem, afinal, ser comunicadas em um sem-número de maneiras.²²⁶

Da mesma forma que as línguas naturais, a LOT emprega um número finito de símbolos e regras sintáticas de forma a gerar um número infinito de sentenças. O significado de tais estruturas é igualmente determinado pelo significado dos componentes e estrutura sintática. O resultado é uma linguagem capaz de expressar qualquer significado em qualquer língua natural. Compreender uma sentença em português, catalão ou romeno envolve a ativação de uma crença sobre o significado da dita sentença. Já a ativação da crença em si requer a ativação de um símbolo da LOT com o significado apropriado, uma sentença nesta linguagem. Se este é o caso, então toda sentença passível de compreensão requer uma sentença na LOT com o mesmo significado.²²⁷

Não obstante as características compartilhadas com as línguas naturais, a LOT não é uma delas. A posse de uma LOT é, de fato, uma pré-condição para o aprendizado das línguas naturais. Este é o caso por que tal aprendizado consiste na elaboração e confirmação de hipóteses sobre significados. Aprender o vocábulo da língua portuguesa “computador” requer a construção e confirmação da seguinte hipótese: “x é um ‘computador’ se e somente se x é um G”. Não é possível formar tais suposições sem o uso de um sistema representacional. Operações mentais deste tipo exigem a estruturação das possibilidades semânticas em exame. Se nenhuma língua natural é inata, deve haver uma língua de caráter diverso a

²²⁶ Idem, p. 52-54.

²²⁷ Idem, p. 54.

alavancar o aprendizado das demais. Dito de outra forma, não é possível aprender línguas naturais sem uma LOT.²²⁸

Uma característica adicional da LOT a distingui-la das linguagens naturais é a sua relação peculiar com estados intencionais. Como vimos, a posse de um determinado estado intencional é a instanciação de uma sentença da LOT, que é processada de determinada forma pelo sistema. Todo estado intencional tem conteúdo preciso, o que significa que as sentenças da LOT também o têm. Este fato as diferencia das sentenças de línguas naturais e sua ambiguidade.²²⁹ Adaptando um exemplo de M.J. Cain,²³⁰ consideremos alguém que afirma “Fang perseguiu o corredor com um pedaço de pau”. Evidentemente, a asserção é ambígua: ela pode significar que Fang usou um pedaço de pau na perseguição ou que o perseguido carregava consigo um pedaço de pau. Se a língua portuguesa fosse a portadora do significado na mente da pessoa em questão, não haveria fatos determinantes da sua crença; a crença ela mesmo seria ambígua. Sabe-se, porém, que crenças não o são; logo, a estrutura fundamental do pensamento e seus significados tem que ser diferente. Considerações sobre ambiguidade também mostram porque representações mentais não são imagísticas; imagens, como crenças, também dão margens a indeterminações interpretativas.²³¹ Seja como for, é interessante notar que a presença desta estrutura rígida já é suficiente para capacitar um organismo a pensar. Os atos do pensamento são anteriores ao aprendizado de línguas naturais; de fato, aprendê-las pressupõe a capacidade de pensar, como mostra a nossa breve discussão sobre a formação de hipóteses sobre significados de uma língua natural.

Neste momento, seria natural para o leitor se indagar sobre o que significa “ter sentenças da LOT ativas na mente”. Recordando o fato da teoria de Fodor ser uma forma de *token physicalism*, será fácil perceber que as estruturas da LOT têm que ser fisicamente realizadas. No caso da mente humana, as sentenças da LOT são realizadas por estados cerebrais. Isto significa que o já mencionado desejo por *steak tartar* implica a existência de uma sentença da LOT sobre o dito repasto que

²²⁸ Idem, p. 56.

²²⁹ Idem, P. 56.

²³⁰ Ibidem.

²³¹ Idem, P.57-58.

é idêntica a um de meus estados cerebrais. Cada pedaço da sentença é uma parte distinta deste estado cerebral; a estrutura deste último espelha a estrutura da sentença realizada. Sendo assim, pode-se supor que sentenças semanticamente assemelhadas em um mesmo cérebro terão estruturas físicas assemelhadas. Assumir a realização múltipla, porém, significa reconhecer que tais sentenças podem ser reproduzidas em sistemas físicos bastante diferentes, parecidos apenas no que se refere a uma estrutura causal/computacional abstrata.²³²

Por que Fodor chama sua teoria de CTM? A teoria em exame é computacional porque aborda processos intencionais como um mecanismo computacional fisicamente realizado processando sentenças-input da LOT, e gerando uma ou mais sentenças na mesma linguagem como output. O processamento classifica as representações resultantes como pertencendo a esta ou aquela “caixa”. Uma vez que computadores são máquinas que manipulam símbolos sintaticamente estruturados por meio de regras, e sentenças da LOT são símbolos deste tipo, a mente é um tipo de computador. Nas palavras de M. J. Cain, temos a mente como “um computador sediado no cérebro”.²³³

E por que Fodor vê a abordagem computacional como um caminho auspicioso para a compreensão da mente? Como se vê na discussão acima sobre aprendizado de línguas naturais, a CTM torna inteligível a aquisição de tais competências, já que elas requerem a presença de um sistema representacional a alavancá-las.

A possibilidade de explicar a relativa consistência lógica do pensamento é outro benefício que a CTM e sua LOT proporcionam. A coerência dos processos cognitivos está para o engenho computacional que é o cérebro como a consistência dos dados processados por um *software* está para um computador “artificial”. Se o cérebro é um computador realizado fisicamente, não surpreende que haja coerência nos processos mentais por ele realizados. A cognição humana pode ser enxergada como um processo mecânico que respeita regras racionais. Pode-se esperar, então, que a CTM esclareça certos dados sobre a maneira como raciocinamos. Por que, por exemplo, podemos perceber que uma relação R entre

²³² Idem, p.55

²³³ Idem, p. 61. Minha tradução.

dois objetos é simétrica? Isto acontece porque crenças sobre esta relação exigem a ativação de símbolos que representam os três componentes da relação aRb. Estes símbolos são idênticos àqueles utilizados para representar bRa; sendo assim, ao representar a primeira relação o sujeito já possui o que precisa para chegar à segunda. Tal é potencial de uma estrutura sintática. Esta mesma estrutura também nos torna aptos e compreender pensamentos onde um a satisfaz determinado predicado F, ao contrário de b, embora a e b tenham a mesma extensão.²³⁴ Um estudante de História particularmente distraído pode pensar que o Primeiro Ministro britânico durante a Segunda Guerra nasceu nos EUA sem pensar que Winston Churchill nasceu nos EUA. A primeira crença requer simplesmente que o estudante realize em seu cérebro a sentença da LOT “O Primeiro-ministro britânico na Segunda Guerra Mundial nasceu nos EUA” e a processe como estando na “caixa” de crenças, enquanto a sentença “Winston Churchill nasceu nos EUA” está fora da dita “caixa” ou ausente por completo. A CTM também mostra como é possível a formação de infinitos pensamentos com um estoque finito de recursos, a saber, por recursão, conforme exposto anteriormente.

Apesar de sua notoriedade, a teoria da mente como um engenho computacional não é exigida nem lógica nem praticamente pelo funcionalismo ou pela ideia de realização múltipla; a ideia essencial é, como explicamos, que os fenômenos mentais são intermediários causais entre *inputs* e *outputs*, mediando as interações do sistema com seu ambiente. Sendo assim, vale mencionar outra formulação influente da teoria, o que deixará claro, mais adiante, que o problema da causação mental aparece também para ela. O mesmo vale, conforme argumentaremos, para todas as demais formulações. Passemos então para a Teoria Causal da Mente de Armstrong.²³⁵

Na seção sobre a teoria da mente como identidade vimos que considerações de evolução científica levaram Smart a concluir que a mente tinha que ser um fenômeno físico. Só assim seria possível evitar a aparição de verrugas nomológicas a desfigurar uma ontologia elegante. O fisicalismo de Armstrong é o fim de um caminho parecido. Armstrong toma como certas as correlações entre eventos cerebrais e mentais, e conclui que uma maneira “muito natural” de

²³⁴ Idem, p. 65-66.

²³⁵ The Causal Theory of Mind (Armstrong, 2002 [1981]).

explicá-las é adotar uma ontologia fisicalista. Neste quadro, “as diferenças entre uma pedra e um corpo humano parecem estar fundamentadas unicamente na disposição material extremamente complexa encontrada no corpo vivo, e que está ausente na pedra”.²³⁶

A teoria de Armstrong é uma tentativa de elucidar os problemas da subjetividade com o uso de uma análise causal. Estados mentais nada mais são que fenômenos capazes de causar determinados efeitos e passíveis de serem causados por determinados fatores. Ao contrário de Smart e Putnam, Armstrong tenta capturar com sua análise não somente sensações, mas também atitudes proposicionais. De fato, Armstrong diz que a clareza de sua concepção do mental aparece sobretudo quando consideramos os objetivos (*purposes*) de um agente.²³⁷ Armstrong observa que é “completamente natural”²³⁸ compreender a vontade de um homem de ir à cozinha fazer um lanche como uma causa que o faz agir de certa maneira, ou seja, leva a determinado comportamento. Não é possível entender desejos (ou objetivos, como quer Armstrong) sem referência aos fatos por eles gerados. Como seria possível diferenciar, prossegue Armstrong, o desejo de ir à cozinha do desejo de ir ao quarto, estando ausente a consideração causal? Evidentemente, existe uma conexão forte entre o mental e comportamentos observáveis; é justamente aí, aliás, que reside a plausibilidade da teorização behaviorista.²³⁹ A ideia ganha em atratividade quando examinamos mais alguns exemplos: ter fome é possuir um estado que tende a gerar busca por comida. Este estado é causado por objetos percebidos pelo sujeito da sensação, ou por outros estados do organismo. Evidentemente, podemos caracterizar estas causas como *inputs*, sendo os comportamentos *outputs*. Sob a mesma perspectiva, *inputs* de agressão conduzem em determinadas circunstâncias (o estado do organismo em um dado momento) a um estado (fúria) causador de comportamentos (luta, ataque verbal) e reações fisiológicas (descargas de adrenalina, postura ameaçadora) que muito naturalmente satisfazem o conceito de *output*. Supondo que haja interesse em uma psicologia válida para espécies diversas, não interessa a realização

²³⁶ Idem, p.81. Minha tradução.

²³⁷ Idem, p. 82-83.

²³⁸ Idem, p. 81. Minha tradução

²³⁹ Ibidem.

material dos estados em questão. Todo sistema capaz de reproduzir estas relações será um sistema que sente fome e fúria.

Evidentemente, os exemplos acima são simplificações. Armstrong deixa claro que sua proposta é abrir caminho para formulações cheias de sutilezas. Os padrões em questão são, afinal, relações causais de extrema complexidade. Um espectro tremendamente variado de eventos mentais controla as sutis variações de comportamento que distinguem animais e homens dos demais objetos. Vale acrescentar que o critério para apontarmos corretamente uma categoria do mental vai além do comportamento ou output mental esperado. É preciso que o output tenha sido gerado de uma maneira específica. Consideremos desejos novamente; o que temos são causas sensíveis a informação (*information-sensitive causes*²⁴⁰). A informação em questão vem com os inputs de percepções e crenças do agente (notar a semelhança com Davidson). O comportamento de ir à cozinha fazer um lanche é idêntico estando o agente com fome ou com vontade de ver sua geladeira nova. Os estados mentais geradores do comportamento, no entanto, são diferentes porque operam sobre as ações de forma diversa. A fome cessa sua operação após uma refeição, o que não ocorre necessariamente com a alternativa. Portanto, comportamentos idênticos estão ligados a itens mentais diversos, cuja particularidade aparece em suas condições de ativação e desativação. Da mesma maneira que seu compatriota Smart, Armstrong lembra que sua formulação é mais um programa de pesquisa do que uma teoria “desenvolvida”, daí sua incapacidade, em um primeiro momento, de gerar explicações mais rigorosas.²⁴¹

Armstrong cita três fortes razões para a adoção de sua teoria. A primeira nos é particularmente interessante, pois é sugerida como demonstração da inadequação da ITM.²⁴² O funcionalismo, diz Armstrong, é capaz de demonstrar *por que* certos processos neurológicos originam nossas sensações. Se nos perguntarmos sobre o porquê de atividades neurais na amígdala gerarem emoções de medo, uma teoria funcionalista avançada nos responderá: *porque estes processos realizam tal e tal cadeia causal*. A ITM presumivelmente não teria uma resposta adequada, tendo que tomar a identidade entre temor e processos da

²⁴⁰ Idem, p. 83. Minha tradução.

²⁴¹ Ibidem.

²⁴² Idem, p. 86.

amígdala como um “fato bruto”. Melhor dizendo, com o funcionalismo estamos aptos a responder a pergunta “Como é possível que estados mentais sejam estados físicos do cérebro?”²⁴³ O funcionalismo de Armstrong, por ele denominado Teoria Causal da Mente (*Causal Theory of Mind*, doravante TCM) é sugerido, portanto, como um esquema que incorpora eventuais avanços neurocientíficos em uma base teórica rica e independentemente. Esta é promessa, mantida mais explicitamente por Putnam, de uma psicologia cuja abrangência supera mesmo barreiras entre espécies biológicas.

Sobre as demais vantagens apregoadas por Armstrong, bastar-nos-á mencioná-las brevemente. Armstrong afirma primeiramente que sua TCM explica satisfatoriamente nossas intuições dualistas. Tendemos para o dualismo, diz Armstrong, devido a nossa incapacidade de capturar a essência da introspecção com a própria introspecção. Tudo o que nosso “olho da mente” consegue enxergar é a conexão causal da mente com o comportamento. As qualidades intrínsecas do mundo mental permanecem invisíveis a este olhar interior. Armstrong exemplifica esta tese refletindo sobre desejos: sabemos que os mesmos são fatores que nos impelem a agir desta ou daquela maneira. Os atos resultantes são passíveis de uma compreensão que nos falta quando nos voltamos para suas causas mentais. Diante desta incompreensão, somos levados a concluir que o mundo mental é distinto de tudo que nos rodeia, e esta diferença é expressa supondo-se um caráter imaterial. Evidentemente, na TCM esta interpretação é um equívoco – o compromisso fisicalista da teoria é explícito.²⁴⁴

A segunda vantagem anunciada por Armstrong é uma explanação simplíssima da intencionalidade dos eventos mentais. “Intencional” aqui significa aquilo que “aponta” ou “é sobre” algo. Desejos de comer morangos “apontam” para morangos, os objetos intencionais da vontade em questão. É claro que objetos intencionais não precisam efetivamente existir; pensamentos sobre a Maldição de Tutankhamon não são menos reais que desejos por morangos, não obstante a inexistência da Maldição. Seja como for, Armstrong, fiel ao gosto pelo macabro tão comum na filosofia analítica, nos pede que consideremos o conceito

²⁴³ Ibidem. Minha tradução

²⁴⁴ Idem, p. 84.

de veneno como conceito causal.²⁴⁵ Ao fazê-lo, obtemos uma perspectiva onde qualquer substância capaz de fazer alguém adoecer ou causar sua morte por meio de um “processo biológico”²⁴⁶ é um veneno. Portanto, a definição de “veneno” é obtida a partir de uma análise de seus efeitos no mundo e *como* estes efeitos são produzidos. Armstrong acrescenta que é legítimo ver tais substâncias como “apontando” para seus efeitos. Algo análogo ocorre com eventos mentais e seus efeitos. Desejar comer morangos é apontar para os frutos de uma dada maneira porque os mesmos estão presentes na cadeia causal que realiza o comportamento indicado pelo desejo. Armstrong reforça esta ideia com uma comparação do mapeamento que um foguete faz de seu destino. O foguete está para seu alvo como um *gourmand* para seu repasto. Sendo assim, a intencionalidade do mental emerge da elaboração das condições de satisfação de um sem-número de cadeias causais.²⁴⁷

Com tantos argumentos sólidos a favor do funcionalismo e teorias aparentadas, o que falta para que tais ideias sejam integradas a uma metafísica consistente? O esquema funcionalista/computacional parece, à primeira vista, razoável como integrante de uma ontologia fisicalista: existem propriedades funcionais caracterizando um *token physicalism*, mas estas propriedades nada estão fazendo de milagroso ou agindo como penduricalhos nomológicos. Tais propriedades, afinal, são realizadas por objetos legitimamente físicos, não havendo espaço para causação ou determinação *sui generis* via leis psicofísicas excêntricas. Desafortunadamente, a cessão de poderes causais às propriedades físicas por parte do domínio mental põe em dúvida a relevância deste no mundo. Uma vez assumida a realização física de todas as propriedades mentais, fica estabelecida uma exclusividade do físico no campo das relações causais. Eventos mentais de fato têm efeitos no mundo, mas não *enquanto* eventos mentais. As propriedades físicas estão encarregadas de todo o trabalho. Melhor dizendo, eventos mentais causam eventos físicos *porque também são eventos físicos* (se assumirmos o *token physicalism*). Tudo indica que o funcionalismo parece empurrar o mental para o abismo da inércia epifenomenal. Os realizadores só influenciam o domínio físico enquanto eventos físicos, sem deixar espaço para o

²⁴⁵ Idem, p. 82-84.

²⁴⁶ Idem, p. 82. Minha tradução.

²⁴⁷ Idem, p. 82, 84-85.

mental, a menos que haja uma superdeterminação sistemática na causação mental (superdeterminação é a junção de duas ou mais cadeias causais distintas, cada uma delas suficiente para o efeito conjunto). Torna-se clara agora uma vantagem da ITM em relação às teorias funcionalistas/de múltipla realização: a questão de propriedades psicológicas inertes cavalgando eventos físicos não existe, pois as propriedades mentais *são* as propriedades físicas da ação. Ainda que motivada por ideias engenhosas, o funcionalismo e seus congêneres está preso no lamaçal da causação mental, e escapar não é uma questão simples, como pretendemos mostrar mais à frente.

David Chalmers é um filósofo que vem tentando barrar o caminho rumo a uma teoria materialista da consciência. Chalmers procura mostrar²⁴⁸ que teóricos reducionistas em geral confundem dois tipos distintos de fenômeno, o “psicológico” e o “fenomenal”, e as teorias resultantes são explicações apenas para o primeiro tipo. Ao descartarem o “aspecto fenomenal da mente”, eles cometem precisamente o erro que Dennett atribui aos chamados “reducionistas gananciosos” [referência aqui] (embora Chalmers não utilize o termo): ignoram algo que jamais poderia ser deixado de lado, sob pena de falsificar o fenômeno. Vejamos o que Chalmers tem em mente quando utiliza esta distinção. O conceito *psicológico* de “mente”²⁴⁹ abarca esta como uma causa ou explicação para o comportamento de um determinado organismo. Um evento mental psicológico é a instanciação de um processo cognitivo (raciocínio, representações mentais de objetos, operações feitas com tais representações etc.). Estes processos podem ou não estar acompanhados de experiência consciente. Já o conceito *fenomenal* de “mente”²⁵⁰ abrange unicamente experiências conscientes; eles caracterizam a mente pela maneira como *sentimos* nossos processos mentais. O conceito psicológico, ao contrário, é caracterizado apenas pelo que a mente *faz*. Ambos os aspectos são reais, e uma boa teoria da consciência, diz Chalmers, não pode excluir nenhum dos dois. Processos mentais conscientes têm uma “vida dupla”;²⁵¹ a ideia de percepção, por exemplo, pode ser entendida psicologicamente, como uma representação mental de certo objeto, ou como uma experiência subjetiva,

²⁴⁸ *The Conscious Mind* (Chalmers, 1996).

²⁴⁹ *Ibidem*, p. 24-31.

²⁵⁰ Chalmers, 1996, p. 24-31

²⁵¹ *Ibidem*, p. 16.

uma sensação privada, daquilo que é representado. Eventos mentais do segundo tipo, embora não sejam redutíveis ao nível psicológico, estão sempre *associados* a este (à ocorrência de um processo cognitivo).²⁵²

Na filosofia de Chalmers, pesquisadores fisicalistas que tentam explicar a subjetividade com um modelo “mecânico”, uma dinâmica física, estão tratando de algo distinto de experiências conscientes, quer saibam ou não.²⁵³ Quem investiga a mente da primeira maneira está abordando o “Problema Fácil”, e quem leva a sério a subjetividade a ponto de não querer reduzi-la a um processo físico ou funcional está tratando do “Problema Difícil”. Existem, portanto, dois problemas mente-corpo. O primeiro é melhor compreendido como um conjunto de problemas, como por exemplo, como agem os mecanismos físicos do aprendizado, como os neurônios armazenam informação, de que forma o cérebro representa eventos etc. Todas estas questões tratam de propriedades funcionais, instanciadas por sistemas físicos. Sendo assim, elas são problemas empíricos, que, embora difíceis, nada possuem de misterioso. A ciência cognitiva pode dar conta do “Problema Fácil”, pois não há nenhum obstáculo metafísico a superar. O “Problema Fácil” abarca, portanto, questões que, por mais interessantes que sejam, não chegam a ser mistérios – existe um programa de pesquisa razoavelmente bem definido para estudar este tipo de questão. Chalmers compara as ciências da mente com a filosofia da biologia, que lida com problemas técnicos de conceitos já bem compreendidos, como evolução, adaptação e especiação. Não consideramos estes conceitos como mistérios metafísicos; o mesmo vale para os aspectos psicológicos, ou cognitivos, da mente.

O “Problema Difícil”, por sua vez, continuaria nos incomodando mesmo que tivéssemos uma teoria completa da mente enquanto cognição. Segundo Chalmers, podemos sempre nos perguntar *como é possível um determinado processo psicológico ser acompanhado por experiência fenomenal*. Não possuímos leis para conectar o nível psicológico ao fenomenal. Além disso, Não parece incoerente conceber uma criatura com processos psicológicos idênticos aos nossos, mas sem experiência subjetiva. Em outras palavras, uma teoria que pode

²⁵² Os detalhes da correlação não nos interessam no momento, mas serão importantes quando a discussão estiver voltada para a superveniência, como veremos mais à frente.

²⁵³ Ibidem, p. 12-13.

explicar nossos processos psicológicos e conectá-los ao nosso comportamento ainda teria que explicar por que estes processos são conscientes, por que nós os *experimentamos*. Dito de outra forma, Chalmers propõe uma investigação do “aspecto qualitativo” da nossa vida mental, nossos *qualia*. Chalmers argumenta que as ciências cognitivas não fizeram progresso nenhum neste sentido e nem poderiam, já que tratam de uma questão diferente.

A filosofia de John Searle também exige uma rejeição do reducionismo fisicalista, movida por razões parecidas com as de Chalmers. Como este, Searle crê que teorias fisicalistas não podem ser teorias da consciência porque partem da negação dos dados.²⁵⁴ A identificação da consciência com processos cerebrais ou funcionais não pode ser uma contribuição para o estudo da subjetividade porque é o comprometimento com sua inexistência. Da mesma forma que Chalmers, Searle distingue os *qualia* dos estados cerebrais objetivos. Searle, porém, não acredita, como Chalmers, que todo vocabulário mentalístico tem um duplo sentido. Quando falamos de “dor”, por exemplo, estamos nos referindo unicamente ao sentido fenomenal, aos *qualia*. Não há dores puramente cognitivas, nem desejos, crenças etc., no sentido puramente “psicológico” de Chalmers. A distinção feita por Searle consiste em diferenciar *causa* e *efeito*. Os estados cerebrais, analisáveis por métodos científicos, *causam* nossas experiências, mas não se confundem com elas. O problema da consciência, diz Searle, consiste em entender o que o cérebro faz para causar nossa subjetividade, sem cometer o erro de confundi-la esta com o cérebro em si. Podemos perceber aqui uma similaridade considerável da posição de Searle com a de Chalmers, que pretende investigar como os processos psicológicos de terceira pessoa (objetivos) são *acompanhados por* experiências conscientes. Tais experiências têm, para Searle, uma “ontologia subjetiva”; sua existência é inseparável de uma perspectiva de primeira pessoa. Em outras palavras, dores, crenças, desejos, e outros estados mentais subjetivos são sempre os estados mentais de alguém. Searle pensa poder concluir daí que todas as teorias fisicalistas da consciência são quiméricas. Tudo o que os seus proponentes obtêm é eliminar os *qualia* por explicação, e isso é inaceitável. Tentar fazê-lo é dar um passo em falso, já que não podemos duvidar que somos conscientes. Searle conclui, como Chalmers, que o fisicalismo está condenado desde o início a

²⁵⁴ *O Mistério da Consciência*, p. 117-133, 137-147.

falsificar o problema. Qualquer processo cerebral (ou conjunto de processos) detectado por métodos científicos e correlacionado com a consciência será candidato à *causa* desta (na visão de Chalmers, o correlato psicológico dela), mas não pode haver identidade entre as duas coisas. Experiências conscientes são intrinsecamente subjetivas.

Desafortunadamente, Searle não mostra a seus leitores como a subjetividade intrínseca às experiências conscientes deve orientar as pesquisas empíricas sobre as mesmas. A filosofia da consciência de Searle contrasta fortemente com as de Fodor, Armstrong, Putnam e Smart, já que lhe falta conteúdo empírico. O foco da investigação de Searle é deixar claro quais são os dados a serem explicados, e para tanto ele nos fornece uma fenomenologia. Não é necessário examinar aqui a longa lista de propriedades da consciência examinadas por Searle nesta fenomenologia (intencionalidade, familiaridade, a dimensão do prazer, entre várias outras).²⁵⁵ O que interessa aqui é mostrar que a sua postura antirreducionista o conduz à armadilha da causalção mental, que ameaça igualmente a filosofia de Chalmers.

Conforme exposto acima, Chalmers defende uma postura explicitamente dualista em sua metafísica da mente. Assumindo que nossas intuições de que a mente consciente dirige nossas ações, o problema da causalção mental reaparece. Já vimos que a dissecação dos processos químicos de corpos animais nada revela de inexplicável acerca das forças envolvidas. Tudo o que existe é englobado por teorias microfísicas fundamentais, não restando papel para forças mentais. Portanto, é correto assumir o fechamento causal do mundo físico e a consequente irrelevância causal do domínio fenomenal de Chalmers. Não é injusto concluir que a consciência irreduzível tem que ser epifenomenal. Interessantemente, o filósofo australiano percebe o problema²⁵⁶ e o abraça: admitindo que sua teoria é incompatível com o fechamento causal do mundo físico, Chalmers tenta mostrar que uma consciência de caráter epifenomenal talvez não seja algo tão desastroso:

²⁵⁵ *The Rediscovery of the Mind*, p.127-150.

²⁵⁶ *The conscious mind*, p.150.

Na segunda parte [de sua discussão do epifenomenalismo], considerarei as razões pelas quais o epifenomenalismo pode ser considerado indesejável, e analisar suas forças enquanto argumentos. Se estas intuições não se traduzirem em argumentos convincentes, pode ser que o epifenomenalismo implicado por esta posição é apenas contra-intuitivo, e no fim das contas um grau de epifenomenalismo pode ser aceito.²⁵⁷

É mais oportuno abordar as respostas de Chalmers em uma discussão mais minuciosa do que é um epifenômeno. O que interessa agora é mostrar a amplitude do problema, que perpassa diversas filosofias. Por isso podemos voltar a Searle, que dá como certa a existência da causação mental, ridicularizando o flerte de Chalmers com os epifenômenos.²⁵⁸ É óbvio, diz ele, que sensações causam movimentos corporais, tendo, portanto, efeitos no mundo físico. Infelizmente, a filosofia de Searle torna incompatíveis este aparente truísmo e o fechamento causal do mundo físico. Se experiências conscientes são metafisicamente distintas dos estados físicos que as causam, desaparece a inteligibilidade das relações causais entre as duas realidades. Os apelos de Searle por uma teoria mais plausível para o senso comum resultaram numa metafísica inconsistente. Não surpreende que Searle não consiga categorizar adequadamente o tipo de relação causal entre estados físicos e estados mentais. Isto se torna claro na sua discussão sobre reducionismo. A mente intrinsecamente subjetiva é uma

propriedade causalmente emergente de sistemas da mesma maneira que solidez e liquidez são propriedades de sistemas de moléculas. A existência da consciência pode ser explicada pelas interações causais entre elementos do cérebro no nível microscópico, mas a consciência ela mesma não pode ser deduzida ou calculada a partir da mera estrutura física dos neurônios sem um exame complementar das relações causais entre eles.”²⁵⁹

A passagem citada descreve um tipo de emergência causal que poderia ser aceita por qualquer teórico da identidade e até mesmo por funcionalistas como Putnam e Armstrong. Se a filosofia da mente de Searle fosse coerente com esta estrutura de emergência, poder-se-ia alegar que ali não há problema de causação mental. O objetivo mais importante de Searle, porém, é mostrar que a filosofia da mente praticada pelos ditos filósofos é uma grande confusão. Dado que é possível localizar a controvérsia na ontologia da emergência, temos uma contradição: Searle utiliza a mesma ontologia que pretende rejeitar. A subjetividade intrínseca

²⁵⁷Idem, p. 150-151. Minha tradução

²⁵⁸ *O Mistério da Consciência*, p. 170-171, 177.

²⁵⁹ Idem, p. 112. Minha tradução

da mente, afinal, contrasta fortemente com as propriedades ordinárias de solidez e liquidez, já que a “emergência” das mesmas preserva o caráter *totalmente objetivo* das suas “causas”. Ambas as propriedades, afinal, são vistas consensualmente como redutíveis a propriedades físicas de determinados sistemas ou pelo menos realizadas por elas. Ao utilizar esta noção de emergência, Searle contradiz seu compromisso com a singularidade do mental. Tal compromisso, como vimos, coloca o teórico no beco sem saída de uma interação indizível. A aceitação, por parte de Chalmers, do epifenomenalismo é, portanto, uma exigência de sua posição, e não a excentricidade grotesca vista por Searle. No entanto, conforme veremos mais adiante, a aparente opção pela coerência também traz suas próprias contradições.