
CONFERÊNCIAS PETRUS HISPANUS 1998

O MENTAL E O FÍSICO

HILARY PUTNAM
UNIVERSIDADE DE HARVARD

CONFERÊNCIA I

CAUSALIDADE MENTAL¹

Hilary Putnam

Peço a vossa indulgência por começar esta Conferência Petrus Hispanus com uma nota autobiográfica. Embora, lamentavelmente, eu não fale a vossa bela língua, o português era uma presença constante em casa dos meus pais. A explicação para isto é dada na introdução à bibliografia do meu pai, intitulada *Samuel Putnam, Latin Americanist*². Nesta introdução, o autor da bibliografia explica:

¹ Esta conferência e a seguinte fazem parte de um volume que contém as três Royce Lectures on the Philosophy of Mind que apresentei na Universidade de Brown em Novembro de 1997. Estas conferências serão publicadas em *Philosophy and Phenomenological Research* e também, juntamente com as Dewey Lectures que apresentei na Universidade de Columbia em 1994, num volume a ser publicado pela Columbia University Press (o título ainda não está decidido).

² *Samuel Putnam, Latin Americanist: A Bibliography*, C. Harvey Gardiner, Carbondale; The Library, Southern Illinois University, 1970.

Algures na Europa, numa data que não pode ser determinada com precisão, Sam Putnam travou conhecimento com Fidelino de Figueiredo. Rejeitando a esterilidade da literatura do seu Portugal natal nos dias da ditadura de Salazar, Figueiredo disse a Putnam: «Porque não vai para o Brasil? Eles é que têm uma literatura interessante». Passaram muitos anos antes que Putnam fosse ao Brasil em pessoa, mas pouco depois do seu regresso aos Estados Unidos, em 1933, dirigiu parcialmente, de facto, as suas atenções intelectuais em direcção ao sul. Assim como uma figura literária portuguesa o tinha levado até ao Brasil, também esse gigante de entre os países da América do Sul o levaria até outras partes da América Latina. Contudo, por muito empenhada que viesse a ser a sua incursão na América de língua espanhola, o seu interesse principal continuaria a ser o Brasil. Quinze anos depois de tudo começar, Sam declararia com convicção que o Brasil era a sua segunda casa.

Não é surpreendente, portanto, que eu tenha ouvido falar português muito frequentemente enquanto o meu pai foi vivo. E fazer uma conferência no país que deu início ao envolvimento do meu pai com essa língua é para mim, peculiarmente, um «regresso a casa». Seja-me também permitido dizer que estou grato pela oportunidade que este convite me proporciona de partilhar a investigação em que estou envolvido neste momento — investigação relacionada com debates actuais na filosofia da mente. Visto que, apesar de eu discordar dos seus pontos de vista, não há um representante desses pontos de vista tão competente, íntegro e genuinamente dedicado à filosofia como Jaegwon Kim, da Brown University, escolhi a sua posição como aquela a examinar e a criticar — e, desse modo, propor uma maneira muito diferente de ver a filosofia da mente. No entanto, começarei com uma citação não de Kim, mas de William James.

William James, a certa altura, faz a notável experiência mental de imaginar um ser humano que (de modo não detectável) é destituído de quaisquer propriedades mentais:

[...] Imaginei aquilo a que chamei uma «namorada automática», querendo com isto significar um corpo sem alma que fosse absolutamente indistinguível de uma moça espiritualmente animada, que risse, falasse, corasse, cuidasse de nós e desempenhasse todas as tarefas femininas com tanto tacto e doçura com se tivesse uma alma. Será que alguém a consideraria como uma igual relativamente a qualquer moça propriamente dita? Certamente que não³.

³ *The Works of William James*, F. Burkhardt, F. Bowers e I. Skrupskelis (orgs.), vol. 2, *The Meaning of Truth*, Cambridge, MA, Harvard University Press, 1975, pág. 103, nota 2.

HILARY PUTNAM

A preocupação parece sem dúvida rebuscada. Contudo, o tópico surgiu de novo num debate recente entre Jaegwon Kim e Donald Davidson.

Nos ensaios reunidos no seu livro *Supervenience and Mind*, Kim dedica uma enorme atenção à filosofia da mente de Donald Davidson. Embora Davidson seja um materialista que vai ao ponto de acreditar que cada acontecimento mental individual é *idêntico* a um acontecimento físico individual, não há, de acordo com ele, leis psicofísicas. No entanto (na maior parte dos artigos de *Supervenience and Mind*), Kim defende a possibilidade de se descobrirem leis desse tipo. Como ele diz, ainda que certos argumentos meus⁴ mostrem que jamais descobriremos leis *irrestritas* da forma $F \leftrightarrow M$, onde F é uma propriedade física, M é uma propriedade mental, e a lei quantifica sobre todos os organismos fisicamente possíveis, pode muito bem haver, todavia, «leis-ponte próprias de uma espécie», i.e., leis da forma $E_i \rightarrow (M \leftrightarrow F_i)$ as quais, *relativamente a uma espécie ou estrutura* E_i , especificam um estado físico F_i como *necessário e suficiente* para a ocorrência do estado mental M .⁵ De facto, a posição de Kim em «The Myth of Nonreductive Physicalism» é a de que só se *houver leis desse tipo* é que as propriedades mentais poderão ter uma eficácia causal genuína.

Antes de mais, eis a descrição que Kim faz da posição de Davidson:

Os acontecimentos mentais, observou Davidson, têm relações causais com os acontecimentos físicos. Mas as relações causais têm de estar apoiadas em leis; isto é, as relações causais entre acontecimentos individuais têm de exemplificar regularidades legiformes⁶. Uma vez que (de acordo com Davidson) não há leis do mental, sejam psicofísicas sejam puramente psicológicas, qualquer relação causal que envolva um acontecimento mental tem de exemplificar uma lei física, de onde se segue que o acontecimento mental tem uma descrição física ou exemplifica

⁴ Kim está a referir-se, em particular, ao meu «The Nature of Mental States», reimpresso nos meus *Philosophical Papers*, vol. 2, *Mind, Language and Reality*, Cambridge University Press, Cambridge, 1975.

⁵ Kim, *loc. cit.*, pág. 273. Kim faz notar correctamente que o funcionalismo que eu defendi em «The Nature of Mental States» *pressupõe* a natureza dessas «leis-ponte próprias de uma espécie». No meu *Representation and Reality*, defendo que tais «leis-ponte próprias de uma espécie», tal como as leis gerais da forma $F \leftrightarrow M$, não podem ser descobertas.

⁶ Embora não se refira a Davidson pelo nome, Elisabeth Anscombe, na sua conferência inaugural em Cambridge («Causality and Determinism» reimpresso no seu *Collected Papers*, Vol. 2, Blackwell, Oxford, 1981, pp. 133-147), proporciona contra-exemplos poderosos a qualquer ponto de vista que implique que todas as asserções causais singulares têm de estar apoiadas numa regularidade legiforme.

uma espécie de acontecimento físico. É que um acontecimento é físico (ou mental) se exemplifica uma espécie de acontecimento físico (ou uma espécie de acontecimento mental). Segue-se, portanto, que todos os acontecimentos são acontecimentos físicos — sob a suposição de que todo o acontecimento participa em pelo menos uma relação causal⁷.

E eis a crítica de Kim:

A ontologia de Davidson reconhece acontecimentos individuais como particulares espaço-temporais. E a estrutura principal subjacente a estes acontecimentos é uma estrutura causal; a rede de relações causais que conecta os acontecimentos uns com os outros é aquilo que confere uma estrutura inteligível a este universo de acontecimentos. Que papel desempenha o mental, segundo o monismo anómalo de Davidson, na determinação dessa estrutura? Nenhum. Pois o monismo anómalo implica o seguinte: *exactamente a mesma rede de relações causais se verificaria no mundo de Davidson se redistribuíssemos as propriedades mentais pelos seus acontecimentos da maneira que quiséssemos; não alteraríamos uma única relação causal se, ao acaso e arbitrariamente, re-atribuíssemos propriedades mentais a acontecimentos, ou até se removéssemos completamente o mental do mundo*. Não esqueçamos que, segundo o monismo anómalo, os acontecimentos são causas ou efeitos apenas na medida em que exemplificam leis físicas, e isto significa que as propriedades mentais de um acontecimento são causalmente ineficazes⁸.

A defesa de Davidson

Começo por defender Davidson contra o argumento de Kim⁹. Não é que eu partilhe, deixem-me dizê-lo desde logo, o monismo anómalo de Davidson ou o ponto de vista de Davidson acerca da causalidade. Acerca do monismo anómalo, a minha posição é a de que os acontecimentos mentais não são nem «idênticos» nem «não idênticos» aos acontecimentos físicos; não creio que tenha aqui sido dado um sentido à noção de «identidade». Creio, além disso, que o artigo

⁷ «The Myth of Nonreductive Materialism», pág. 269.

⁸ *Ibid.*, pp. 269-270.

⁹ Há, contudo, outros argumentos de Kim que me parecem reveladores da inadequação da tese de Davidson. Especificamente, parece, com efeito, que apesar de (segundo a tese de Davidson) os acontecimentos mentais poderem causar acontecimentos físicos, o facto de esses acontecimentos mentais terem *predicados* mentais não desempenha qualquer papel em explicar por que razão eles o podem fazer. Isto decorre do compromisso de Davidson com a ideia de que são as leis estritas e apenas as leis estritas que fundamentam as relações causais. Ver os artigos de Davidson, Kim, McLaughlin e Sosa em John Heil e Alfred Mele, orgs., *Mental Causation*, Oxford University Press, Oxford, 1993.

HILARY PUTNAM

«Causality and Determination»¹⁰, de Elizabeth Anscombe, contém contra-exemplos conclusivos contra a tese de Davidson de que todas as asserções causais singulares verdadeiras têm de ser suportadas por uma lei estrita¹¹. Contudo, creio que a objecção de Kim à posição de Davidson pode ser rebatida, e que tentar rebatê-la é uma excelente maneira de analisar alguns dos problemas centrais da filosofia da mente.

O argumento de Kim depende, de modo evidente, da seguinte notável condicional:

(Supondo que o ponto de vista de Davidson é correcto.) Não alteraríamos uma única relação causal se, ao acaso e arbitrariamente, re-atribuíssemos propriedades mentais a acontecimentos, ou mesmo se removéssemos completamente o mental do mundo.

Uma consequência imediata desta condicional é a seguinte:

(AUTÓMATOS) — (Supondo que o ponto de vista de Davidson é correcto.) Mesmo que certas pessoas não tivessem quaisquer propriedades mentais, desde que todas as suas propriedades físicas e ambientes físicos fossem os mesmos, ocorreriam os mesmos acontecimentos físicos.

Na prática, Kim convidou-nos a fazer a mesma experiência mental que James: (supondo que a filosofia da mente de Davidson é correcta) imaginemos que havia um mundo em que um corpo exactamente como o meu (ou como o vosso) estava exactamente no mesmo estado físico e exactamente no mesmo ambiente físico que o meu (ou que o vosso), mas era desprovido de quaisquer propriedades mentais. Não teria ele o mesmo comportamento que o meu (ou que o vosso)? Esse corpo seria exactamente como o corpo da «namorada automática» imaginada por James, um eu automático ou um cada-um-de-vocês automático. E do facto de que (como Kim tem de defender) o Hilary Putnam automático abriria a torneira do banho tão eficientemente como um com propriedades mentais, ainda que o acontecimento mental, «a decisão de abrir a torneira do banho», não tivesse tido lugar, Kim conclui que (se a filosofia da mente de David-

¹⁰ Reimpresso em Ernest Sosa, org., *Causation and Conditionals*, Oxford University Press, Oxford, 1995.

¹¹ Um outro problema da tese de Davidson sobre causalidade é o de que nem todas as leis físicas são leis *causais*, e Davidson não oferece nenhuma maneira de distinguir uma conexão entre acontecimentos que seja genuinamente causal de uma que seja nomológica sem ser causal.

son é correcta) esse acontecimento mental *não tem eficácia causal*. As propriedades mentais são meros «epifenómenos».

Claro que o argumento pode ser bloqueado negando-se que a contrafactual «faça sentido». E para a contrafactual não fazer sentido, Kim acha que teria de ser verdade que as nossas propriedades mentais fossem de algum modo redutíveis às nossas propriedades físicas (ou, mais precisamente, fossem «fortemente sobrevenientes» em relação a elas de algum modo satisfatório¹²). Mas a redução (incluindo a «sobreveniência forte») exige leis psicofísicas, as quais a posição de Davidson exclui. Assim, Kim confronta-nos com um dilema: ou Davidson não tem razão e há leis psicofísicas, ou então as nossas propriedades mentais são epifenoménicas. Ou o epifenomenalismo ou o reducionismo!

Preparando o terreno: os defeitos da contrafactual de Jaegwon Kim

Suponhamos que eu estou a fazer correr a água para o banho. Se alguém me perguntasse «Porque é que a água está a correr?», eu poderia responder «Estou a encher a banheira porque quero tomar um banho», mas também podia dizer simplesmente «Decidi tomar um banho». E ninguém duvidaria que eu tinha *explicado* porque é que a água estava a correr na casa de banho.

John Haldane fez a observação inteligente¹³ de que «há tantos tipos de causa quantos os sentidos de “por causa de” ou de “porque”». (Ele acrescentou que «a doutrina aristotélica das quatro causas era apenas uma classificação preliminar».) Se o «porque» da explicação precedente está correcto, então, num certo sentido de «causa», a minha decisão de abrir a torneira até à posição intermédia (entre «quente» e «frio») é a *causa* de estar agora água a correr (com uma certa temperatura agradável). As decisões (as quais estão

¹² «Postscripts on Supervenience», secção 2, pp. 167-168.

¹³ Mas nem ele nem eu nos lembramos onde! Em «Reductionism and the Nature of Psychology», defendi o ponto de vista de que as explicações buscam o seu próprio nível; as explicações em termos de microcausalidade não entram em conflito com, por exemplo, explicações psicológicas, *porque generalizam a partir de classes de casos muito diferentes*, e uma vez que um caso pode pertencer a classes de generalização muito diferentes, pode simultaneamente ter mais do que uma explicação. Em conversa, Jaegwon Kim disse-me que alimenta agora a ideia de que a noção fundamental é a de *explicação* e não a de causalidade, e que pensar desta maneira pode tornar inteligível que um mesmo acontecimento tenha «causas» diferentes. Concordo inteiramente!

HILARY PUTNAM

entre as coisas que os filósofos designam de acontecimentos «mentais» ou «psicológicos») podem ser a causa de águas a correr (que estão entre as coisas que os filósofos designam de «acontecimentos físicos»). E supor que o «porque» não é correcto — isto é, supor que *não* é por causa da minha decisão que está neste momento a correr água com a tal temperatura agradável e que de facto a minha decisão foi completamente «ineficaz» — seria rejeitar maneiras de falar que são essenciais a toda a concepção que temos de nós próprios como seres num mundo.

Segue-se então que as nossas propriedades mentais têm, de algum modo, de ser redutíveis a propriedades físicas? Antes de tentarmos desfazer este nó, analisemos primeiro a questão de saber como podemos tornar inócua a contrafactual de Kim.

Bom, não queremos limitar-nos a aceitar, como uma possibilidade completamente inteligível, que algumas pessoas possam não ter quaisquer propriedades mentais e que todas as suas propriedades físicas e os seus ambientes físicos possam ser os mesmos que seriam se elas as tivessem, ao mesmo tempo que defendemos que a consequente da condicional de Kim («Os mesmos acontecimentos físicos ocorreriam») é falsa; i.e., não queremos afirmar:

(NÃO-AUTÓMATOS) — Se certas pessoas não tivessem quaisquer propriedades mentais e todas as suas propriedades físicas e ambientes físicos fossem os mesmos que seriam se elas as tivessem, ocorreriam acontecimentos físicos *diferentes*¹⁴.

(NÃO-AUTÓMATOS) não é algo que consideremos credível.

Claro que o facto de não considerarmos (NÃO-AUTÓMATOS) credível tem tudo a ver com o facto de já não considerarmos o dualismo cartesiano forte (o «interaccionismo») credível. É que (NÃO-AUTÓMATOS) é exactamente aquilo em que Descartes acreditava: ele acreditava que a mente, concebida como uma entidade imaterial que «é o sustentáculo» de todas as nossas propriedades mentais, faz com que os nossos corpos tenham um comportamento diferente daquele que teriam se a mente desaparecesse.

Na conferência de hoje e na seguinte defenderei que nem os problemas clássicos da filosofia da mente, nem os «pontos de vista filosóficos» a que eles dão origem são completamente inteligíveis. Penso que toda esta ideia da mente como um objecto imaterial a «interagir» com o corpo é um excelente exemplo de um ponto de

¹⁴ Repare-se que (AUTÓMATOS) e (NÃO-AUTÓMATOS) são contrárias e não contraditórias.

vista ininteligível em filosofia da mente; mas esta não é a crítica que habitualmente se faz ao interaccionismo.

A crítica habitual foi, de uma maneira muito bela e incisiva, formulada por Bertrand Russell¹⁵. Se o dualismo «interaccionista» for verdadeiro, então o meu corpo tem de ter uma trajectória diferente da trajectória que as leis da física preveriam para ele com base na totalidade das forças físicas que agem sobre ele. O próprio Descartes tinha consciência disto, e uma vez que nem mesmo ele queria postular que, quando os corpos humanos estão envolvidos em interacções físicas, se violam leis básicas como a da conservação da quantidade de movimento, ele supôs que embora a mente possa alterar a direcção em que o corpo se move (agindo sobre o corpo na região da glândula pineal), ela não pode alterar a quantidade total (escalar) de movimento. Mas pouco depois da época de Descartes descobriu-se que *a quantidade de movimento total em cada direcção no espaço é também constante. Assim, se o interaccionismo for verdadeiro, algumas leis físicas da conservação da energia serão violadas quando os seres humanos agem com base em decisões ou outros pensamentos.* Em resumo, o «interaccionismo» implica que *o modo como os corpos humanos se comportam viola as leis da física.* Visto que não há o mais pequeno indício de que isto seja verdade, somos obrigados a rejeitar o interaccionismo do mesmo modo que rejeitamos o vitalismo e outras teorias desactualizadas que postulam que vários fenómenos (a «vida» era um exemplo popular no século XIX) são «excepções às leis da física».

O argumento de Russell é *empírico*. Supõe que o interaccionismo é completamente inteligível (ao contrário do que vou defender) e argumenta que os dados empíricos o infirmam. Mas quer rejeitemos o interaccionismo por razões empíricas, como Russell, quer por acharmos que é parcialmente destituído de sentido, como defenderei que devemos, o facto é que as pessoas com formação científica rejeitam hoje o interaccionismo. Portanto, se também quisermos rejeitar (AUTÓMATOS), não podemos fazê-lo afirmando simplesmente (NÃO-AUTÓMATOS).

Há, de facto, um número apreciável de pontos de vista filosóficos que implicam que (AUTÓMATOS) não é verdade, mas que não implicam (NÃO-AUTÓMATOS) — por exemplo, o fiscalismo reducionista, o behaviorismo lógico e o verificacionismo. Os filósofos que defendem

¹⁵ Veja-se Bertrand Russell, *History of Western Philosophy*, George Allen & Unwin, Londres, 1945.

estes pontos de vista atacariam (AUTÓMATOS) por encontrarem qualquer coisa de errado na possibilidade que (AUTÓMATOS) admite, a qual formulei atrás como «*algumas pessoas poderiam não ter quaisquer propriedades mentais, mas todas as suas propriedades físicas e os seus ambientes físicos poderiam ser os mesmos que seriam se elas as tivessem*». Mais precisamente, esses filósofos aceitariam que, se supusermos que o interaccionismo deve ser excluído, então esta suposta «possibilidade» (ou melhor, a conjunção da «possibilidade» com a negação do interaccionismo) sofre de algum — a ininteligibilidade (de acordo com os verificacionistas), ou a autocontradição (de acordo com os behavioristas lógicos), ou a impossibilidade metafísica (de acordo com alguns fisicalistas reducionistas — como veremos, outros fisicalistas reducionistas argumentariam que essa «possibilidade» é *irrelevante* se estivermos interessados no que acontece no mundo actual).

O verificacionismo defende que se a «possibilidade» descrita fosse actual^{*}, então, visto que rejeitámos o interaccionismo, as pessoas em questão teriam de agir exactamente como se tivessem propriedades mentais (repare-se que isto é exactamente a condicional de Kim; mas isso não significa que o verificacionista aceite a condicional de Kim; pelo contrário, ele está a apresentar um argumento por *reductio*). Mas este estado de coisas seria inverificável (se as pessoas em questão agissem exactamente como se tivessem propriedades mentais, seria em princípio impossível verificar que elas não as tinham) e, logo, todo este cenário é cognitivamente destituído de significado. Ao contrário do verificacionismo, o behaviorismo lógico não supõe necessariamente que todas as proposições com significado são verificáveis em princípio, mas defende que todas as proposições acerca de propriedades *mentais* são logicamente equivalentes a proposições acerca de comportamentos (físicos). Portanto, se supuséssemos que poderíamos imaginar um mundo em que algumas pessoas não têm quaisquer propriedades mentais, mas não apresentam um comportamento diferente daquele que apresentariam se de facto possuísem mentes, estaríamos a tentar imaginar que *as condições logicamente necessárias para a presença de propriedades mentais estavam satisfeitas mas as propriedades não estavam presentes*, o que é uma contradição. Assim, se o verificacionismo está correcto, a conjunção do «não-interaccionismo» com a antecedente de (AUTÓMATOS) é cognitivamente destituída de significado, e se o behaviorismo lógico está correcto, essa conjunção é auto-contradi-

^{*} Na acepção modal (e não temporal) do termo. (*N. do T.*)

tória. Se qualquer um desses pontos de vista estiver correcto, podemos recusar considerar (AUTÓMATOS) como uma contrafactual verdadeira sem sermos forçados a aceitar (NÃO-AUTÓMATOS). Infelizmente, nem o verificacionismo nem o behaviorismo lógico parecem já ser pontos de vista sustentáveis.

O ponto de vista de Donald Davidson é o de que a razão de ser do nosso discurso acerca do mental é permitir-nos «racionalizar» o comportamento dos agentes humanos. Além disso, sustenta ele, essa é a sua *única* função; é por isso que não pode haver leis psicofísicas. Se este ponto de vista estiver correcto, a conjunção de «não interaccionismo» com a antecedente de (AUTÓMATOS) descreve um estado de coisas em que a única justificação para aplicar predicados mentais se verifica, mas em que esses predicados mentais não têm «realmente» aplicação. Davidson consideraria ininteligível, certamente, esta sugestão!

Davidson não é, claramente, um verificacionista, e também não é, certamente, um behaviorista lógico; mas o seu ponto de vista tem uma relação interessante com ambos. Tal como os behavioristas lógicos, ele acredita que se nos comportamos em todos os aspectos como se nos fossem aplicáveis predicados mentais («todos os aspectos» inclui aqui aspectos microfísicos, se os acontecimentos microfísicos puderem tornar-se relevantes para a interpretação do nosso uso da linguagem ou de outro comportamento qualquer), então esses predicados mentais aplicam-se-nos *de facto*; todavia, ao contrário dos behavioristas lógicos, Davidson não acredita que possamos formular verdades conceptuais que estabeleçam ligações entre predicados mentais particulares e predicados comportamentais particulares. Como os verificacionistas, ele acredita que, se um predicado mental se nos aplica, então um observador que fosse «omnisciente» quanto a todos os factos físicos acerca de nós e do nosso ambiente seria capaz de verificar que esse predicado se nos aplica¹⁶. Com efeito, podemos apelidar Davidson de psicoverificacionista — verificacionista acerca do mental. A minha rejeição da inteligibilidade dos vários cenários construídos a partir de (AUTÓMATOS) que considerámos partirá, porém, da estratégia que consiste em defender Davidson nesse ponto. Não adoptarei qualquer das suposições metafísicas (acerca da única finalidade do nosso discurso acerca do mental, etc.) que fazem parte da metafísica da mente de Davidson.

¹⁶ Esta presunção figura explicitamente no artigo de Davidson «A Coherence Theory of Truth and Knowledge».

Voltando agora a nossa atenção para o reducionismo físico (e trata-se de uma versão — altamente sofisticada — do reducionismo físico que Kim defende), aqueles que adoptam esta perspectiva dispõem de diversas estratégias possíveis para enfrentar (AUTÓMATOS) e a ameaça que ele põe à eficácia causal do mental. A mais simples de todas consiste em argumentar que (AUTÓMATOS) é *irrelevante*. Afinal de contas, se for verdade que *no mundo actual* as nossas chamadas «propriedades mentais» são apenas um subconjunto das nossas propriedades físicas, então qualquer mundo em que a antecedente de (AUTÓMATOS) seja verdadeira é um mundo em que as propriedades mentais são «realizadas» de uma maneira muito diferente do mundo actual. Mesmo que as propriedades mentais *fossem* «epifenoménicas» *nesse* «mundo possível», no mundo actual elas certamente não são, porque são físicas e as propriedades físicas são, por definição, não epifenoménicas¹⁷.

Contudo, como tentarei mostrar na próxima conferência, o fisicalismo reducionista é incoerente. Se isto for correcto, e se não quisermos ser verificacionistas (ou «psicoverificacionistas») ou behavioristas lógicos, que outras razões poderemos ter para não aceitar (AUTÓMATOS) e, com (AUTÓMATOS), a conclusão de que, uma vez que todos os acontecimentos físicos ocorreriam do mesmo modo mesmo

¹⁷ Alternativamente (mas de modo mais controverso), o fisicalista reducionista pode apoiar-se em ideias avançadas por um filósofo que é, ele próprio, um opositor deste tipo de reducionismo: Saul Kripke. No seu brilhante e enormemente influente *Naming and Necessity*, Kripke defendeu que, quando descobrimos que uma certa substância ou uma certa propriedade ou magnitude tem uma natureza descritível na linguagem da física ou da química — quando descobrimos, por exemplo, que a água é H_2O , ou que a luz é radiação electromagnética com um comprimento de onda situado algures no espectro electromagnético, ou que a temperatura é energia cinética molecular média — descobrimos também o que essa substância ou propriedade ou magnitude são *em qualquer mundo metafisicamente possível*. Se aceitarmos a tese de Kripke (mas veja-se «Is Water Necessarily H_2O ?» no meu *Realism with a Human Face*), segue-se que *não existe nenhum mundo possível em que a água não seja H_2O* . *A fortiori*, qualquer contrafactual cuja antecedente seja «Se a água não fosse H_2O » é problemática. E o mesmo se aplica a contrafactuais com antecedentes como «Se a luz não fosse radiação electromagnética» e «Se a temperatura não fosse energia cinética molecular média». E, do mesmo modo, se as nossas propriedades mentais forem de facto redutíveis a propriedades físicas da mesma maneira como a luz, a água e a temperatura o são, então qualquer contrafactual cuja antecedente seja «se as nossas propriedades mentais não fossem redutíveis a propriedades físicas» ou «se as nossas propriedades mentais estivessem ausentes e as nossas propriedades físicas se mantivessem» — em particular, a própria contrafactual (AUTÓMATOS) — é problemática. É que, deste ponto de vista, se é um facto empírico que as nossas propriedades mentais são redutíveis a propriedades físicas (coisa em que Kripke não acredita), então a antecedente de (AUTÓMATOS) postula uma *impossibilidade metafísica*.

que as nossas propriedades mentais nos fossem retiradas, essas propriedades mentais são, sem exceção, *epifenômic*as? Como podemos evitar a disjunção de Kim: ou as nossas propriedades mentais são propriedades físicas disfarçadas, ou são epifenômicas? (Kim afirmou-me que ele próprio considera ambas as disjuntas pouco atraentes!)

A inteligibilidade da antecedente de (AUTÓMATOS) reavaliada

Alguns chamariam à perspectiva que vou adoptar nestas conferências uma perspectiva «wittgensteiniana»; mas ainda que a minha discussão não seja completamente destituída de referências à filosofia tardia de Ludwig Wittgenstein, prometo evitar citações das *Investigações* e usar a terminologia de Wittgenstein tão pouco quanto possível¹⁸. Talvez esta promessa alivie a ansiedade que alguns de vocês possam sentir! Especificamente, a minha abordagem ao estimulante dilema de Kim consistirá — o que não é surpresa nenhuma — em questionar a *inteligibilidade* da antecedente de (AUTÓMATOS). Defenderei que essa antecedente

(SEM ALMA) — Certas pessoas não têm quaisquer propriedades mentais, mas todas as suas propriedades físicas são as mesmas que seriam se elas as tivessem, e os seus ambientes físicos são os mesmos.

e também a conjunção de (SEM ALMA) com «a negação do interacionismo», não chegam a ser completamente inteligíveis¹⁹. E defenderei que, visto que nenhum destes supostos «estados de coisas possíveis» é suficientemente inteligível, a questão de saber o que aconteceria se algum deles se verificasse é destituída de sentido. Fazendo uma analogia, a afirmação «a carruagem da Gata Borracheira transformou-se numa abóbora» satisfaz os nossos critérios de inteligibilidade quando *entendida como proferida no contexto de um conto de fadas* (sabemos «o que fazer» com ela, como reagir a ela, como «entrar no jogo»); mas se retirarmos a afirmação «algumas vezes as

¹⁸ Evitarei nestas conferências tanto quanto possível a terminologia característica de Wittgenstein (e.g., «jogo de linguagem»), não porque seja crítico do uso que Wittgenstein faz dessa terminologia, mas porque há agora no mercado tantas exegeses concorrentes que o seu uso num contexto filosófico que não seja especificamente acerca de Wittgenstein cria provavelmente mais confusões do que as que dissipa.

¹⁹ Cf. Cavell acerca da distinção entre ser simplesmente «ininteligível» e «não claramente inteligível». A minha noção do que é a inteligibilidade *completa* é influenciada pela discussão de Cavell, mas ele não deve ser considerado responsável por ela.

carruagens transformam-se em abóboras» de um tipo muito particular de contexto, e tentarmos discutir a questão «o que aconteceria se uma carruagem se transformasse numa abóbora?», como se fosse, por exemplo, uma questão científica séria, estaríamos a dizer algo sem sentido. Do mesmo modo, poderíamos achar interessante uma obra de ficção na qual alguém estivesse apaixonado por uma «namorada automática»; mas isso não implica que falar de «mundos possíveis» em que todos ou alguns de nós fossemos «namorados automáticos» faça sentido²⁰. Porém, isto vai exigir bastante análise, e eu prevejo ter de enfrentar alguns mal-entendidos.

Os mal-entendidos que eu receio têm a ver com a antiga popularidade e actual merecida impopularidade do verificacionismo e do behaviorismo lógico. Como a luta contra o behaviorismo lógico que teve lugar nos anos 50 e 60, e a luta contra o verificacionismo que ainda decorre, foram objecto de muita atenção e ainda ocupam um espaço considerável em colectâneas amplamente usadas de artigos de filosofia da mente, espera-se imediatamente que quem questione a inteligibilidade de um cenário como (SEM ALMA) (ou (SEM ALMA) mais a negação do interaccionismo) tem de ser um verificacionista ou um behaviorista. A seguir, espero conseguir eliminar este mal-entendido mostrando como há razões muito diferentes para questionar a inteligibilidade desses cenários, razões essas que não dependem de um compromisso prévio com uma teoria filosófica (como o verificacionismo) que supostamente proporciona um método geral para avaliar se uma afirmação arbitrária tem significado, ou de um compromisso com as teses do behaviorismo lógico. Na próxima conferência espero mostrar que há razões intimamente relacionadas com estas para questionar a inteligibilidade do reducionismo, e, assim, para considerar *ambas* as alternativas do dilema de Kim como estando bastante longe de ser completamente inteligíveis.

²⁰ Discutirei mais tarde uma sugestão notável feita por Wittgenstein segundo a qual «nada de mental» se passa em algumas pessoas. Podemos aceitar que elas são «autômatos» por serem «uma tribo que queremos escravizar». («O governo e os cientistas revelam que as pessoas desta tribo não têm alma; de modo que podem ser usadas sem escrúpulos para qualquer propósito». *Remarks on the Philosophy of Psychology*, vol. I, pp. 20-22.)

«As palavras só têm sentido no fluxo da vida»²¹

Aquilo para que quero chamar a atenção, agora e na próxima conferência, é para a maneira como diferentes imagens filosóficas acerca do funcionamento da linguagem e do que são os significados (ou melhor, acerca daquilo em que consiste o conhecimento dos significados) afectam as nossas atitudes em relação à esmagadora maioria dos debates filosóficos. Isto é algo sobre o qual Charles Travis, um filósofo cuja obra não é, nem de perto nem de longe, tão conhecida como devia, escreveu profusamente, e também algo sobre o qual Stanley Cavell escreveu bastante, especialmente na sua obra-prima, *The Claim of Reason*. Se me perdoarem o que pode parecer uma digressão, gostaria de dizer algo acerca de certas questões gerais levantadas por estes autores antes de voltar às questões acerca da redutibilidade e da causalidade mental.

A diferença entre as duas imagens filosóficas a que me acabei de referir é descrita dum modo muito claro no livro de Charles Travis sobre a filosofia tardia de Wittgenstein, *The Uses of Sense*²². (A resenha-discussão que Travis faz da filosofia da linguagem de Grice, «Annals of Analysis»²³, é talvez a melhor introdução abreviada a estas questões, ao mesmo tempo que é uma crítica poderosa aos métodos de Grice.) Travis apresenta aquilo a que estou a chamar duas «imagens filosóficas» como diferentes concepções daquilo que poderia ser a semântica de uma elocução numa língua natural. A uma das concepções (aquela da qual Paul Grice foi um dos principais representantes), Travis chama a concepção «clássica»; à outra concepção (menos familiar), que ele atribui a Wittgenstein e a Austin, ele chama «semântica sensível a ocasiões de fala», porque o núcleo desta segunda concepção é a tese de que o conteúdo de uma elocução depende do contexto específico em que é dita, da ocasião específica de fala.

A segunda concepção não nega que as palavras tenham «significados» (isto é, que haja algo a que se possa correctamente chamar «conhecer o — ou um — significado» de uma palavra, e que este conhecimento restrinja os conteúdos que possam ser expressos ao usar-se a palavra com aquilo que se pode considerar como esse

²¹ L. Wittgenstein, *Last Writings*, vol. I. §913. No vol. II, pág. 30. Wittgenstein escreve «O que se passa cá dentro também só tem significado no fluxo da vida».

²² Charles Travis, *The Uses of Sense*, Oxford University Press, Oxford, 1989.

²³ Charles Travis, «Annals of Analysis», *Mind*, vol. 100, 398 (Abril de 1991), pp. 237-263.

significado particular). Aquilo que ela nega é que o significado (ou o conhecimento em questão) determine completamente o que está a ser dito (aquilo que se supõe que seja verdadeiro ou falso, ou se se está, de facto, a dizer algo que seja verdadeiro ou falso) quando uma frase é usada para fazer uma asserção²⁴.

Alguns exemplos podem clarificar esta questão. Eu conheço, certamente, o significado de palavras como «há», «café», «muito», «em cima de», «a» e «mesa». Mas esse conhecimento não determina, por si só, o conteúdo da frase «Há muito café em cima da mesa»; de facto, a frase, enquanto frase, não *tem* um conteúdo determinado independente de elocuições específicas. Além disso, o conteúdo da frase «Há muito café em cima da mesa» que é avaliável quanto ao valor de verdade é altamente sensível às ocasiões em que a frase é proferida: conforme as circunstâncias, a frase pode ser usada para dizer que há muitas chávenas de café em cima numa mesa contextualmente identificável (*há muito café em cima da mesa; sirvam-se*), ou que há sacas de café empilhadas sobre a mesa (*há muito café em cima da mesa; ponham-no na camioneta*), ou que alguém entornou café em cima da mesa (*há muito café em cima da mesa; limpem-na*), etc.

Eis um outro tipo de exemplo: eu tenho no meu jardim uma árvore ornamental com folhas cor de bronze. Suponhamos que um brincalhão as pinta de verde. Consoante a pessoa que a disser e a pessoa ou pessoas a quem for dita e a razão por que é dita, a frase «A árvore tem folhas verdes», dita com a minha árvore em mente, tanto pode ser verdadeira ou falsa, como não ser claramente nenhuma das duas!

(Reagindo ao exemplo do café, um filósofo da linguagem meu conhecido — adepto da distinção de Grice entre o significado padrão de uma elocução e as suas implicaturas conversacionais — sugeriu que o «significado padrão» de «há muito café em cima da mesa» é que há muitas (quantas?) moléculas de café em cima da mesa. Mas, se isto é verdade, o significado «padrão» é um no qual as palavras nunca são usadas!)

A concepção clássica não nega, é claro, que a referência exacta de algumas palavras (e.g., os marcadores temporais e os indexicais

²⁴ James Conant defende que uma concepção semelhante já intervém no *Tractatus* de Wittgenstein, ao contrário do que dizem as muitas tentativas de classificar o primeiro Wittgenstein como defensor de uma teoria da linguagem mais próxima da de Rudolf Carnap.

comuns) seja sensível a ocasiões de fala; mas trata essa sensibilidade como um fenómeno especial, facilmente integrável numa explicação geral. A concepção clássica da semântica de uma língua natural é «tarskiana»: as condições de verdade são recursivamente associadas a todas as frases de uma língua natural (como na famosa adaptação e modificação que Davidson fez do trabalho de Tarski). A tese que Travis defende é a de que a sensibilidade a ocasiões de fala, longe de ser um fenómeno especial, é a norma.

Recentemente tem surgido uma quantidade apreciável de literatura em linguística acerca deste mesmo fenómeno (que agora é habitualmente chamado «sensibilidade ao contexto» em vez de «sensibilidade a ocasiões de fala»). Os meus exemplos do «café» e do «verde» ilustram como os nomes comuns e os adjectivos podem ter referências muito diferentes em contextos diferentes, *sempre de modo compatível* com o que «significam». Para determinar o que está a ser dito com «Há muito café em cima da mesa» ou com «A árvore agora tem folhas verdes» num contexto específico, é necessário conhecer o «significado das palavras», as restrições implícitas sobre o que pode e o que não pode ser dito usando aquelas palavras, e *também* usar o bom senso, de maneira a perceber o que está a ser dito no contexto dado; e, como Kant disse há muito tempo (embora não nestes termos), não há regras recursivas para o «bom senso» (pelo menos regras que possamos formular!). Como Cavell defende amplamente em *The Claim of Reason*, a nossa «adesão» a um uso novo, o nosso sentido partilhado do que é e do que não é uma projecção natural dos nossos usos anteriores de uma palavra num contexto novo está sempre presente e é essencial à própria possibilidade da linguagem — sem que seja algo que possa ser captado por um sistema de «regras».

Há muitos outros tipos de exemplos para além dos dois que apresentei. Se uma superfície é «plana» ou não, depende do que seja um padrão *razoável* de «planura» no contexto particular. (Contra isto, Peter Unger argumentou uma vez que só um plano euclidiano é «literalmente plano» — de modo que quando digo que o tampo de uma mesa é plano estou a falar em «modo figurado»! Repare-se que a semântica que Unger propõe para «plano» viola completamente o princípio de caridade!) Se executar um concerto é «difícil» ou não depende de quem esteja a perguntar a quem (imaginem uma criança a perguntar a um professor de violino e um violinista profissional a perguntar a outro). Se uma saca de açúcar pesa «um quilo» ou não

HILARY PUTNAM

depende de a pergunta ser feita no supermercado ou no laboratório. Mas chegou a altura de dizer o que é que tudo isto tem a ver com a filosofia.

Consideremos o seguinte cenário discutido por Wittgenstein (*Last Writings* I, §93- §101; cito apenas o §96):

96. Uma tribo que queremos escravizar. O governo e os cientistas revelam que os membros desta tribo não têm alma; de modo que podem ser usados sem escrúpulos para qualquer propósito. Estamos, mesmo assim, naturalmente interessados na sua linguagem; pois é claro que precisamos de lhes dar ordens e de obter deles informação. Também queremos saber o que eles dizem uns aos outros, visto que isso está relacionado com o resto do seu comportamento. Mas também temos de estar interessados naquilo que, neles, corresponde às nossas «elocuições psicológicas», visto que queremos mantê-los capazes de trabalhar; e assim as suas expressões de dor, de incómodo, de prazer na vida, etc., têm importância para nós. De facto, descobrimos também que estas pessoas podem com sucesso ser usadas como objectos experimentais em laboratórios fisiológicos e psicológicos, uma vez que as suas reacções (incluindo as reacções linguísticas) são totalmente idênticas às dos homens dotados de alma [*seelenbegabten Menschen*]. Estou a supor que se descobriu que se pode ensinar a estes *autómatos* a nossa linguagem em vez de deles através de um método bastante semelhante à nossa «instrução».

Aqui Wittgenstein imagina que, com a motivação de explorar e escravizar certas pessoas, as representamos para nós próprios como sendo exactamente como «certas pessoas» que postulámos atrás. Lembram-se?

(SEM ALMA) — Certas pessoas não têm quaisquer propriedades mentais, mas todas as suas propriedades físicas são as mesmas que seriam se elas as tivessem, e os seus ambientes físicos são os mesmos.

Quando deparo com (SEM ALMA) no contexto de um argumento como o de Kim, confesso que fico sem saber o que fazer com ele. Sinto que não consigo conceber qualquer circunstância em que usasse a descrição «essas pessoas não têm quaisquer propriedades mentais, mas todas as suas propriedades físicas são as mesmas que seriam se elas as tivessem, e os seus ambientes físicos são os mesmos». E justamente porque está ausente um contexto, uma «ocasião de fala» — e se nos falta um contexto no qual uma elocução *E* faça sentido, não conseguimos proporcionar um dizendo apenas «Agora estamos a falar filosoficamente», ou dizendo «Consideremos a possibilidade de *E*!» — justamente porque está ausente um contexto, eu não sou capaz de dizer o que significaria (SEM ALMA) ser

verdadeira. Contudo, quando deparo com o §96 de *Last Writings* de Wittgenstein citado atrás, tomo subitamente consciência de que, infelizmente, há uma circunstância na qual podíamos muito bem usar uma tal descrição (ou a versão mais curta, a versão «autómatos sem alma»). Nesse contexto, compreendo-a até bem de mais. Compreendo as elocuições a que Wittgenstein se refere («as pessoas desta tribo não têm alma»; no §97 ele acrescenta «se alguém de entre vós exprimir a ideia de que se tem de estar a passar algo *dentro* destes seres, algo de mental, isso seria objecto de troça, na qualidade de superstição estúpida»), no sentido de ser perfeitamente capaz de acompanhar o que se passa.

Significa isto que eu sou capaz de atribuir um «mundo possível» a (SEM ALMA) no contexto que Wittgenstein construiu? Que sou capaz de dizer o que significaria ser *verdadeira* a versão revelada «pelo governo e pelos cientistas»? De modo nenhum²⁵. Que um acto de propaganda não descreve, depois de pensarmos bem, um estado de coisas que possamos conceber, não significa que ele não funcione eficazmente como propaganda. (Wittgenstein tinha perfeita consciência do que os nazis disseram dos judeus.) Um dos ensinamentos que podemos tirar deste caso é que a «compreensão» é, ela própria, sensível ao contexto; num sentido, eu «compreendo» o que significa dizer que as pessoas da tribo em questão são «autómatos sem alma», compreendo a *função* das palavras, o efeito que se pretende que elas tenham e o efeito que elas de facto têm; mas isso não implica que eu as compreenda no contexto do argumento de Jaegwon Kim! Pelo contrário, para as compreender nesse contexto, eu teria de as compreender *independentemente* do argumento de Kim, uma vez que esse argumento pressupõe a inteligibilidade *prévia* da ideia de que certas pessoas sejam «autómatos sem alma». (Ou, mais precisamente, pressupõe que, se rejeitarmos o reducionismo, temos de considerar que (SEM ALMA) faz sentido — sem nos dizer *como* entender esse suposto «sentido».)

Note-se que, se isto é verdade, o problema do argumento de Kim não é que a antecedente não poderia ser usada para fazer uma asserção que compreendêssemos. Se eu dissesse de um grupo de burocratas «eles são fantásticos — não têm quaisquer propriedades

²⁵ No vol. II de *Last Writings*, §47, Wittgenstein diz que «construímos coisas como a “tribo sem alma” — que no fim deixámos de ter em conta. Seria preciso mostrar que as deixamos de ter em conta».

HILARY PUTNAM

mentais, mas todas as suas propriedades físicas são as mesmas que seriam se eles as tivessem, e os seus ambientes físicos são os mesmos, embora sejam de facto autómatos sem alma», eu poderia estar a dizer algo perfeitamente inteligível (e algo que eu não poderia dizer com a mesma eficácia se dissesse «eles são absolutamente desprovidos de capacidades de juízo, simpatia e compreensão humana»). *Nesse contexto*, as palavras que formam (SEM ALMA) podem ter um conteúdo perfeitamente claro; mas não um conteúdo que seja relevante para as discussões acerca da causalidade mental. O problema da antecedente de Kim não está nas próprias palavras; está em que nós não sabemos o que Kim está a fazer com elas.

Tradução de Pedro Santos