

CÉREBROS EM UMA CUBA¹

Hilary Putnam

Tradução: André L. P. Leones.

Uma formiga rasteja por uma faixa de areia. Conforme o faz, ela traça uma linha no chão. Por puro acaso, a linha traçada faz curvas e se entrecruza de tal forma que acaba parecendo uma caricatura reconhecível de Winston Churchill. A formiga traçou uma imagem de Winston Churchill, uma imagem que *retrata* Churchill?

Após refletir um pouco, a maioria das pessoas dirá que não. Afinal, a formiga nunca viu Churchill ou sequer uma imagem de Churchill, e não teve a intenção de retratar Churchill. Ela apenas traçou uma linha (e mesmo *isso* não foi intencional), uma linha que *nós* podemos “enxergar como” uma imagem de Churchill.

Podemos expressar isso dizendo que a linha não é, “em si mesma”, uma representação² do que quer que seja em vez de qualquer outra coisa. A similaridade (de um certo tipo bastante complicado) com as características de Winston Churchill não é suficiente para fazer com que algo represente ou se refira a Winston Churchill. E isso tampouco é necessário: em nossa comunidade, a forma impressa “Winston Churchill”, o proferimento das palavras “Winston Churchill” e várias outras coisas são usadas para representar Churchill (embora não pictoricamente), ainda que não tenham o tipo de similaridade com Churchill que uma imagem — mesmo uma linha traçada — tem. Se a *similaridade* não é necessária ou suficiente para fazer com que algo represente outra coisa, como *qualquer coisa* pode ser necessária ou suficiente para esse propósito? Como, diabos, uma coisa pode representar (ou “significar” etc.) algo diferente dela?

A resposta pode parecer fácil. Suponha que a formiga viu Winston Churchill, e suponha que ela teve a inteligência e a habilidade para desenhar uma imagem dele. Suponha que ela produziu a caricatura *intencionalmente*. Então, a linha representaria Churchill.

Por outro lado, suponha que a linha tinha a forma WINSTON CHURCHILL. E suponha que isso foi um mero acidente (ignorando a improbabilidade envolvida). Então, a “forma impressa” WINSTON CHURCHILL *não* representaria Churchill, embora a forma impressa represente Churchill quando aparece em quase qualquer livro hoje em dia.

Logo, ao que parece, aquilo que é necessário para a representação, ou o que é essencialmente necessário para a representação, é a *intenção*.

¹ Publicado em *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981, pp. 1-21).

² Neste livro, os termos “representação” e “referência” sempre se referem a uma relação entre uma palavra (ou outro tipo de signo, símbolo ou representação) e algo que efetivamente existe (i.e., não só um “objeto do pensamento”). Existe um sentido de “referir” pelo qual posso me “referir” àquilo que não existe; não é nesse sentido que “referir” é usado aqui. Uma palavra mais antiga para o que chamo de “representação” ou “referência” é *denotação*.

Em segundo lugar, siga o costume dos lógicos modernos e uso “existir” como significando “existe no passado, presente ou futuro”. Assim, Winston Churchill “existe” e podemos nos “referir a” ou “representar” Winston Churchill, embora ele não esteja mais vivo.

Mas, para ter a intenção de que *qualquer coisa*, mesmo a linguagem privada (mesmo as palavras “Winston Churchill” ditas mentalmente, e não em voz alta), *represente* Churchill, devo ter sido capaz de, em primeiro lugar, *pensar sobre* Churchill. Se linhas traçadas na areia, ruídos etc. não podem, “em si mesmos”, representar nada, então como é que formas do pensamento podem, “em si mesmas”, representar o que quer que seja? Ou será que eles podem? Como o pensamento alcança e “compreende” o que lhe é externo?

No passado, alguns filósofos saltaram desse tipo de consideração para o que viram como uma prova de que a mente é *essencialmente não física por natureza*. O argumento é simples: o que dissemos sobre o traçado da formiga se aplica a qualquer objeto físico. Nenhum objeto físico pode, em si mesmo, referir-se a uma coisa ao invés de outra; no entanto, os *pensamentos na mente* obviamente são bem-sucedidos em se referir a uma coisa em vez de outra. Logo, os pensamentos (e, por conseguinte, a mente) são de uma natureza essencialmente distinta daquela dos objetos materiais. Os pensamentos têm a característica da *intencionalidade* — eles podem se referir a algo mais; nada físico tem “intencionalidade”, salvo quando a intencionalidade deriva de algum emprego daquela coisa física por uma mente. Ou assim é afirmado. Isso é muito apressado; o mero postular dos misteriosos poderes da mente não resolve nada. Mas o problema é muito real. Como são possíveis a intencionalidade e a referência?

Teorias mágicas da referência

Vimos que a “imagem” da formiga não tem conexão necessária com Winston Churchill. O mero fato de que a “imagem” guarda uma “semelhança” com Churchill não a torna uma imagem real ou tampouco uma representação de Churchill. A menos que a formiga seja inteligente (o que não é) e conheça Churchill (não é o caso), a curva que traçou não é uma imagem ou sequer uma representação de nada. Alguns povos primitivos acreditavam que certas representações (*nomes*, em particular) têm uma conexão necessária com seus portadores; que saber o “nome verdadeiro” de alguém ou algo dá poder sobre a pessoa ou a coisa. Esse poder advém da *conexão mágica* entre o nome e o portador do nome; uma vez que se compreende que um nome tem *apenas* uma conexão contextual, contingente e convencional com seu portador, é difícil entender por que o conhecimento do nome deveria ter qualquer significância mística.

O que é importante compreender é que aquilo que ocorre com as imagens físicas também acontece com as imagens mentais e as representações mentais em geral; a exemplo das representações físicas, as representações mentais não têm uma conexão necessária com aquilo que representam. A suposição contrária é remanescente do pensamento mágico.

Talvez o ponto seja mais fácil de compreender no caso das *imagens* mentais. (Talvez o primeiro filósofo a compreender a enorme importância desse ponto, embora não tenha sido o primeiro a elaborá-lo, foi Wittgenstein.) Suponha que haja um planeta em algum lugar no

qual os seres humanos tenham evoluído (ou sido deixados lá por astronautas alienígenas, ou o que você achar melhor). Suponha que esses humanos, de resto semelhantes a nós, nunca tenham visto *árvores*. Suponha que eles nunca tenham imaginado árvores (talvez a vida vegetal só exista no tal planeta na forma de bolores). Suponha que certo dia a fotografia de uma árvore caia no planeta, deixada acidentalmente por uma espaçonave que não faz nenhum outro tipo de contato com eles. Imagine-os intrigados com a imagem. Que diabo é isso? Toda sorte de especulação lhes ocorre: um edifício, uma canópia, até mesmo alguma espécie de animal. Mas suponha que eles jamais se aproximem da verdade.

Para nós, a fotografia é a representação de uma árvore. Para aqueles humanos, ela representa apenas um objeto estranho, de natureza e função desconhecidas. Suponha que, por ter visto a fotografia, um deles tenha uma imagem mental idêntica a uma das minhas imagens mentais de uma árvore. A imagem mental dele não é uma *representação de uma árvore*. Ela é apenas uma representação do objeto estranho (seja lá o que for) que a fotografia misteriosa representa.

Mesmo assim, alguém pode argumentar que a imagem mental é, *de fato*, uma representação de uma árvore simplesmente porque a fotografia que a ocasionou é, ela própria e para começo de conversa, uma representação de uma árvore. Há uma cadeia causal que vai das árvores reais até a imagem mental, por mais estranha que seja a imagem.

Mas é possível imaginar até mesmo essa cadeia causal como ausente. Suponha que a “fotografia da árvore” deixada pela espaçonave não seja realmente a imagem de uma árvore, mas o resultado acidental do derramamento de algumas tintas. Mesmo que se pareça exatamente com a imagem de uma árvore, ela o é tanto quanto a “caricatura” de Churchill feita pela formiga era uma imagem de Churchill. Podemos até imaginar que a espaçonave que deixou a “fotografia” veio de um planeta cujos habitantes não saibam nada a respeito de árvores. Então, os humanos ainda teriam imagens mentais qualitativamente idênticas à minha imagem de uma árvore, mas elas não seriam imagens que representariam uma árvore ou qualquer outra coisa.

A mesma coisa é verdadeira no que diz respeito às *palavras*. Um texto em uma folha de papel pode ser uma descrição perfeita das árvores, mas, caso tenha sido escrito por macacos que pressionaram teclas aleatórias de uma máquina de escrever por milhões de anos, as palavras não se referem a nada. Se houvesse uma pessoa que memorizasse essas palavras e as dissesse mentalmente sem compreendê-las, elas tampouco se refeririam a coisa alguma quando pensadas na mente.

Imagine que a pessoa que está dizendo mentalmente essas palavras foi hipnotizada. Suponha que as palavras estejam em japonês, e que foi dito à pessoa que ela compreende japonês. Suponha que, ao pensar nessas palavras, ela tenha uma “sensação de entendimento”. (No entanto, se alguém invadissem sua linha de raciocínio e perguntasse o que *significam* as palavras nas quais pensa, a pessoa descobriria que não consegue explicá-las.) Talvez a ilusão fosse tão perfeita que ela conseguiria até mesmo enganar um telepata japonês! Mas, se não

conseguisse utilizar as palavras nos contextos corretos, respondesse a questões sobre o que está “pensando” etc., então a pessoa não as teria compreendido.

Ao combinar essas histórias de ficção científica que estou contando, podemos inventar um caso no qual alguém pense em palavras que sejam, de fato, uma descrição de árvores em alguma língua *e* simultaneamente tenha imagens mentais apropriadas, mas *não* compreenda as palavras *nem* saiba o que é uma árvore. Podemos até imaginar que as imagens mentais foram ocasionadas por respingos de tinta (embora a pessoa foi hipnotizada para pensar que são imagens de algo apropriado ao seu pensamento — ainda que, se for perguntada, ela não será capaz de dizer do que se tratam). E podemos imaginar que a língua na qual a pessoa pensa seja uma que ela e o hipnotizador desconhecem por completo — talvez seja apenas coincidência que essas “frases sem sentido”, como o hipnotizador as supõe, sejam uma descrição de árvores em japonês. Em resumo, tudo o que passa pela mente da pessoa pode ser qualitativamente idêntico ao que passaria pela mente de um falante de japonês que estivesse *realmente* pensando sobre árvores — mas nada disso se referiria a árvores.

Claro que tudo isso é, na verdade, impossível, da mesma forma como seria impossível que macacos datilografassem por acaso uma cópia de *Hamlet*. Ou seja, as probabilidades contrárias são tão altas que isso realmente jamais aconteceria (nós achamos). Mas não é logicamente impossível, ou sequer fisicamente impossível. *Poderia* acontecer (é compatível com as leis da física, e talvez seja compatível com as reais condições do universo, caso haja muitos seres inteligentes em outros planetas). E, caso acontecesse, seria uma demonstração incrível de uma importante verdade conceitual: que mesmo um grande e complexo sistema de representações verbais e visuais não tem uma conexão *intrínseca*, interna e mágica com aquilo que representa — uma conexão independente de como ele foi engendrado e de quais são as disposições do falante ou pensante. E isso é verdade tanto para o sistema de representações (palavras e imagens, no caso do exemplo) fisicamente concretizado — as palavras são escritas ou faladas, e as imagens são imagens físicas — quanto para o sistema efetivado apenas na mente. Palavras pensadas e imagens mentais não representam *intrinsecamente* aquilo de que se tratam.

O caso dos cérebros em uma cuba

Eis aqui uma possibilidade de ficção científica discutida pelos filósofos: imagine que um ser humano (pode imaginar que seja você) foi submetido a uma cirurgia por um cientista perverso. O cérebro da pessoa (seu cérebro) foi removido do corpo e colocado em uma cuba cheia de nutrientes que o mantêm vivo. As terminações nervosas foram conectadas a um computador supercientífico que faz com que o cérebro da pessoa tenha a ilusão de que tudo está perfeitamente normal. Parecem haver pessoas, objetos, o céu etc.; mas, na realidade, tudo o que a pessoa (você) está vivenciando é resultado de impulsos eletrônicos que trafegam do computador até as terminações nervosas. O computador é tão astuto que, se a pessoa tenta

levantar a mão, um *feedback* fará com que ela “veja” e “sinta” a mão sendo erguida. Além disso, usando outros programas, o cientista perverso pode fazer a vítima “vivenciar” (ou alucinar) qualquer situação ou ambiente que ele queira. O cientista também pode obliterar a lembrança da cirurgia cerebral, de tal forma que à vítima pareça que ela sempre viveu naquele cenário. Pode até parecer a ela que está sentada, lendo sobre essa suposição divertida, mas totalmente absurda, de que há um cientista perverso que remove os cérebros dos corpos das pessoas e os coloca em uma cuba com nutrientes para mantê-los vivos. As terminações nervosas supostamente estão conectadas a um computador supercientífico que faz com que o cérebro da pessoa tenha a ilusão de...

Claro que, ao ser mencionada uma possibilidade desse tipo em uma aula sobre a Teoria do Conhecimento, o propósito é levantar de uma forma moderna o clássico problema do ceticismo com relação ao mundo exterior. (*Como você sabe que não está nessa situação desagradável?*) Mas essa situação desagradável também é um artifício útil para abordar questões acerca do relacionamento mente/mundo.

Em vez de apenas um cérebro em uma cuba, podemos imaginar que todos os seres humanos (talvez todos os seres sencientes) são cérebros em uma cuba (ou sistemas nervosos em uma cuba, caso alguns seres com aparatos mínimos também sejam tidos como “sencientes”). Claro que o cientista perverso teria de estar do lado de fora — onde mais estaria? Talvez não haja um cientista perverso, talvez (embora isso seja absurdo) apenas calhou de o universo consistir em um maquinário automatizado que supervisiona uma cuba cheia de cérebros e sistemas nervosos.

Dessa vez, vamos supor que o maquinário automatizado está programado para nos dar a todos uma alucinação *coletiva*, em vez de várias alucinações separadas e não relacionadas entre si. Assim, quando me parece estar falando contigo, você tem a impressão de ouvir as minhas palavras. Claro que não é o caso de as minhas palavras efetivamente chegarem aos seus ouvidos — pois você não tem ouvidos (reais), nem eu tenho boca e língua reais. Em vez disso, quando elaboro minhas palavras, o que acontece é que os impulsos eferentes viajam do meu cérebro para o computador, o qual faz com que eu “ouça” a minha própria voz proferindo aquelas palavras e “sinta” a minha língua se mexendo etc., e faz com que você “ouça” as minhas palavras e me “veja” falando etc. Nesse caso, de certo modo, nós estamos efetivamente nos comunicando. Não estou enganado quanto à sua existência real (apenas sobre a existência do seu corpo e do “mundo exterior”, à exceção dos cérebros). Sob um certo ponto de vista, sequer importa que “o mundo inteiro” seja uma alucinação coletiva; pois, no fim das contas, você de fato ouve as minhas palavras quando falo contigo, mesmo que o mecanismo não seja o que suponhamos que ele seja. (Claro que, se formos dois amantes fazendo amor, em vez de duas pessoas tendo apenas uma conversa, a sugestão de que somos apenas dois cérebros em uma cuba pode ser perturbadora.)

Agora, quero responder a uma questão que parecerá muito tola e óbvia (ao menos para algumas pessoas, incluindo certos filósofos bastante sofisticados), mas que nos levará de

forma incrivelmente rápida para verdadeiras profundezas filosóficas. Suponha que toda essa história fosse mesmo verdadeira. Nós poderíamos, caso fôssemos cérebros em uma cuba, *dizer* ou *pensar* que somos tal coisa?

Argumentarei que a resposta é: “Não, não poderíamos”. Com efeito, argumentarei que a suposição de que somos mesmo cérebros em uma cuba, embora não viole nenhuma lei física e seja perfeitamente consistente com tudo o que temos vivenciado, não pode ser verdadeira. *Ela não pode ser verdadeira* porque é, de certa maneira, autorrefutante.

O argumento que apresentarei é inusual, e levei vários anos para me convencer de que está mesmo certo. Mas ele é um argumento correto. O que o faz parecer tão estranho é o fato de estar ligado a algumas das questões mais profundas da filosofia. (Ele primeiro me ocorreu quando pensava sobre um teorema da lógica moderna, o “Teorema de Löwenheim-Skolem”, e de repente vi uma conexão entre esse teorema e alguns dos argumentos de Wittgenstein nas *Investigações Filosóficas*.)

Uma “suposição autorrefutante” é aquela cuja verdade implica a sua própria falsidade. Por exemplo, considere a tese de que *todos os enunciados gerais são falsos*. Ela é um enunciado geral. Portanto, se for verdadeira, deverá ser falsa. Logo, ela é falsa. Às vezes, uma tese é chamada de “autorrefutante” se a *suposição de que a tese é acolhida ou enunciada* implicar a sua falsidade. Por exemplo: “Eu não existo” é autorrefutante caso seja pensada por *mim* (por qualquer “*mim*”). Assim, um indivíduo pode estar certo de que existe caso pense a respeito disso (como Descartes argumentou).

O que demonstrarei é que a suposição de que somos cérebros em uma cuba tem justamente essa propriedade. Se podemos ponderar se é verdadeira ou falsa, então ela não é verdadeira (demonstrarei). Logo, ela não é verdadeira.

Antes de apresentar o argumento, pensemos sobre por que parece tão estranho que um argumento desses possa ser formulado (ao menos para os filósofos partidários de uma concepção da verdade como “cópia”). Admitimos ser compatível com as leis físicas a existência de um mundo no qual todos os seres sencientes são cérebros em uma cuba. Como dizem os filósofos, existe um “mundo possível” no qual todos os seres sencientes são cérebros em uma cuba. (Essa conversa sobre “mundos possíveis” soa como se houvesse um *lugar* onde qualquer suposição absurda fosse verdadeira, razão pela qual pode ser bastante falaciosa em filosofia.) Os humanos naquele mundo possível têm exatamente as mesmas experiências que *nós* temos. Eles pensam os mesmos pensamentos que nós pensamos (as mesmas palavras, imagens, formas-de-pensamento etc. passam pelas mentes deles, pelo menos). No entanto, afirmo que há um argumento que podemos apresentar e que demonstra que não somos cérebros em uma cuba. Como pode haver algo assim? E por que as pessoas no mundo possível que realmente são cérebros em uma cuba não podem apresentá-lo também?

A resposta será (basicamente) esta: embora as pessoas naquele mundo possível possam pensar e “dizer” quaisquer palavras que nós pensamos e dizemos, elas não podem (eu

afirmo) *referir-se* ao que podemos nos referir. Especificamente, elas não podem pensar ou dizer que são cérebros em uma cuba (mesmo ao pensar “*nós somos cérebros em uma cuba*”).

O teste de Turing

Suponha que alguém consiga inventar um computador que possa realmente ter uma conversa inteligente com uma pessoa (sobre tantos assuntos quanto queira falar uma pessoa inteligente). Como se poderia decidir que o computador é “consciente”?

O lógico britânico Alan Turing propôs o seguinte teste³: deixe alguém conversar com o computador e com uma pessoa que não conheça. Se ele não conseguir distinguir o computador do ser humano, então (pressuponha que o teste foi repetido um número suficiente de vezes com diferentes interlocutores) o computador é consciente. Em resumo, uma máquina computacional é consciente se passar no “teste de Turing”. (Claro que as conversas não devem ocorrer cara a cara, uma vez que o interlocutor não deve saber qual é a aparência visual de nenhum de seus dois parceiros de conversação. Tampouco a voz pode ser usada, pois uma voz mecânica simplesmente soaria distinta de uma voz humana. Imagine, antes, que as conversas ocorram por meio de uma máquina de escrever elétrica. O interlocutor datilografa suas afirmações, questões etc., e os dois parceiros — a máquina e a pessoa — respondem usando o teclado elétrico. Além disso, a máquina pode *mentir* — diante da pergunta “você é uma máquina?”, ela pode responder: “Não, sou assistente aqui no laboratório”).

A ideia de que isso seja realmente um teste de consciência tem sido criticada por vários autores (os quais, em princípio, não são de modo algum hostis à ideia de que uma máquina possa ser consciente). Mas esse não é o nosso tópico no momento. Desejo usar a ideia geral do teste de Turing, a ideia geral de um *teste dialógico de competência*, para um propósito diferente, o propósito de explorar a noção de *referência*.

Imagine uma situação na qual o problema não é determinar se o parceiro é realmente uma pessoa ou uma máquina, mas, antes, determinar se o parceiro usa as palavras para referir como nós usamos. O teste óbvio é, de novo, levar adiante uma conversa e, caso não surjam problemas, se o parceiro “passar” no sentido de ser indistinguível de alguém previamente certificado como um falante da mesma língua, referindo-se aos tipos usuais de objetos etc., concluir que o parceiro se refere aos objetos da mesma forma como nós nos referimos. Quando o propósito do teste de Turing for conforme descrevemos, isto é, determinar a existência de referência (compartilhada), devo designá-lo como o *Teste de Turing para a Referência*. E, assim como os filósofos têm debatido a questão sobre se o teste original de Turing é *definitivo* para a consciência, i.e., a questão sobre se uma máquina que “passa” no teste não apenas uma vez, mas regularmente, é *necessariamente* consciente, então, do mesmo modo, desejo discutir

³ A. M. Turing, “Computing Machinery and Intelligence”, *Mind* (1950), reimpresso em A. R. Anderson (ed.), *Minds and Machines*.

a questão sobre se o Teste de Turing para a Referência sugerido há pouco é um teste definitivo para a referência compartilhada.

A resposta será “não”. O Teste de Turing para a Referência não é definitivo. Por certo, na prática, é um excelente teste; mas não é logicamente impossível (embora com certeza seja altamente improvável) que alguém consiga passar no Teste de Turing para a Referência sem referir-se a nada. Segue-se daí, como veremos, que podemos estender a nossa observação de que as palavras (e textos e discursos inteiros) não têm uma conexão necessária com seus referentes. Mesmo que consideremos não as palavras por si mesmas, mas regras decidindo quais palavras podem apropriadamente ser elaboradas em certos contextos — mesmo que consideremos, no jargão computacional, *programas para o uso de palavras* — a menos que tais programas *se refiram a algo extralinguístico*, essas palavras ainda não possuem nenhuma referência determinada. Eis um passo crucial no processo para se chegar à conclusão de que os Habitantes do Mundo do Cérebro-em-uma-Cuba não podem se referir a nada externo (e, por conseguinte, não podem dizer que são Habitantes do Mundo do Cérebro-em-uma-Cuba).

Suponha, por exemplo, que estou na situação de Turing (jogando o “Jogo da Imitação”, na terminologia criada por ele) e o meu parceiro é, de fato, uma máquina. Suponha que essa máquina seja capaz de vencer o jogo (“passar” no teste). Imagine que a máquina foi programada para produzir belas respostas em nossa língua para afirmações, questões, observações etc., mas não tem órgãos dos sentidos (além da conexão com a máquina de escrever elétrica) nem órgãos motores (além da máquina de escrever elétrica). (Até onde entendo, Turing não pressupõe que a posse de órgãos dos sentidos ou motores seja necessária para se ter consciência ou inteligência.) Suponha que não só falem à máquina olhos e ouvidos eletrônicos etc., mas que também não haja recursos no programa da máquina, o programa para jogar o Jogo da Imitação, para incorporar *inputs* desses órgãos dos sentidos ou controlar um corpo. O que deveríamos dizer sobre uma máquina como essa?

Para mim, parece evidente que não podemos e não devemos atribuir referência a tal dispositivo. É verdade que a máquina pode discursar lindamente sobre, digamos, a paisagem da Nova Inglaterra. Mas ela não conseguiria reconhecer uma macieira ou uma maçã, uma montanha ou uma vaca, um campo ou um campanário, caso estivesse diante de uma dessas coisas.

O que temos é um dispositivo para produzir sentenças em respostas a sentenças. Mas nenhuma dessas sentenças está, de modo algum, conectada ao mundo real. *Se alguém acoplasse duas dessas máquinas e as deixasse jogar o Jogo da Imitação uma com a outra, elas ficariam “enganando-se” mutuamente pela eternidade afora, mesmo que resto do mundo desaparecesse!* Há tanta razão para considerar a conversa da máquina sobre maçãs como referindo-se às maçãs do mundo real quanto para considerar o “desenho” da formiga como referindo-se a Winston Churchill.

Aqui, o que produz uma ilusão de referência, significado, inteligência etc. é o fato de que há uma convenção de representação que *nós* temos e em cujos termos o discurso da

máquina refere-se a maçãs, campanários, Nova Inglaterra etc. De modo similar, e pela mesma razão, há a *ilusão* de que a formiga caricaturou Churchill. Mas somos capazes de perceber, manusear, lidar com maçãs e campos. Nossa fala sobre maçãs e campos está intimamente conectada às nossas transações *não verbais* com maçãs e campos. Existem “regras de entrada da linguagem” que nos levam de experiências com maçãs para elocuições como “Eu vejo uma maçã”, e “regras de saída da linguagem” que nos levam de decisões expressas em forma linguística (“Estou indo comprar algumas maçãs”) para ações distintas da fala. Na falta das regras de entrada ou das regras de saída da linguagem, não há razão para considerar a conversa da máquina (ou entre as duas máquinas, no caso que imaginamos onde duas máquinas rodam o Jogo da Imitação) como algo mais do que um jogo sintático. Um jogo sintático que *se assemelha* ao discurso inteligente, está certo; mas apenas como (e não mais do que) o traçado da formiga se assemelha a uma caricatura sarcástica.

No caso da formiga, poderíamos argumentar que ela teria traçado a mesma linha mesmo se Winston Churchill jamais tivesse existido. No caso da máquina, não é possível argumentar de forma mais ou menos análoga; se maçãs, árvores, campanários e campos não existissem, presume-se que os programadores não teriam criado o mesmo programa. Embora a máquina não *perceba* maçãs, campos ou campanários, seus criadores-projetistas perceberam. Existe *alguma* conexão causal entre a máquina e as maçãs etc. do mundo real por meio da experiência perceptiva e do conhecimento dos criadores-projetistas. Mas uma conexão tão fraca dificilmente seria suficiente para a referência. Não só é logicamente possível, embora fantasticamente improvável, que a mesma máquina *pudesse* existir mesmo que maçãs, campos e campanários não existissem; mais importante, a máquina é totalmente insensível à existência *continuada* de maçãs, campos, campanários etc. Mesmo que todas essas coisas *cessassem* de existir, a máquina seguiria discursando alegremente da mesma forma. Eis porque não se pode considerar que a máquina se refira ao que quer que seja.

O ponto relevante para a nossa discussão é que não há nada no teste de Turing que descarte uma máquina programada para não fazer nada *além* de jogar o Jogo da Imitação, e que uma máquina que não possa fazer nada *além* de jogar o Jogo da Imitação *claramente* faz tantas referências quanto um gravador.

Cérebros em uma cuba (de novo)

Vamos comparar os hipotéticos “cérebros em uma cuba” com as máquinas descritas há pouco. É óbvio que existem diferenças importantes. Os cérebros em uma cuba não têm órgãos dos sentidos, mas têm *dispositivos* para os órgãos dos sentidos, ou seja, há terminações nervosas aferentes, há *inputs* dessas terminações nervosas aferentes, e esses *inputs* figuram no “programa” dos cérebros em uma cuba da mesma forma como estão no programa dos nossos cérebros. Os cérebros em uma cuba são *cérebros*; mais do que isso, eles são cérebros *funcionais*, e funcionam segundo as mesmas regras que regem os cérebros no mundo real. Por

essas razões, pareceria absurdo negar que eles têm consciência ou inteligência. Mas o fato de que são conscientes e inteligentes não significa que suas palavras se refiram àquilo que as nossas palavras se referem. A questão em que estamos interessados é esta: suas verbalizações contendo, digamos, a palavra “árvore” se referem, de fato, às *árvores*? De forma mais geral: eles podem se referir a quaisquer objetos *externos*? (Em oposição, por exemplo, aos objetos produzidos pelo maquinário automatizado.)

Para ajustar nossas ideias, vamos especificar que o maquinário automatizado supostamente passou a existir graças a algum tipo de acaso ou coincidência cósmica (ou talvez sempre tenha existido). Nesse mundo hipotético, supõe-se que o próprio maquinário automatizado não tem nenhum criador-projetista. Com efeito, conforme dissemos no começo deste capítulo, podemos imaginar que todos os seres sencientes (por mínima que seja a sua sciência) estão dentro da cuba.

Essa suposição não ajuda. Pois não há nenhuma conexão entre a *palavra* “árvore” tal como usada por esses cérebros e as árvores reais. Eles ainda utilizariam a palavra “árvore” tal como fazem, pensariam exatamente o que pensam, teriam a mesmas imagens que têm, mesmo que não existissem árvores na realidade. Suas imagens, palavras etc. são qualitativamente idênticas às imagens, palavras etc. que representam as árvores no *nosso* mundo; mas já vimos (a formiga de novo!) que similaridade qualitativa com algo que representa um objeto (Winston Churchill ou uma árvore) não torna algo uma representação por si só. Em resumo, os cérebros em uma cuba não estão pensando sobre árvores reais quando pensam “há uma árvore na minha frente”, pois não há nada em virtude do que seu pensamento “árvore” represente árvores reais.

Caso isso soe precipitado, reflita sobre o seguinte: vimos que as palavras não se referem necessariamente a árvores mesmo que estejam arranjadas em uma sequência idêntica a um discurso que (se ocorresse em uma de nossas mentes) inquestionavelmente *seria sobre árvores* no mundo real. Tampouco o “programa”, no sentido das regras, práticas e disposições dos cérebros ao comportamento verbal, necessariamente se refere a árvores ou faz referência a árvores por meio das conexões que estabelece entre palavras e palavras, ou entre pistas *lingüísticas* e respostas *lingüísticas*. Se esses cérebros pensam sobre, referem-se a, representam árvores (árvores reais, fora da cuba), então deve ser por causa da forma como o “programa” conecta o sistema da linguagem a *inputs* e *outputs não verbais*. Há, com efeito, tais *inputs* e *outputs não verbais* no mundo dos Cérebros-em-uma-Cuba (aquelas terminações nervosas eferentes e aferentes de novo!), mas também vimos que os “dados dos sentidos” produzidos pelo maquinário automatizado não representam árvores (ou qualquer coisa externa) mesmo quando se parecem exatamente com as nossas imagens de árvores. Tal como um respingo de tinta pode se assemelhar a uma imagem de árvore mesmo não *sendo* uma imagem de árvore, vimos que um “dado dos sentidos” pode ser qualitativamente idêntico à “imagem de uma árvore” mesmo não sendo a imagem de uma árvore. Como é possível o fato de que, no caso dos cérebros em uma cuba, a linguagem seja conectada pelo programa com

inputs sensoriais que não representam, de forma intrínseca ou extrínseca, árvores (ou qualquer coisa externa) e faça com que todo o sistema de representações, a linguagem-em-uso, *refira-se* a ou represente árvores ou qualquer coisa externa?

A resposta é que isso não é possível. Todo o sistema de dados dos sentidos, os sinais motores para as terminações eferentes, e o pensamento verbal ou conceitualmente mediado e conectado pelas “regras de entrada da linguagem” aos dados dos sentidos (ou coisa que o valha) como *inputs*, e pelas “regras de saída da linguagem” aos sinais motores como *outputs*, têm tanta conexão com as *árvores* quanto o traçado da formiga tem para com Winston Churchill. Uma vez que entendemos que a *similaridade qualitativa* (correspondendo, se você quiser, à identidade qualitativa) entre os pensamentos dos cérebros em uma cuba e os pensamentos de alguém no mundo real não implica, de modo algum, a mesmidade da referência, não é difícil ver que não existe base alguma para considerar que o cérebro em uma cuba se refira a coisas externas.

As premissas do argumento

Agora já apresentei o argumento prometido para demonstrar que os cérebros em uma cuba não podem pensar ou dizer que são cérebros em uma cuba. Resta apenas explicitá-lo e examinar sua estrutura.

Pelo que eu disse há pouco, quando o cérebro em uma cuba (no mundo onde todo ser senciente é e sempre foi um cérebro em uma cuba) pensa “há uma árvore na minha frente”, seu pensamento não se refere a árvores reais. Em algumas teorias que discutiremos, ele pode se referir a árvores na imagem, ou aos impulsos eletrônicos que produzem as experiências com árvores, ou às características do programa que são responsáveis por esses impulsos eletrônicos. Essas teorias não são descartadas pelo que acabei de dizer, pois há uma conexão causal próxima entre o uso da palavra “árvore” em português-cuba e a presença de árvores na imagem, a presença de impulsos eletrônicos de um certo tipo, e a presença de certas características no programa da máquina. Segundo essas teorias, o cérebro está *certo*, não *errado*, ao pensar “há uma árvore na minha frente”. Dado a que “árvores” e “na frente de” se referem em português-cuba, e presumindo que uma dessas teorias esteja correta, então as condições de verdade para “há uma árvore na minha frente” quando ela ocorre em português-cuba são simplesmente de que uma árvore na imagem esteja “na frente” do “eu” em questão — na imagem —, ou talvez que o tipo de impulso eletrônico que normalmente produz essa experiência esteja vindo do maquinário automatizado, ou talvez que o recurso do maquinário que deve produzir a experiência da “árvore na minha frente” esteja operando. E essas condições de verdade certamente são cumpridas.

Segundo o mesmo argumento, “cuba” refere-se a cubas na imagem em português-cuba ou a algo relacionado (impulsos eletrônicos ou recursos do programa), mas certamente não a cubas reais, uma vez que o uso de “cuba” em português-cuba não tem conexão causal

com cubas reais (à exceção da conexão na qual os cérebros em uma cuba não seriam capazes de usar a palavra “cuba” não fosse pela presença de uma cuba específica — a cuba na qual estão; mas essa conexão é obtida entre o uso de *toda* palavra em português-cuba e aquela cuba específica; não é uma conexão *especial* entre o uso da palavra *específica* “cuba” e as cubas). De modo similar, “fluido nutriente” refere-se a um líquido na imagem em português-cuba ou a algo relacionado (impulsos eletrônicos ou recursos do programa). Disso se segue que, caso seu “mundo possível” seja mesmo o mundo real, e nós sejamos de fato cérebros em uma cuba, então o que agora queremos dizer com “somos cérebros em uma cuba” é que *nós somos cérebros em uma cuba na imagem* ou algo desse tipo (se é queremos mesmo dizer alguma coisa). Mas parte da hipótese de que somos cérebros em uma cuba é que não somos cérebros em uma cuba na imagem (i.e., não estamos “alucinando” que somos cérebros em uma cuba). Assim, caso sejamos cérebros em uma cuba, então a sentença “somos cérebros em uma cuba” diz algo falso (se é que diz alguma coisa). Em resumo, se nós somos cérebros em uma cuba, então a sentença “somos cérebros em uma cuba” é falsa. Portanto, ela é (necessariamente) falsa.

A suposição de que tal possibilidade faz sentido surge de uma combinação de dois erros: 1) levar a *possibilidade física* muito a sério; e 2) operar inconscientemente com uma teoria mágica da referência, uma teoria segundo a qual certas representações mentais necessariamente se referem a certas coisas e tipos de coisas externas.

Há um “mundo fisicamente possível” no qual somos cérebros em uma cuba — o que isso significa, exceto que existe uma *descrição* de tal estado de coisas que é compatível com as leis da física? Assim como há uma tendência em nossa cultura (e é assim desde o século XVII) de tomar a *física* como a nossa metafísica, isto é, de encarar as ciências exatas como as descrições há muito buscadas do “verdadeiro e definitivo mobiliário do universo”, também existe, como uma consequência imediata, uma tendência a se considerar a “possibilidade física” como a pedra angular do que poderia ser efetiva e realmente o caso. Segundo essa visão, a verdade é a verdade física; a possibilidade, possibilidade física; e a necessidade, necessidade física. Mas vimos há pouco, embora apenas no caso de um exemplo bastante artificial, que essa visão é errada. A existência de um “mundo fisicamente possível” no qual somos cérebros em uma cuba (e sempre fomos e seremos) não significa que possamos realmente, de fato ou possivelmente *sermos* cérebros em uma cuba. O que exclui essa possibilidade não é a física, mas a *filosofia*.

Alguns filósofos, ansiosos para simultaneamente asseverar e minimizar as demandas de sua profissão (o típico estado mental da filosofia anglo-americana no século XX), diriam: “Certo. Você demonstrou que algumas coisas que parecem possibilidades físicas são, na verdade, impossibilidades *conceituais*. O que há de tão surpreendente nisso?”.

Bem, não há dúvida de que o meu argumento possa ser descrito como “conceitual”. Mas descrever a atividade filosófica como a busca por verdades “conceituais” faz tudo isso soar

como uma *investigação sobre o significado das palavras*. E não é, de forma alguma, nisso que estamos engajados.

O que estamos fazendo é examinar as *precondições* para se *pensar em, representar, referir-se* a algo etc. Temos investigado essas precondições *não* pela investigação do significado dessas palavras e frases (como faria um linguista, por exemplo), mas por meio do *raciocínio a priori*. Não no velho sentido “absoluto” (uma vez que não afirmamos que as teorias mágicas da referência estejam erradas *a priori*), mas no sentido de inquirir sobre o que é *racionalmente* possível se *pressupusermos* certas premissas gerais ou fizermos determinadas pressuposições teóricas muito abrangentes. Tal procedimento não é “empírico” nem exatamente “a priori”, mas tem elementos de ambas as formas de investigação. A despeito da falibilidade do meu procedimento e de sua dependência de certos pressupostos que podem ser descritos como “empíricos” (*e.g.*, o pressuposto de que a mente não tem acesso a coisas ou propriedades externas além daquelas fornecidas pelos sentidos), ele tem uma relação próxima com o que Kant chamou de investigação “transcendental”, pois, repito, é uma investigação das *precondições* da referência e, por conseguinte, do pensamento — precondições embutidas na natureza das nossas próprias mentes, embora não (como Kant esperava) independentes por completo dos pressupostos empíricos.

Uma das premissas do argumento é óbvia: que as teorias mágicas da referência estão erradas, e erradas no que diz respeito às representações mentais e não só às físicas. A outra premissa é que alguém não pode se referir a certos tipos de coisas, *e.g.*, *árvores*, caso não tenha nenhuma interação causal com elas⁴ ou com coisas em cujos termos elas possam ser descritas. Mas por que deveríamos aceitar essas premissas? Uma vez que elas constituem o amplo arcabouço dentro do qual estou argumentando, chegou o momento de examiná-las mais de perto.

As razões para negar conexões necessárias entre representações e seus referentes

Mencionei antes que alguns filósofos (dos quais o mais célebre é Brentano) atribuíram um poder à mente, a “intencionalidade”, o qual é precisamente aquilo que possibilita a ela *referir*. É evidente que rejeitei isso como uma solução. Mas o que me dá esse direito? Fui, talvez, muito precipitado?

Esses filósofos não afirmam, de modo algum, que podemos pensar sobre as coisas ou propriedades externas sem usar representações. E o argumento que apresentei antes, comparando dados sensíveis visuais à “imagem” da formiga (a argumentação por meio da

⁴ Se os cérebros em uma cuba terão uma conexão causal com, digamos, as *árvores no futuro*, talvez, então, eles possam se referir *agora* às árvores por meio da descrição “as coisas a que me referirei como ‘árvores’ em tal e tal tempo futuro”. Mas imaginamos um caso no qual os cérebros em uma cuba *nunca* sairão da cuba e, portanto, *jamaiz* terão uma conexão causal com árvores etc.

história de ficção científica sobre a “imagem” de uma árvore que surgiu de um respingo de tinta e deu origem a dados dos sentidos qualitativamente similares às nossas “imagens visuais de árvores”, mas desacompanhados de qualquer *conceito* de uma árvore), seria aceito como demonstrando que *imagens* não referem necessariamente. Se existem representações mentais que necessariamente se referem (a coisas externas), elas devem ser da natureza dos *conceitos*, não da natureza das *imagens*. Mas o que são *conceitos*?

Quando introspectamos, não percebemos “conceitos” fluindo como tais pela nossa mente. Ao interromper o fluxo de pensamento quando ou onde quisermos, o que capturamos são palavras, imagens, sensações, sentimentos. Quando expresso meus pensamentos em voz alta, não penso neles duas vezes. Ouço as minhas palavras tal como você. Para ser franco, a sensação é diferente quando profiro palavras nas quais acredito e quando profiro palavras nas quais não acredito (mas, às vezes, quando estou nervoso ou perante uma audiência hostil, tenho a sensação de mentir quando sei que estou dizendo a verdade); e a sensação é diferente quando profiro palavras que entendo e quando profiro palavras que não entendo. Mas posso, sem dificuldade, imaginar alguém pensando nessas mesmas palavras (no sentido de dizê-las mentalmente) e tendo a exata sensação de entendimento, asseveração etc. que tenho, e percebendo um minuto depois (ou sendo acordada por um hipnotizador) que não entendeu nada do que acabou de passar pela sua mente, que sequer compreendeu a língua em que estão essas palavras. Não afirmo que isso seja muito provável; apenas quero dizer que não há nada de inimaginável aqui. E o que isso demonstra não é que *conceitos* são palavras (ou imagens, sensações etc.), mas que atribuir um “conceito” ou “pensamento” a uma pessoa é bem diferente de atribuir a ela qualquer “apresentação” mental, qualquer entidade ou evento introspectível. *Conceitos* não são apresentações mentais que se referem intrinsecamente a objetos externos pela razão bastante decisiva de que eles não são, de forma alguma, apresentações mentais. *Conceitos* são signos usados de uma certa maneira; os signos podem ser públicos ou privados, entidades mentais ou entidades físicas, mas, mesmo quando os signos são “mentais” e “privados”, o signo em si, separado de seu uso, não é o conceito. E próprios os signos não referem intrinsecamente.

Podemos entender isso fazendo um experimento mental muito simples. Suponha que você seja como eu e não consiga distinguir um olmo de uma faia. Ainda dizemos que a referência de “olmo” em meu discurso é a mesma que a referência de “olmo” no discurso de qualquer pessoa, a saber, os olmos, e que o conjunto de todas as faias é a extensão de “faia” (i.e., o conjunto de coisas das quais a palavra “faia” é verdadeiramente predicada) tanto no seu discurso quanto no meu. É mesmo crível que a diferença entre aquilo a que “olmo” se refere e aquilo a que “faia” se refere seja ocasionada por uma diferença em nossos *conceitos*? Meu conceito de um olmo é exatamente o mesmo que meu conceito de uma faia (enrubesço ao confessá-lo). (Isso, aliás, mostra que a determinação da referência é social e não individual; eu e você diferimos dos especialistas que *conseguem* diferenciar olmos de faias.) Se alguém tentar heroicamente sustentar que a diferença entre a referência de “olmo” e a referência de

“faia” no *meu* discurso é explicada por uma diferença no meu estado psicológico, então deixe-o imaginar uma Terra Gêmea onde as palavras são trocadas. A Terra Gêmea é bastante parecida com a Terra; com efeito, à exceção do fato de que “olmo” e “faia” são trocadas entre si, o leitor pode imaginar que a Terra Gêmea é exatamente como a Terra. Suponha que eu tenha um *Doppelgänger* na Terra Gêmea que seja, molécula por molécula, idêntico a mim (no sentido de que duas gravatas possam ser “idênticas”). Se você for um dualista, então suponha que meu *Doppelgänger* pense os mesmos pensamentos verbalizados que eu, tenha os mesmos dados dos sentidos, as mesmas disposições etc. É absurdo pensar que seu estado psicológico seja minimamente distinto do meu; e, no entanto, a palavra “olmo” representa *faias* para ele, e a minha palavra “olmo” representa olmos. (De modo similar, se a “água” na Terra Gêmea for um líquido diferente — digamos, XYZ e não H₂O —, então “água” representa um líquido diferente quando usado na Terra Gêmea e quando usado na Terra etc.) Contrariando uma doutrina que nos acompanha desde o século XVII, *os significados simplesmente não estão na cabeça*.

Vimos que possuir um conceito não é uma questão de possuir imagens (digamos, de árvores — ou até mesmo imagens, “visuais” ou “acústicas”, de sentenças ou discursos inteiros, a propósito), uma vez que alguém poderia possuir qualquer sistema de imagens que você quisesse, mas sem a *habilidade* de usar as sentenças de formas situacionalmente apropriadas (considerando tanto os fatores linguísticos — conforme dito antes — quanto os fatores não linguísticos como determinantes para a “adequação situacional”). Um homem pode ter todas as imagens que você quiser e ainda ficar completamente perdido quando alguém disser a ele “aponte para uma árvore”, mesmo que haja muitas árvores ao redor. Ele pode até mesmo ter a imagem do que tem de fazer e, ainda assim, não saber o que tem de fazer. Pois a imagem, caso não esteja acompanhada pela habilidade de agir de uma certa maneira, é apenas uma *imagem*, e agir em conformidade com uma imagem é, em si, uma habilidade que uma pessoa pode ou não possuir. (O homem pode se imaginar apontando para uma árvore, mas apenas com o propósito de contemplar algo logicamente possível; ele mesmo apontando para uma árvore após alguém ter produzido a — para ele sem significado — sequência de sons “por favor, aponte para uma árvore”.) Ele ainda não saberia que tinha que apontar para uma árvore, e ainda não *entenderia* “apontar para uma árvore”.

Examinei a habilidade de se usar certas sentenças como o critério para a posse de um conceito completo, mas isso poderia ser facilmente liberalizado. Poderíamos permitir um simbolismo que consistisse em elementos que não são palavras em uma linguagem natural, por exemplo, e poderíamos permitir fenômenos mentais tais como imagens e outros tipos de eventos internos. O que é essencial é que eles tenham a mesma complexidade e habilidade de serem combinados uns com os outros etc. como sentenças em uma linguagem natural. Pois, embora uma apresentação específica — digamos, um lampejo azul — possa servir a um determinado matemático como a expressão interior de uma prova completa do Teorema dos Números Primos, ainda não haveria a tentação de dizer disso (e seria falso dizer isso) se o

matemático não conseguisse extrair de seu “lampejo azul” passos separados e conexões lógicas. Mas, não importa que espécies de fenômenos internos nós aceitemos como *expressões* possíveis do pensamento, argumentos bastante similares ao precedente demonstrarão que não são os fenômenos em si que constituem o entendimento, mas, antes, a habilidade do pensante de *empregar* esses fenômenos, de obter o fenômeno certo nas circunstâncias certas.

O que foi exposto acima é uma versão muito abreviada do argumento de Wittgenstein nas *Investigações Filosóficas*. Caso esteja correto, então a tentativa de entender o pensamento pelo que é chamado de investigação “fenomenológica” é algo fundamentalmente equivocado; pois o que os fenomenólogos não conseguem perceber é que estão descrevendo a *expressão* interior do pensamento, mas o *entendimento* dessa expressão — o entendimento que uma pessoa tem de seus próprios pensamentos — não é uma *ocorrência*, mas uma *habilidade*. Nosso exemplo de um homem que finge pensar em japonês (e engana um telepata japonês) já mostrou a futilidade de uma abordagem fenomenológica do problema do *entendimento*. Pois, mesmo que haja alguma qualidade introspectível presente quando, e somente quando, alguém *realmente* entende (isso, na verdade, parece falso sobre a introspecção), essa qualidade permanece apenas *correlacionada* com o entendimento, e ainda é possível que o homem que engana o telepata também possua tal qualidade e, *mesmo assim*, não compreenda uma palavra japonesa sequer.

Por outro lado, considere a hipótese perfeitamente possível de um homem que não tenha nenhum “monólogo interior”. Ele fala português fluentemente e, se perguntado acerca de um dado tema, dará sua opinião de forma minuciosa. Mas ele nunca pensa (em palavras, imagens etc.) quando não está falando em voz alta; e nada tampouco “passa por sua cabeça”, exceto que (é claro) ele ouve a própria voz ao falar e tem as impressões sensoriais habituais dos arredores, além de uma “sensação de entendimento” geral. (Talvez ele tenha o hábito de falar sozinho.) Quando datilografa uma carta ou vai a uma loja etc., ele não tem um “fluxo de pensamento” interior; mas suas ações são inteligentes e dotadas de propósito, e, caso alguém se aproxime e pergunte “o que você está fazendo?”, ele dará respostas absolutamente coerentes.

Esse homem parece perfeitamente imaginável. Ninguém hesitaria em dizer que ele era lúcido, desgostava de rock n’ roll (se ele expressasse com frequência uma forte aversão ao rock n’ roll) etc. só porque não tinha pensamentos conscientes, exceto ao falar em voz alta.

O que se segue de tudo isso é que (a) nenhum conjunto de eventos mentais — imagens ou acontecimentos e qualidades mentais mais “abstratos” — *constituem* o entendimento; e (b) nenhum conjunto de eventos mentais é *necessário* para o entendimento. Especificamente, os *conceitos não podem ser idênticos a objetos mentais de nenhuma espécie*. Pois, supondo que por objeto mental queiramos dizer algo introspectível, nós acabamos de ver que, seja o que isso for, ele pode estar ausente em um homem que entende o mundo de forma apropriada (e, portanto, possui o conceito completo) e presente em um homem que não tem o conceito de modo algum.

Voltando agora à nossa crítica das teorias mágicas da referência (um tópico que também interessava a Wittgenstein), vemos que, por um lado, esses “objetos mentais” que podemos detectar de forma introspectiva — palavras, imagens, sensações etc. — não referem intrinsecamente, a exemplo da imagem da formiga (e pelas mesmas razões), enquanto que as tentativas de postular objetos mentais especiais, “conceitos”, os quais *têm* uma conexão necessária com seus referentes e apenas os fenomenólogos treinados conseguem detectar, cometem um disparate *lógico*; pois os conceitos são (ao menos em parte) *habilidades*, e não ocorrências. A doutrina segundo a qual existem apresentações mentais que necessariamente se referem a coisas externas não é apenas má ciência natural; ela também é má fenomenologia e confusão conceitual.