

Cap. I

Funcionalismo vs. substancialismo

Uma máquina pode ser consciente?

1. Definição de consciência

O que é a consciência? Em que momento surgiu na evolução biológica? Como é possível a subjetividade em um mundo físico? Antes de prosseguir, é bom tentarmos definir o conceito geral de “consciência”, seguindo alguns dicionários de filosofia.¹

“Consciência” é a intuição, mais ou menos clara, que o sujeito tem dos seus estados e dos seus atos (LALANDE, p. 195). Possibilidade que cada um tem de dar atenção aos seus próprios modos de ser e às suas próprias ações, de estar ciente² dos próprios estados, percepções, ideias, sentimentos, volições etc. (ABBAGNANO, p. 217). A mente consciente, em oposição à mente inconsciente ou subconsciente (RUNES, p. 64).

Consideremos, porém, uma citação clássica do filósofo escocês William Hamilton³:

A consciência não pode ser definida: nós mesmos podemos estar totalmente cientes do que seja a consciência, mas não conseguimos sem confusão transmitir para os outros uma definição do que nós mesmos apreendemos com clareza. A razão é simples: a consciência está na raiz de todo conhecimento.

Cientistas e engenheiros muitas vezes exigem que definamos claramente o conceito de “consciência”, por meio de palavras, antes de aceitarem prosseguir na discussão. Mas as palavras foram desenvolvidas para a comunicação intersubjetiva, e o fenômeno que queremos tratar é subjetivo, privativo a cada indivíduo. O sentido desse conceito, “consciência fenomênica”, é melhor fixado por ostensão (i.e., apontando): é aquilo que estamos vivenciando agora. Thomas Nagel coloca da seguinte maneira: “what is it like to be an X” (como é ser um X).⁴ A partir desta definição ostensiva geral, mais para frente buscaremos distinguir entre diferentes tipos de consciência.

Para explorar esta situação, consideremos um primeiro experimento mental.

¹ LALANDE, A. (1999), *Vocabulário técnico e crítico da filosofia*, 3ª ed., trad. F.S. Correia, M.E.V. Aguiar, J.E. Torres & M.G. Souza, Martins Fontes, São Paulo. ABBAGNANO, N.(2007), *Dicionário de filosofia*, 5ª ed. revista e ampliada, trad. A. Bosi & I.C. Benedetti, Martins Fontes, São Paulo. RUNES, D.D. (1942), *The dictionary of philosophy*, 4ª ed., Philosophical Library, New York.

² O ato de “estar ciente de” é o ato de “ciência”, mas este termo é evitado em português, pois designa também a atividade científica. Podemos denotar esta por CIÊNCIA₁, e o ato de estar ciente por CIÊNCIA₂. O termo em inglês para ciência₂ é *awareness*, e em italiano *consapevolezza*. Alzira Allegro traduz por “consciência [A]”, em HILL, C.S. (2010), *Consciência*, Ed. Unesp, São Paulo. Essas e muitas outras informações sobre os termos usados em filosofia da mente estão apresentados nos *Arquivos lexicográficos* desta disciplina.

³ HAMILTON, W. (1877), *Lectures on metaphysics and logic*, 6ª ed., vol. I. Blackwood, Edinburgh, p. 191, (disponível online). O curso de metafísica (vol. I) foi redigido originalmente em 1836-37, na Universidade de Edimburgo.

⁴ NAGEL, T. (1974), “What is it like to be a bat?” *Philosophical Review* 83: 435-50. Em português: “Como é ser um morcego?”, trad. P. Abrantes & J. Orione, *Cadernos de História e Filosofia da Ciência* 15, 245-62, 2005.

2. O teste de Turing robótico

Considere uma situação como a do filme *Metrópolis*, de Fritz Lang (1927), em que um robô metálico é construído pelo cientista malvado e transformado na ginoide (androide feminina) Maria, interpretada pela atriz Brigitte Helm, indistinguível de um ser humano. Vamos supor que, em seu interior, Maria é feita de válvulas (ou chips de circuito integrado), fios e motores metálicos, mas julgando por seu comportamento, suas expressões faciais e sua fala, todos a consideram um ser humano normal. Nesse sentido, pode-se dizer que ela passou no “teste de Turing robótico”.

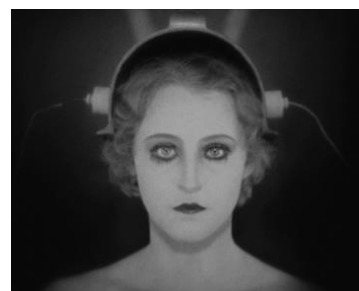


Fig. I.1. A ginoide Maria.

Alan Turing foi um genial matemático e computeiro, que em 1950 escreveu um artigo discutindo se máquinas poderiam “pensar”. Ao invés de tentar dar uma definição lexicográfica de “pensamento”, propôs um jogo ou teste para estabelecer se uma máquina pensa. Simplificando um pouco, o “jogo da imitação” envolveria um ser humano que faria perguntas ou dialogaria com a máquina ou com um humano escondido atrás de uma parede, e teria que fazer a identificação correta.⁵

Acredito que, dentro de cerca de cinquenta anos [no ano 2000], será possível programar computadores, com uma capacidade de memória de cerca de 10^9 [dígitos binários; em torno de 10 MB], para fazê-los jogar o jogo da imitação tão bem que um interrogador médio não terá mais de 70% de probabilidade de chegar à identificação correta, após 5 minutos de interrogatório. (TURING, 1973, p. 61; 1950, p. 442)

Ou seja, se uma máquina conseguir enganar um ser humano, de maneira que este julgue que está dialogando com outro ser humano, então a máquina deve ser considerada pensante ou inteligente. Turing não está se referindo diretamente à “consciência”, mas na discussão das possíveis críticas à sua abordagem, considera o que chama “o argumento da consciência”:

Este argumento parece ser uma negação da validade de nosso teste. De acordo com a forma mais extremada de tal concepção, a única maneira de alguém estar seguro de que a máquina pensa é *ser* ele a própria máquina e sentir-se pensando. Poderia então descrever estes sentimentos [*feelings*] ao mundo, mas naturalmente não se justificaria que alguém lhe desse atenção. De modo semelhante, de acordo com a mesma concepção, a única maneira de saber-se se um *homem* pensa é ser esse homem. Trata-se, de fato, do ponto de vista solipsista. (TURING, 1973, p. 65; 1950, p. 446)

Turing conclui então que o proponente do argumento da consciência não iria querer assumir uma posição “solipsista”, ou seja, uma posição que assume que só existiria a sua própria mente. Finalizando a discussão deste argumento:

⁵ TURING, A.M. (1950), “Computing machinery and intelligence”, *Mind* 59, pp. 433-60. Em português: “Computadores e inteligência”, trad. M. Epstein, in EPSTEIN, I. (org.) (1973), *Cibernética e comunicação*, Cultrix, São Paulo, pp. 45-82. A expressão “teste de Turing robótico” encontra-se em HARNAD, S. & SCHERZER, P. (2008), “First, scale up to the robotic Turing test, then worry about feeling”, *Journal Artificial Intelligence in Medicine* 44, pp. 83-89. Sobre a previsão de 2029: KURZWEIL, R. (2005), *The singularity is near*, Penguin, New York, p. 295.

Não quero dar a impressão de que penso que não exista nenhum mistério no que respeita à consciência. Existe, por exemplo, algo assim como que um paradoxo vinculado às tentativas de localizá-la. Mas não acredito que tais mistérios tenham de ser necessariamente resolvidos antes de podermos responder à pergunta que nos preocupa neste artigo. (TURING, 1973, p. 67; 1950, p. 447)

As condições específicas colocadas na primeira citação de Turing ainda não foram plenamente satisfeitas, mas a cada ano o Prêmio Loebner é oferecido para a melhor performance de máquina. O futurólogo Ray Kurzweil prevê que o teste de Turing será plenamente superado em torno de 2029.

Voltemos agora para o teste de Turing robótico envolvendo a ginoide Maria. A pergunta a ser considerada é a seguinte: a ginoide é consciente? Por hipótese, ela se comporta como uma humana, fala como uma humana, é maquiavélica como nós, tem a pele igual à de um humano e tem cheiro de humano, como Ava, do filme *Ex machina* (2015). Mas Ava tem um cérebro feito de matéria especial, ao passo que a ginoide que estamos discutindo é feita de circuitos integrados. Maria teria consciência?

3. Comportamentalismo filosófico

A concepção de que a natureza de uma coisa se esgota com sua aparência, ou com seu comportamento externo, acoplado a informações sobre a sua história, pode ser chamada de “comportamentalismo”, ou *behaviorismo*. Como é bem sabido, este termo está associado a uma corrente específica da psicologia, que na verdade se divide em diversas correntes. O’DONAHUE & KITCHENER (1999) descrevem pelo menos catorze versões do comportamentalismo, dez psicológicas e quatro filosóficas. Não exploraremos este tema aqui e agora, mas definiremos sucintamente o “comportamentalismo filosófico” como a visão que afirma que um estado mental só pode ser atribuído a um sistema baseado em seu comportamento observável. Isso é resumido na frase de Wittgenstein: “Um ‘processo interno’ necessita de critérios externos” (*Investigações filosóficas*, §580). Esta concepção é expressa na seguinte citação do psicólogo John Staddon, referindo-se à corrente que propõe, chamada “behaviorismo teórico”.⁶

O behaviorismo teórico adota a visão do teste de Turing da consciência. Porém, esta visão não é aceita por todos. John Searle (1992), se o compreendi corretamente, argumenta que se um dispositivo fosse encontrado que conseguisse passar no teste de Turing, mesmo assim ele não seria consciente. Tenho três reações a isso. Primeiro, a suposição de que tal dispositivo *pode* ser criado somente a partir de hardware pode ser falsa, e neste caso nada mais precisa ser dito. Em segundo lugar está uma pergunta óbvia: supondo que ele possa ser criado, *como você sabe* que ele não é consciente? A única resposta a esta pergunta é: porque ele não passa no teste de Turing, o que contraria a primeira suposição. Em outras palavras, se a única maneira de saber que alguém (ou alguma coisa) é consciente é porque ela responde às nossas perguntas de

⁶ STADDON, J.E.R. (1999), “Theoretical behaviorism”, in O’DONAHUE, W. & KITCHENER, R. (orgs.), *Handbook of behaviorism*, Academic, San Diego, pp. 217-41. Em português, ver ABIB, J.A.D. (2015), “Psicologia, comportamentalismo e subjetividade”, in CHITOLINA, C.L.; PEREIRA, J.A. & PINTO, R.H. (orgs.), *Mente, cérebro e consciência: um confronto entre filosofia e ciência*, Paco, Jundiaí, pp. 279-96. A definição sucinta de comportamentalismo filosófico é baseada em OPPY, G. & DOWE, D. (2016), “The Turing test”, *Stanford Encyclopedia of Philosophy* (online). WITTGENSTEIN, L. (1979), *Investigações filosóficas*, Coleção Os Pensadores, 2ª ed., trad. J.C. Bruni, Abril Cultural, São Paulo (orig. em inglês: 1953). Mencionado na citação: SEARLE, J. (1992), *The rediscovery of the mind*, MIT Press, Cambridge (MA); em português: *A redescoberta da mente*, trad. E.P. Ferreira, Martins Fontes, São Paulo, 1997.

maneira apropriada, então, por definição, uma máquina que consegue passar no teste de Turing deve ser consciente.

Minha terceira reação é simplesmente esperar para ver. Se uma máquina for criada que passe no teste de Turing, as pessoas logo irão tratá-la como alguém de seu tipo. Se estivermos dispostos a atribuir consciência para um cachorro, ou para alguém cuja habilidade para comunicar esteja tão prejudicada como para Helen Keller, será que deixaremos de atribuí-la para um dispositivo que fale e responda de maneira indistinguível de um ser humano? (STADDON, 1999, p. 230)

4. Mentalismo

A posição que nega que a ginoide tenha consciência, na falta de um nome melhor, pode ser chamada de “mentalismo”, em oposição ao comportamentalismo filosófico. O *mentalismo*, então, seria a tese de que há uma perspectiva qualitativa ou subjetiva que caracteriza a consciência, e que vai além do comportamento externamente verificável. Esta tese está encapsulada na citação de Hamilton da seção I.1.

Um debate opondo mentalistas e comportamentalistas se deu na década de 1980 com relação à consciência animal. O estudioso do comportamento animal e psicólogo gestaltista Wolfgang Köhler havia colhido na ilha de Tenerife, durante a Primeira Guerra Mundial, indícios de comportamento inteligente em chimpanzés, que conseguiam empilhar caixas para subir e pegar bananas presas no alto da jaula. Ele concluiu que isso teria sido feito não por mera tentativa e erro, mas por “insight” (*Einsicht*), sugerindo que os chimpanzés possuem consciência como nós (a menos da linguagem e outras funções cognitivas superiores). Em 1945, Herbert G. Birch mostrou que experiência prévia com os instrumentos era necessária para a ocorrência deste comportamento inteligente. Em 1981, Epstein, Lanza & Skinner também criticaram a conclusão de Köhler, mostrando que é possível condicionar pombos com um repertório de comportamentos de maneira que eles possam reproduzir o comportamento sofisticado observado em chimpanzés. Para eles, isso sugeria que era leviano atribuir uma mente aos macacos, pois o seu comportamento poderia ser completamente explicado pelo seu repertório de condicionamentos, conforme sugerido por Birch. Este resultado é muito interessante, mas não afeta a tese de que há um estado mental que sirva de causa intermediária na cadeia causal ligando condicionamentos e comportamento.⁷ Mais tarde discutiremos a questão de se animais não humanos têm consciência, e a evidência de inteligência em pássaros.

⁷ EPSTEIN, R.; LANZA, R.P. & SKINNER, B.F. (1981), “‘Self-awareness’ in the pigeon”, *Science* 212: 695-6. Ver também o curto filme lançado por BAXLEY, N. (1982), *Cognition, creativity and behavior: the columban simulations*, disponível em <https://www.youtube.com/watch?v=QKSvu3mj-14> (parte 1) e ... [v=erhmslcHvaw](https://www.youtube.com/watch?v=erhmslcHvaw) (parte 2), com os comentários de Skinner sobre Köhler. Sobre o trabalho deste: KÖHLER, W. (1917), *Intelligenzprüfungen an Anthropoiden*, Königlich-Preußische Akademie der Wissenschaften, Berlin; tradução para o inglês: *The mentality of apes*, trad. Ella Winter, Kegan Paul, Trench & Trubner, London, 1925. Para uma resenha do assunto, ver SHETTLEWORTH, Sara J. (2012), “Do animals have insight, and what is insight anyway?”, *Canadian Journal of Experimental Psychology* 66: 217-26

5. Funcionalismo de estados mentais

Na filosofia da mente, o funcionalismo é a posição que define estados mentais por suas *funções*, e não por sua natureza material.⁸ Assim, os estados mentais seriam constituídos pelas relações causais existentes entre eles, além de pelas entradas sensoriais e pelas saídas comportamentais. Não importaria qual é o substrato da mente – matéria orgânica, matéria inorgânica, espírito –, o que importa seria a organização do sistema, ou o estado informacional. O chamado “funcionalismo de máquina” compara estados mentais com estados lógicos (*software*) de um computador, que existiriam independentemente da natureza física do computador (*hardware*).

O funcionalismo foi articulado na década de 1960 por autores como Hilary Putnam e Jerry Fodor,⁹ mas a sua tese principal remonta à Antiguidade (ver seção IV.3, sobre Aristóteles).

A intuição por trás do funcionalismo é que o que determina o tipo psicológico ao qual um particular mental pertence é o papel causal do particular na vida mental do organismo. Individualização funcional é diferenciação com respeito ao papel causal. Uma dor de cabeça, por exemplo, é identificada com o tipo de estado mental que, entre outras coisas, causa uma disposição para tomar aspirina em pessoas que acreditam que a aspirina alivia uma dor de cabeça, causa um desejo de se livrar da dor que se está sentindo, frequentemente causa alguém que fala português a dizer coisas como “Estou com dor de cabeça”, e surge a partir de excesso de trabalho, fadiga ocular e tensão. Esta lista presumivelmente não está completa. Mais será conhecido sobre a natureza de uma dor de cabeça, à medida que pesquisas em psicologia e fisiologia descobrirem mais sobre o seu papel causal (FODOR, 1981, p. 128).

Diversas faculdades mentais são definidas de maneira funcional, como por exemplo “capacidade de cálculo”. Um indivíduo com “síndrome de *savant*” tem uma capacidade de cálculo assombrosa, mas um computador de última geração tem uma capacidade ainda maior. Neste exemplo, a faculdade mental é definida pela eficiência de realizar um cálculo matemático, não importando como e nem qual o substrato material.

A tese funcionalista forte é que todos os estados mentais são definíveis em termos funcionais. Isso incluiria a própria consciência. A tese de que máquinas mecânicas ou computadores de circuitos integrados de silício podem ter consciência é chamada “funcionalismo de máquina” ou “inteligência artificial forte” (vimos, porém, que o debate se complica pelo fato de diferentes definições de consciência serem usados). Uma maneira de exprimir esta posição é dizer que computadores não apenas podem “simular” a mente humana, como se tenta fazer com o *Blue Brain Project* (Projeto Cérebro Azul) de Lausanne, mas poderão “emular” a consciência, ou seja, fazê-la emergir a partir de uma computação altamente complicada. Certamente um computador só pode simular um furação, não emulá-lo, pois o seu interior não fica molhado; será que um computador poderia ir além da simulação de um cérebro, emulando a consciência subjetiva? O funcionalismo de máquina defende que sim.

⁸ Um exemplo (já antiquado) de definição funcional é a de “carburador”. Um carburador é definido como qualquer coisa que misture combustível e ar em um motor, usando a sucção para introduzir o combustível. Em princípio, pode-se construir um carburador com qualquer material, desde que ele cumpra a função que o define.

⁹ Em nosso curso, estudamos a tradução de: FODOR, J.A. (1981), “The mind-body problem”, *Scientific American* 244(1): 124-32, 148 (janeiro). Ver também PUTNAM, H.W. (1967), “Psychological predicates”, in CAPITAN, W.H. & MERRILL, D.D. (orgs.) (1967), *Art, mind and religion*, U. Pittsburgh Press, pp. 37-48 (republicado com o título “The nature of mental states”). Uma discussão aprofundada do funcionalismo encontra-se em BLOCK, N. (org.) (1980), *Readings in the philosophy of psychology*, vol. 1, Harvard University Press, Cambridge, pp. 171-306.

6. Aspecto qualitativo do mental

O maior problema para o funcionalismo é dar conta do aspecto qualitativo do mental, o que Fodor chama de “conteúdo qualitativo”:

Não é fácil dizer o que é o conteúdo qualitativo. De fato, segundo algumas teorias, não é possível sequer dizer o que ele é porque ele não pode ser conhecido por descrição, mas apenas através da experiência direta. Apesar disso, eu tentarei descrevê-lo. Tente imaginar-se olhando para uma parede em branco através de um filtro vermelho. Agora, troque o filtro vermelho por um verde e deixe o resto exatamente como estava antes. Alguma coisa relativa ao caráter de sua experiência se altera quando o filtro é trocado, e é esse tipo de coisa que os filósofos chamam de conteúdo qualitativo. [...]

A razão pela qual o conteúdo qualitativo representa um problema para o funcionalismo é evidente. O funcionalismo está comprometido com a definição dos estados mentais em termos de suas causas e seus efeitos. Parece possível, no entanto, que dois estados mentais possam ter as mesmas relações causais, diferindo, porém, em seu conteúdo qualitativo (FODOR, 1981, p. 130).

Este é um ponto central a ser explorado ao longo de todo curso: a natureza “qualitativa” da vivência subjetiva consciente. Termos como “dados dos sentidos” e “qualia” serão usados na discussão deste ponto. Argumentos como o do “espectro invertido”, que remontam a John Locke e que examinaremos mais tarde, são usados para questionar a validade do funcionalismo de estados mentais, como colocado por Fodor na última oração citada acima.

7. Experimento mental da troca de células cerebrais por chips

Uma boa maneira de explorar as posições em filosofia da mente é apresentando experimentos mentais (*Gedankenexperimenten*, *thought experiments*, também traduzido por “experimentos de pensamento”), ou seja, experimentos que não podem ser realizados, pelo menos não hoje em dia (por razões técnicas), ou então por uma questão de princípio. Neste curso, diversos experimentos mentais nos guiarão na exploração da consciência e do cérebro.

Já examinamos o experimento mental da ginoide feita de circuitos integrados, que definiu duas posições: o comportamentalismo filosófico e o mentalismo. Consideremos agora o experimento mental da troca de células cerebrais por chips de silício, explorada por John SEARLE (1997, p. 98)¹⁰:

Imagine que seu cérebro comece a degenerar-se de tal forma que, aos poucos, você vai ficando cego. Imagine que os médicos, desesperados, ansiosos por aliviar seu sofrimento, experimentem qualquer método para recuperar sua visão. Como último recurso, tentam implantar circuitos integrados de silício dentro de seu córtex visual. Suponha que, para seu assombro e também deles, os circuitos integrados de silício devolvam sua visão a seu estado normal. Agora imagine que, para sua maior depressão, seu cérebro continue a degenerar-se, e que os médicos continuem a implantar mais circuitos integrados de silício. Você já pode perceber aonde o experimento de pensamento vai chegar: no final, podemos imaginar que seu cérebro estará inteiramente

¹⁰ SEARLE, J. ([1992] 1997), op. cit. (nota 6), Cap. 3, seção I.

substituído por circuitos integrados de silício; que, ao balançar a cabeça, você poderá ouvir os circuitos integrados chocalhando por todos os lados dentro de seu crânio.

Não exploraremos a discussão subsequente de Searle, mas iremos supor que cada célula de encéfalo é substituída por um chip que reproduz as mesmas relações causais (de entrada e saída) conhecidas hoje em dia, e que os chips tenham a mesma plasticidade que as células reais, estabelecendo novas conexões sinápticas, etc. Ao final da substituição, a “pessoa” terá consciência ou não?

Não temos certeza do que aconteceria, pois o experimento é apenas imaginado. Mas podemos definir duas posições antagônicas com relação ao resultado esperado do experimento. A suposição de que a consciência se manteria intacta caracteriza um *funcionalismo no nível de células biológicas*: a organização entre as células (desprezando os detalhes internos) seria suficiente para fazer emergir a consciência subjetiva. Esta é uma posição mentalista, pois não define “consciência” apenas a partir do comportamento externo. Em oposição a este tipo de funcionalismo, temos a posição que Searle chama de *naturalismo biológico*: alguma coisa que ocorre nos processos biológicos dentro das células seria essencial para a emergência da consciência (esta também é uma posição mentalista). Assim, das três alternativas colocadas na citação de Staddon, a primeira seria a correta.

8. O espectro funcionalista / substancialista

Para fechar este capítulo, consideremos com a Tabela I.1 as várias escalas em que se dá a oposição entre, de um lado, o comportamentalismo e funcionalismo, e de outro, o mentalismo e o naturalismo biológico. Representamos cinco posições (linhas na Tabela), mas outras intermediárias são possíveis. O que caracteriza o lado funcionalista do espectro é considerar que a consciência possuída por um sistema define-se somente pelas *relações* entre partes do sistema (ou seja, pela organização do sistema, o “arranjo” dos atomistas, a “forma” aristotélica), ou simplesmente pelo comportamento do sistema, sem importar qual é a constituição material do sistema (ou a constituição espiritual).

No limite, a posição metafísica mais radical tem sido chamada de “estruturalismo ontológico” (ou realismo estrutural ontológico), e considera que tudo em nosso Universo são relações ou estruturas, e que no fundo a própria matéria (e outras grandezas físicas) emerge de relações sem *relata*, ou seja, de relações de relações de relações, etc., sem a presença de elementos básicos.¹¹ A negação desta concepção pode ser chamada de *substancialismo*, ou seja, a noção de que há uma substância, que pode ser a matéria (ou outra entidade física) ou outra categoria de entidade, irreduzível a relações. O hilemorfismo de Aristóteles seria uma visão substancialista (mesmo que mínima), na medida em que na natureza não haveria forma sem matéria. O naturalismo biológico seria também uma expressão do substancialismo, de maneira que o espectro complementar ao espectro funcionalista pode ser chamado de “espectro substancialista”.

¹¹ LADYMAN, J. (2014), “Structural realism”, in *Stanford encyclopedia of philosophy*, online. Em português: STEINLE, W. (2010), “O realismo estrutural ontológico e o problema das relações sem os relata”. *Analytica 14*: 29-51.

EXPERIMENTO MENTAL	PERGUNTA	SIM	NÃO
Ginoide de circuitos integrados	A ginoide é consciente?	Comportamentalismo filosófico	Mentalismo (genérico)
Programa que simula estados mentais ou células cerebrais (<i>Blue Brain Project</i>)	Um simulador da mente ou cérebro pode ser consciente?	Funcionalismo de máquina (ou de estados mentais) (I.A. forte)	Mentalismo anti-funcionalista
Substituição de células cerebrais por chips	Substituindo cada célula por chips com mesma função preserva a consciência?	Funcionalismo de células biológicas	Naturalismo biológico celular
Substituição de partes das células (biologia sintética)	Substituindo protoplasma e organelas de células preserva a consciência?	Funcionalismo subcelular	Naturalismo biológico subcelular
-----	A realidade consiste só de relações, sem <i>relata</i> ?	Estruturalismo ontológico radical	Substancialismo mínimo

Tabela I.1: O espectro funcionalista é representado na terceira coluna, e seu complementar na quarta coluna (o espectro substancialista).

9. Uma máquina pode ser consciente?

Vemos que pergunta sobre se “máquinas podem ter consciência” pode obter diversas respostas. Se definirmos consciência no sentido do comportamentalismo filosófico, então parece muito plausível que atribuiremos consciência a robôs em torno de 2030 (segundo a estimativa de Kurzweil). Porém, em nosso curso, adotaremos a definição mentalista de consciência, que considera que vivenciamos subjetivamente um estado qualitativo que escapa (num certo sentido) às palavras e que não pode (pelo menos até hoje em dia) ser observado diretamente do ponto de vista externo, objetivo, por outras pessoas.

Da perspectiva mentalista, coloca-se então a questão de como surge a consciência, a partir do corpo ou a partir de duas substâncias separadas (o debate entre materialismo e espiritualismo, a ser a explorado no próximo capítulo)? Será que a consciência é fruto da organização entre as partes, sendo irrelevante a natureza física (ou espiritual) que compõe essas partes, ou será que a natureza das partes é essencial para a consciência? A posição de que a natureza da matéria é essencial (opondo-se ao funcionalismo) é uma forma de “substancialismo”.

Vimos que este debate se dá em diversas escalas (Tabela I.1), e em cada uma delas há uma resposta diferente à questão de se máquinas podem ter consciência. Para o funcionalismo de estados mentais, bastaria a máquina reproduzir as relações causais entre estados mentais (o chamado “funcionalismo de máquina”) para ela ter consciência. Para o funcionalismo de células biológicas, seria preciso reproduzir as relações causais entre as células do encéfalo para que emergisse a consciência em um robô.

A posição substancialista que se opõe a isso (chamada por Searle¹² de “naturalismo biológico”) supõe que há algo dentro da célula biológica que é essencial para a consciência.

¹² SEARLE, J. ([1992] 1997), op. cit. (nota 6), Cap. 1, seção I.

Mas se descobríssemos que uma parte da célula é essencial, então as outras partes poderiam ser substituídas por partes sintéticas, como explorada hoje em dia na área chamada de “biologia sintética”. Neste caso, para esta visão, um robô artificial poderia ter consciência se ele carregasse em suas células esta parte material essencial. Esta seria uma solução substancialista.

O funcionalismo, porém, nunca é completamente derrotado, constatação esta que pode ser chamado de “paradoxo das matrioscas”.¹³ Nessa altura, o funcionalista poderia descer um nível abaixo, e argumentar que aquela parte da célula, essencial para a consciência, é também composta de partes, por exemplo macromoléculas (proteínas, etc.), e que seria possível substituir essas macromoléculas por elementos artificiais, conservando-se as relações entre as partes. Nesta escala, o funcionalismo subcelular argumentaria que a consciência é fruto apenas da organização das partes do sistema, sendo irrelevante a natureza do substrato material. Neste caso, para esta visão, seria possível uma máquina completamente artificial ter consciência. Mas o substancialista nega que isso seja possível.

¹³ A analogia com as matrioscas (ou matriochkas), as bonecas russas que se encaixam uma dentro da outra, originou-se com o sociólogo estadunidense Talcott Parsons, que desenvolveu o funcionalismo na sociologia.