

II

PODEM OS COMPUTADORES PENSAR?

No capítulo anterior, forneci, pelo menos, as leis gerais de uma solução para o chamado «problema da ‘Mente-Corpo’». Embora não saibamos em pormenor como funciona o cérebro, conhecemos o bastante para ter uma ideia das relações gerais entre os processos cerebrais e os processos mentais. Os processos mentais são causados pelo comportamento dos elementos do cérebro. Ao mesmo tempo, realizam-se na estrutura que é constituída por esses elementos. Penso que esta resposta se harmoniza com as abordagens biológicas correntes aos fenómenos biológicos. Sem dúvida, é uma espécie de resposta do senso comum à questão, dado o que conhecemos acerca do modo como o Mundo funciona. No entanto, é um ponto de vista de uma minoria. A concepção predominante em Filosofia, Psicologia e Inteligência Artificial, é a que realça as analogias entre o funcionamento do cérebro humano e o funcionamento dos computadores digitais. Segundo a versão mais extrema desta concepção, o cérebro é justamente um computador digital e a mente

é um programa de computador. Poder-se-ia resumir esta concepção — dou-lhe o nome de «Inteligência Artificial forte» ou «IA forte» — dizendo que a mente está para o cérebro tal como o programa está para o *hardware* do computador.

Esta concepção tem a consequência de que nada existe de essencialmente biológico acerca da mente humana. Acontece que o cérebro é um de entre o número indefinidamente vasto de diferentes tipos de computadores materiais que poderiam apoiar os programas constitutivos da inteligência humana. Nesta concepção, qualquer sistema físico que tivesse um programa correcto com as entradas e saídas correctas teria uma mente, no mesmíssimo sentido em que vocês e eu temos mentes. Assim, por exemplo, se fizéssemos um computador de velhas latas de cerveja com energia fornecida por moinhos de vento, se ele tivesse o programa correcto teria de ter uma mente. E o importante não é que ele, por tudo o que sabemos, poderia ter pensamentos e sentimentos, mas antes que deve ter pensamentos e sentimentos, porque o ter pensamentos e sentimentos consiste justamente nisto: levar a cabo o programa correcto.

A maior parte dos que defendem esta concepção pensa que não projectámos ainda programas que sejam mentes. Mas existe entre eles um acordo muito geral de que é apenas uma questão de tempo, até que os cientistas de computadores e os que trabalham na Inteligência Artificial projectem o *hardware* apropriado e os programas que serão o equivalente dos cérebros e das mentes humanas. Serão esses os cérebros e mentes artificiais que de todos os modos constituem o equivalente dos cérebros e mentes humanas.

Muitas pessoas fora do campo da Inteligência Artificial ficam deveras espantadas por descobrir que alguém possa acreditar numa tal concepção. Assim, antes de a criticar, permitam que eu lhes forneça alguns exemplos das coisas que os que trabalham neste campo efectiva-

mente disseram. Herbert Simon da Carnegie-Mellon University diz que já temos máquinas que podem literalmente pensar. Já que não é preciso esperar por alguma máquina futura, porque os computadores digitais existentes já têm pensamentos, no mesmíssimo sentido em que vocês e eu temos. Ora vejam lá! Os filósofos preocuparam-se durante séculos acerca de se ou não uma máquina podia pensar e agora descobrimos que eles já têm tais máquinas na Carnegie-Mellon. O colega de Simon, Alan Newell afirma que «já descobrimos» (notem que Newell diz «descobrimos», não «supusemos» ou «considerámos a possibilidade», mas *descobrimos*) que a inteligência é justamente uma questão de manipulação de símbolos físicos; não tem nenhuma ligação essencial com qualquer tipo de material ou humidade biológica ou física. Antes, qualquer sistema que seja capaz de manipular símbolos físicos de modo correcto é capaz de inteligência no mesmo sentido literal que a inteligência humana dos seres humanos. Simon e Newell sublinham, pela sua honra, que não existe nada de metafórico nestas pretensões; proferem-nas de um modo inteiramente literal. Freeman Dyson é citado como tendo dito que os computadores têm uma vantagem sobre todos nós, no tocante à evolução. Visto que a consciência é uma questão de processos formais, nos computadores esses processos formais podem ocorrer em substâncias que estão muito mais capacitadas para sobreviver num universo que está a arrefecer do que seres como nós, feitos de materiais húmidos e sujos. Marvin Minsky do MIT diz que a próxima geração de computadores será tão inteligente que «teremos muita sorte se eles permitirem manter-nos em casa como animais de estimação domésticos». O meu preferido de sempre, na literatura das afirmações exageradas em prol do computador digital, vem de John McCarthy, o inventor do termo «Inteligência Artificial». McCarthy diz que mesmo «máquinas tão simples como termostatos têm — pode dizer-se — crenças». E, efectivamente, segundo ele, quase toda a máquina

capaz de resolver problemas tem, pode dizer-se, crenças. Admiro a coragem de McCarty. Uma vez perguntei-lhe: «Que crenças tem o seu termostato?», e ele respondeu: «O meu termostato tem três crenças — está demasiado quente aqui, está demasiado frio aqui e está bem aqui.» Como filósofo, aprecio estas afirmações por uma simples razão. Diferentemente da maior parte das teses filosóficas, elas são razoavelmente claras e admitem uma simples e decisiva refutação. É essa refutação que eu vou emprender neste capítulo.

A natureza da refutação nada tem a ver com qualquer estádio particular da tecnologia dos computadores. É importante sublinhar este ponto, porque a tentação é sempre pensar que a solução para os nossos problemas deve esperar alguma maravilha tecnológica ainda não criada. Mas, de facto, a natureza da refutação é totalmente independente de qualquer estado da tecnologia. Tem a ver com a justa definição de um computador digital, com aquilo que um computador digital é.

Essencial à nossa concepção de um computador digital é que as suas operações possam ser especificadas em termos puramente formais; isto é, especificamos os passos na operação do computador em termos de símbolos abstractos — sequências de zeros e uns impressos numa fita, por exemplo. Uma «regra» típica do computador determinará que, quando a máquina está num certo estádio e tem um certo símbolo na sua fita, então realizará uma certa operação, como o apagamento de um símbolo ou a impressão de outro símbolo, e então ocorrerá um outro estado, como o movimento da fita um quadrado para a esquerda. Mas os símbolos não têm significado; não têm conteúdo semântico; não são acerca de qualquer coisa. Têm de ser especificados unicamente em termos da sua estrutura formal ou sintática. Os zeros e os uns, por exemplo, são simples numerais; nem sequer estão em vez de números. Efectivamente, é esta característica dos computadores digitais que os torna tão poderosos.

Um e o mesmo tipo de *hardware*, se for apropriadamente projectado, pode utilizar-se para executar um âmbito indefinido de programas diferentes. Um e o mesmo programa pode passar num âmbito indefinido de diferentes tipos de *hardwares*.

Mas esta característica dos programas, que se definem em termos puramente formais ou sintáticos, é fatal para a concepção de que os processos mentais e os processos de programa são idênticos. E a razão pode formular-se de um modo muito simples. É muito mais complexo ter uma mente do que ter processos formais ou sintáticos. Os nossos estados mentais internos têm, por definição, certos tipos de conteúdos. Se estou a pensar em Kansas City, ou se desejo beber uma cerveja fresca, ou se estou a imaginar que vai haver uma baixa nas taxas de juro, em cada caso, o meu estado mental tem um certo conteúdo mental, além de quaisquer estruturas formais que possa ter. Isto é, mesmo se os meus pensamentos ocorrem em séries de símbolos, deve haver algo mais no pensamento do que as séries abstractas, porque as séries por si mesmas não têm qualquer significado. Se os meus pensamentos são *acerca de alguma coisa*, então as séries devem ter um *significado*, que faz que os pensamentos sejam a propósito dessas coisas. Numa palavra, a mente tem mais do que uma sintaxe, possui também uma semântica. A razão por que nenhum programa de computador pode *alguma vez* ser uma mente é simplesmente porque um programa de computador é apenas sintático, e as mentes são *mais* do que sintáticas. As mentes são semânticas, no sentido de que possuem *mais* do que uma estrutura formal, *têm* um conteúdo.

Para ilustrar este ponto, concebi uma certa experiência intelectual. Imaginemos que um grupo de programadores de computador escreveram um programa que capacitará um computador para simular a compreensão do chinês. Assim, por exemplo, se ao computador se puser uma questão em chinês, ele conferirá a questão

com a sua memória ou a base de dados e produzirá respostas apropriadas para as perguntas em chinês. Suponhamos, em vista da discussão, que as respostas do computador são tão boas como as de um falante chinês nativo. Ora bem, entenderá o computador nesta base o chinês, compreende ele literalmente o chinês tal como os falantes chineses entendem o chinês? Bem, imaginemos que alguém está fechado num quarto e que neste quarto há vários cestos cheios de símbolos chineses. Imaginemos que alguém, como eu, não compreende uma palavra de chinês, mas que lhe é fornecido um livro de regras em inglês para manipular os símbolos chineses. As regras especificam as manipulações dos símbolos de um modo puramente formal em termos da sua sintaxe e não da sua semântica. Assim a regra poderá dizer: «Tire do cesto número um um símbolo esticado e ponha-o junto de um símbolo encolhido do cesto número dois.» Suponhamos agora que alguns outros símbolos chineses são introduzidos no quarto e que esse alguém recebe mais regras para passar símbolos chineses para o exterior do quarto. Suponhamos que, sem ele saber, os símbolos introduzidos no quarto se chamam «perguntas» feitas pelas pessoas que se encontram fora do quarto e que os símbolos mandados para fora do quarto se chamam «respostas às perguntas». Suponhamos, além disso, que os programadores são tão bons a escrever programas e que alguém é igualmente tão bom em manipular os símbolos que muito depressa as suas respostas são indistinguíveis das de um falante chinês nativo. Lá está ele fechado no quarto manipulando os símbolos chineses e passando cá para fora símbolos chineses em resposta aos símbolos chineses que são introduzidos. Com base nesta situação tal como a descrevi, de nenhum modo se pode aprender chinês pela simples manipulação desses símbolos formais.

Ora, o cerne da história, é apenas este: em virtude da realização de um programa formal de computador do ponto de vista de um observador externo, esse alguém

comporta-se exactamente como se entendesse chinês, mas de qualquer modo não compreende uma só palavra de chinês. Mas, se a efectivação do programa apropriado do computador para a compreensão do chinês não é suficiente para nos *dar* uma compreensão do chinês, então também não basta dar a qualquer outro computador digital *uma compreensão* do chinês. E, mais uma vez, a razão para isso, pode enunciar-se de um modo muito simples. Se não compreendemos o chinês, então nenhum outro computador pode compreender o chinês, porque nenhum computador digital, em virtude da simples execução de um programa, tem algo que nós não tenhamos. Tudo o que o computador tem, como nós temos, é um programa formal para manipular símbolos chineses não interpretados. Repetindo, um computador tem uma sintaxe, mas não uma semântica. Tudo o que a parábola do quarto chinês pretende é lembrar um facto que já conhecíamos. Entender uma língua ou, sem dúvida, ter estados mentais, implica mais do que a simples posse de um feixe de símbolos formais. Implica ter uma compreensão ou um significado associado a esses símbolos. E o computador digital, como foi definido, só pode ter símbolos formais, porque a operação de um computador, como eu disse antes, define-se em termos da sua capacidade para realizar programas. E estes programas só podem especificar-se de um modo puramente formal — isto é, não têm conteúdo semântico.

Podemos ver a força deste argumento, se contrastarmos o que é ser interrogado e dar respostas em inglês e ser interrogado e dar respostas numa língua em que não temos conhecimento de qualquer dos significados das palavras. Imaginemos que, no quarto chinês, nos fazem também perguntas em inglês acerca de coisas como a nossa idade ou a história da nossa vida, e que nós respondemos a essas questões. Qual a diferença entre o caso chinês e o caso inglês? Ora bem, se, como eu, vocês não compreendem chinês e entendem inglês, então a dife-

rença é óbvia. Vocês compreendem as perguntas em inglês porque são expressas em símbolos cujos significados são conhecidos. De modo semelhante, quando vocês fornecem respostas em inglês, produzem símbolos que são significativos para vocês. Mas, no caso do chinês, nada disso se tem. No caso do chinês, vocês simplesmente manipulam símbolos formais segundo um programa de computador e não atribuem nenhum significado a qualquer dos elementos.

Várias objecções foram sugeridas contra este argumento por aqueles que trabalham em Inteligência Artificial, em Psicologia e em Filosofia. Todas têm algo em comum; todas são inadequadas. Existe uma razão óbvia por que é que elas têm de ser inadequadas, visto que o argumento se baseia numa verdade lógica muito simples, a saber, a sintaxe sozinha não basta para a semântica e os computadores digitais na medida em que são computadores têm, por definição, apenas uma sintaxe.

Quero esclarecer bem isto mediante a consideração de alguns argumentos que, muitas vezes, me contrapõem.

Há pessoas que tentam responder ao exemplo do quarto chinês dizendo que todo o sistema compreende chinês. A ideia aqui é que, embora eu, a pessoa que no quarto manipula os símbolos não entenda chinês, sou justamente a unidade central do processamento do sistema do computador. Elas afirmam que é todo o sistema, incluindo o quarto, os cestos cheios de símbolos e os discos que contêm os programas, e talvez outros artigos também, tomados como uma totalidade, que compreende o chinês. Mas isto encontra-se exactamente sujeito à mesma objecção que antes fiz. Não há maneira de o sistema poder passar da sintaxe para a semântica. Eu, enquanto unidade central de processamento, não tenho maneira de calcular qual o significado de qualquer um desses símbolos; mas também não o consegue todo o sistema.

Outra resposta de sentido comum é imaginar que temos o programa da compreensão do chinês dentro

de um robô. Se o robô se deslocasse e interagisse causalmente com o Mundo, não seria isso suficiente para garantir que ele compreendia o chinês? Mais uma vez, a inexorabilidade da distinção semântica/sintaxe supera esta manobra. Enquanto supusermos que o robô tem apenas o computador por um cérebro então, mesmo que se pudesse comportar exactamente como se compreendesse o chinês, não conseguiria ainda passar da sintaxe para a semântica do chinês. Vocês podem ver isso se imaginarem que eu sou um computador. Numa divisão no crânio do robô, eu manipulo símbolos sem saber que alguns deles vêm até mim de câmaras de televisão, ligadas à cabeça do computador, e outras saem para mover os braços e as pernas do robô. Enquanto eu tiver apenas um programa formal de computar, não tenho maneira alguma de ligar qualquer significado a qualquer dos símbolos. E o facto de o robô estar inserido em interacções causais com o mundo exterior não me ajudará a ligar qualquer significado com o símbolo, a não ser que eu tenha algum modo de descobrir esse facto. Suponhamos que o robô apanha um hambúrguer e isso dispara e faz aparecer no quarto o símbolo para hambúrguer. Enquanto eu tive apenas o símbolo sem qualquer conhecimento das suas causas ou do modo como ele ali apareceu, não tenho maneira de conhecer o que ele significa. As interacções causais entre o robô e o resto do Mundo são irrelevantes, a não ser que essas interacções causais sejam representadas em alguma ou noutra mente. Mas, não há nenhuma possibilidade de assim ser, se tudo aquilo em que consiste a chamada Mente é apenas um conjunto de operações puramente formais e sintáticas.

É importante ver exactamente o que se pretende e o que não se pretende com o meu argumento. Suponhamos que eu faço a pergunta que mencionei no princípio: «Pode uma máquina pensar?» Bem, num certo sentido, naturalmente, todos nós somos máquinas. Podemos imaginar a matéria dentro das nossas cabeças como uma máquina

de carne. E, naturalmente, todos podemos pensar. Assim, no sentido de «máquina», a saber, o sentido em que uma máquina é justamente um sistema físico que é capaz de realizar certos tipos de operações, nesse sentido, todos somos máquinas e podemos pensar. Assim, de uma maneira trivial, há máquinas que podem pensar. Mas esta não era a questão que nos preocupava. Assim, tentemos uma diferente formulação. Pode um artefacto pensar? Pode uma máquina feita pelo homem pensar? Bem, mais uma vez, depende do tipo de artefacto. Suponhamos que projectámos uma máquina que era indistinguível, molécula a molécula, de um ser humano. Ora bem, se se podem duplicar as causas, também se presumivelmente duplicar os efeitos. Assim, de novo, a resposta a esta questão é, pelo menos em princípio, trivialmente sim. Se se pudesse construir uma máquina que tivesse a mesma estrutura de um ser humano, então presumivelmente essa máquina seria capaz de pensar. Na realidade, seria um ser humano de substituição. Ora, tentemos novamente.

A questão não é: «Pode uma máquina pensar?», ou «Pode um artefacto pensar?». A questão é: «Pode um computador digital pensar?». Mas, mais uma vez, temos de ser muito cuidadosos em relação à maneira de interpretarmos a questão. De um ponto de vista matemático, qualquer coisa se pode descrever *como se fosse* um computador digital. E isso é porque ele pode descrever-se como ilustrando ou levando a cabo um programa de computador. Num sentido extremamente trivial, a caneta que está diante de mim, em cima da secretária, pode descrever-se como um computador digital. Por acaso, tem até um programa de computador muito chato. O programa diz: «Permaneça aí.» Ora, visto que nesse sentido qualquer coisa é um computador digital, porque qualquer coisa pode descrever-se como realizando um programa de computador, então, mais uma vez, a nossa questão obtém uma resposta trivial. Naturalmente, os nossos cérebros são computadores digitais, porque rea-

lizam qualquer número de programas de computador. E, naturalmente, os nossos cérebros podem pensar. Assim, de novo, existe uma resposta trivial para a questão. Mas não era esta efectivamente a pergunta que tentávamos fazer. A questão que quisemos fazer é esta: «pode o computador digital, tal como foi definido, pensar?» Isto é, o ilustrar ou realizar o correcto programa de computador com as entradas e saídas correctas é suficiente para ou constitutivo do pensamento?» E para esta questão, diferentemente do que acontecia com as suas predecessoras, a resposta é claramente «Não». E é «Não» pela razão que antes indicámos, a saber, o programa de computador define-se apenas em termos sintáticos. Mas pensar é mais do que apenas uma questão de eu manipular símbolos sem significado; implica conteúdos semânticos significativos. Estes conteúdos semânticos são aquilo que nós indicamos por «significado».

É importante sublinhar mais uma vez que não estamos a falar acerca de um estágio particular da tecnologia dos computadores. O argumento nada tem a ver com os avanços futuros espantosos na ciência dos computadores. Nada tem a ver com a distinção entre processos seriais e paralelos ou com o tamanho dos programas ou a velocidade das operações do computador, ou com computadores que podem interagir causalmente com o seu ambiente, ou mesmo com a invenção de robôs. O progresso tecnológico é sempre grosseiramente exagerado, mas, mesmo eliminando o exagero, o desenvolvimento dos computadores foi muito notável e podemos sensatamente esperar que, no futuro, ainda se farão progressos mais notáveis. Sem dúvida, estaremos muito mais capacitados para simular o comportamento humano em computadores do que o podemos fazer agora e certamente muito melhor do que o conseguimos fazer no passado. O que eu quero realçar é que, se estamos a falar da existência de estados mentais ou de uma mente, todas essas simulações são simplesmente irrelevantes. Não inte-

ressa a boa qualidade da tecnologia ou a rapidez com que os cálculos são feitos pelo computador. Se é realmente um computador, as suas operações têm de definir-se sintaticamente, ao passo que a consciência, os pensamentos, os sentimentos, as emoções e tudo o resto implicam mais do que uma sintaxe. O computador é, por definição, incapaz de *duplicar* essas características, por mais poderosa que possa ser a sua habilidade em *simular*. A distinção essencial tem aqui lugar entre duplicação e simulação. E nenhuma simulação, por si mesma, alguma vez constitui a duplicação.

O que até agora fiz foi dar uma base à opinião de que aquelas citações com que comecei esta conferência são realmente tão absurdas como parecem. Há, no entanto, uma questão intrigante nesta discussão e é: «Porque é que alguém terá cogitado que os computadores podem pensar ou ter sentimentos e emoções e tudo o mais.» Ao fim e ao cabo, podemos fazer simulações de computador de qualquer processo do qual se possa fornecer uma descrição formal. Assim, podemos fazer uma simulação por computador da circulação do dinheiro na economia britânica ou do modelo de distribuição do poder no Partido Trabalhista. Podemos fazer a simulação por computador das chuvadas nos contados à volta de Londres ou dos incêndios de armazéns no Leste de Londres. Ora, em cada um destes casos, ninguém supõe que a simulação por computador é efectivamente uma coisa real; ninguém supõe que a simulação por computador de uma tempestade, de uma trovoada, nos deixará todos molhados, ou que a simulação por computador de um incêndio vai queimar toda a nossa casa. Por que diabo alguém no seu completo juízo, havia de supor que a simulação por computador dos processos mentais teria efectivamente processos mentais? Por meu lado, não sei que hei-de responder a isto, uma vez que a ideia, para falar com franqueza, me parece inteiramente louca desde o início. Mas posso fazer algumas especulações.

Em primeiro lugar, no tocante à mente, há ainda muitas pessoas tentadas por alguma espécie de behaviorismo. Pensam que, se um sistema se comporta como se entendesse chinês, então ele deve efectivamente entender chinês. Mas já refutámos esta forma de behaviorismo com o argumento do quarto chinês. Outra suposição feita por muitas pessoas é a de que a mente não é uma parte do mundo biológico, não é uma parte do mundo da natureza. É nisto que se baseia. A concepção da inteligência artificial forte, na sua concepção de que a mente é puramente formal e que, de algum ou outro modo, não pode ser tratada como um produto concreto de processos biológicos, como qualquer outro produto biológico. Há nestas discussões, em suma, uma espécie de dualismo residual. Os partidários da IA crêem que a mente é mais do que uma parte do mundo biológico natural; pensam que a mente é especificável em termos puramente formais. O paradoxo de tudo isto é que a literatura da IA está cheia de fulminações contra a concepção chamada «dualismo», mas, na realidade, toda a tese da IA forte se baseia numa espécie de dualismo. Funda-se numa rejeição da ideia de que a mente é justamente um fenómeno biológico natural do mundo, como qualquer outro.

Quero concluir este capítulo apresentando juntamente as teses do primeiro e do segundo capítulos. Ambas as teses se podem enunciar de um modo muito simples. E, na realidade, vou enunciá-las com uma crueza talvez excessiva. Mas, se as pusermos conjuntamente, penso que obtemos uma concepção bastante poderosa das relações entre mentes, cérebros e computadores. E o argumento tem uma estrutura lógica muito simples, de maneira que vocês podem ser se ele é válido ou inválido. A primeira premissa é:

1. *Os cérebros causam mentes.*

Ora, naturalmente, isto é de facto demasiado rude. O que queremos dizer é que os processos mentais que, na nossa consideração, constituem a mente são causados,

inteiramente causados, por processos que ocorrem dentro do cérebro. Mas sejamos rudes, abreviemos isto com três palavras — cérebros causam mentes. E isto é justamente facto sobre o modo como o Mundo actua. Escrevamos agora a proposição número dois:

2. *A sintaxe não é suficiente para a semântica.*

Esta proposição é uma verdade conceptual. Articula precisamente a nossa distinção entre a noção daquilo que é puramente formal e aquilo que tem conteúdo. Ora, a estas duas proposições, que os cérebros causam mentes e que a sintaxe não é suficiente para a semântica, acrescentemos uma terceira e uma quarta:

3. *Os programas de computador são inteiramente definidos pela sua estrutura formal ou sintática.*

Esta posição, na minha opinião, é verdadeira por definição; é parte do que significamos com a noção de um programa de computador.

4. *As mentes têm conteúdos mentais; especificamente, têm conteúdos semânticos.*

E isto é para mim um facto óbvio acerca do modo como as mentes agem. Os meus pensamentos, crenças e desejos são acerca de alguma coisa, ou referem-se a alguma coisa ou dizem respeito a estados de coisas no Mundo; e fazem isso porque o seu conteúdo os dirige para esses estados de coisas no Mundo. Ora, a partir destas quatro premissas, podemos tirar a nossa primeira conclusão; ela segue-se obviamente das premissas dois, três e quatro:

CONCLUSÃO 1. *Nenhum programa de computador é, por si mesmo, suficiente para dar uma mente a um sistema. Os programas, em suma, não são mentes e por si mesmos não chegam para ter mentes.*

Ora, esta é uma conclusão muito poderosa, porque significa que o projecto de tentar criar mentes unicamente

mediante o projectar programas está condenado, desde o início. E é importante tornar a sublinhar que isto nada tem a ver com qualquer estado particular da tecnologia ou qualquer estado particular da complexidade do programa. É um resultado puramente formal ou lógico, a partir de um conjunto de axiomas que são aceites por todos (ou quase por todos) os disputantes em questão. Isto é, mesmo a maioria dos mais violentos entusiastas da Inteligência Artificial reconhece que, de facto, enquanto assunto de biologia, os processos cerebrais causam estados mentais e reconhecem também que os programas se definem de um modo puramente formal. Mas, se se juntarem todas estas conclusões com algumas outras coisas que conhecemos, então, segue-se imediatamente que o projecto de IA forte é incapaz de realização e de cumprimento.

No entanto, já que obtivemos estes axiomas, vejamos o que é que podemos ainda derivar mais. Eis uma segunda conclusão:

CONCLUSÃO 2. *A maneira como as funções cerebrais causam mentes não pode ser apenas em virtude da activação de um programa de computador.*

E esta segunda conclusão segue-se da junção da primeira premissa com a nossa primeira conclusão. Isto é, a partir do facto de que cérebros causam mentes e que os programas não são suficientes para esse trabalho, segue-se que a maneira como os cérebros causam mentes não pode ter lugar apenas em virtude da activação de um programa de computador. Ora, também considero isto um resultado importante, porque tem a consequência de que o cérebro não é ou, pelo menos, não é justamente um computador digital. Vimos antes que qualquer coisa se pode descrever, de um modo trivial, como se fosse um computador digital, e os cérebros não são excepção. Mas a importância desta conclusão é que as propriedades

computacionais do cérebro não são simplesmente suficientes para explicar o seu funcionamento para produzir estados mentais. E, efectivamente, isto deveria parecer-nos, de qualquer maneira, uma conclusão científica bastante trivial porque tudo o que faz é lembrar-nos o facto de que os cérebros são máquinas biológicas; a sua biologia tem importância. Não é, como vários praticantes da Inteligência Artificial afirmaram, um facto irrelevante que a mente se realize em cérebros humanos.

A partir da nossa primeira premissa, podemos agora também derivar uma terceira conclusão:

CONCLUSÃO 3. *Tudo o mais que causou mentes deveria ter poderes causais, pelo menos, equivalentes aos do cérebro.*

E esta terceira conclusão é uma consequência trivial da nossa primeira premissa. É um pouco como dizer que, se o meu motor a gasolina impele o carro a cem quilómetros por hora, então, qualquer motor a *diesel* que fosse capaz de fazer o mesmo deveria ter também uma saída energética, pelo menos, equivalente à do meu motor a gasolina. Decerto, algum outro sistema poderá causar processos mentais utilizando características químicas ou bioquímicas inteiramente diferentes das que o cérebro efectivamente usa. Pode ser que venha a descobrir-se que, noutros planetas ou noutros sistemas solares, existem seres com estados mentais que utilizam uma bioquímica inteiramente diversa da nossa. Suponhamos que os marcianos chegaram à Terra e concluímos que eles têm estados mentais. Mas suponhamos que, quando as suas cabeças fossem abertas, se descobria que tudo o que têm dentro era apenas lama verde. Pois bem, mesmo então a lama verde, se funcionasse de maneira a produzir consciência e tudo o mais que é característico da vida mental, deveria ter poderes causais iguais aos do cérebro humano. Mas agora, da nossa primeira conclusão, de que os programas não são sufi-

cientes, e da nossa terceira conclusão, de que qualquer outro sistema deveria ter poderes causais iguais ao cérebro, segue-se imediatamente a:

CONCLUSÃO 4. *Para qualquer artefacto que pudéssemos construir, o qual tivesse estados mentais equivalentes aos estados mentais humanos, a realização de um programa de computador não seria por si só suficiente. Antes, o artefacto deveria ter poderes equivalentes aos poderes do cérebro humano.*

O resultado desta discussão é, creio, lembrar-nos de algo que já sabemos há muito: a saber, os estados mentais são fenómenos biológicos. A consciência, a intencionalidade, a subjectividade e a causação mental fazem todos parte da nossa história vital biológica, juntamente com o crescimento, a reprodução, a secreção da bÍlis e a digestão.