

IA – Inteligência Artificial ou Imbecilidade Automática? As máquinas podem pensar e sentir?

Valdemar W. Setzer *

Sumário

A visão padrão científica atual do mundo é materialista: sua hipótese fundamental é que o universo é puramente físico, e todos os seus fenômenos devem-se tão somente a processos físicos. Este artigo assume a hipótese de trabalho oposta: há também processos não-físicos. Isso é proposto como uma ampliação da visão científica atual, sem recair em um pensamento religioso, no misticismo, em fé, dogmas ou sectarismo. Evidências são dadas como suporte dessa visão ampliada. Algumas delas são universais, tais como a origem da matéria e da energia; outras são pessoais, tais como a unidade da percepção sensorial, a dependência da percepção em associar o que é percebido a conceitos (considerados como "objetos" não físicos no mundo platônico das idéias), a possibilidade de autodeterminar o próximo pensamento e o livre arbítrio, a subjetividade de sentimentos e particularidades da memória humana. A mente e suas atividades interiores de pensar, sentir e querer são examinadas utilizando tanto suas características gerais como a hipótese de trabalho fundamental, expondo porque elas não podem ser totalmente inseridas em uma máquina. O artigo também abrange o papel essencial do cérebro físico no processo do pensar e a dependência da consciência em relação ao pensar e sentir, com ênfase na autoconsciência. Novas classificações de inteligência e inteligência artificial são propostas. A "sala chinesa" de Searle, o teste de Turing, o xadrez eletrônico e as profecias de Kurzweil são descritos e discutidos. Dois filmes recentes nos quais os papéis de robôs são desempenhados por atores humanos são discutidos, e também suas influências sobre pessoas leigas e crianças. O artigo conclui que considerar seres humanos e seres vivos como máquinas representa um grande perigo para a humanidade e para o mundo, mostrando como a hipóteses fundamental pode ajudar a reverter essa tendência.

1. Introdução

O crescente poder de processar dados dos computadores modernos tem permitido a implementação de tarefas que teriam sido quase impossíveis dez anos atrás. Alguns exemplos foram a derrota do campeão mundial de xadrez pelo Deep Blue da IBM, reconhecimentos de voz, escrita e imagens, etc. Isso fez surgir questões sobre os limites dos computadores: Eles irão substituir toda atividade humana intelectual, talvez até mesmo manual? Vão revelar comportamento inteligente e substituir os seres humanos em tarefas criativas? Exercitarão o mesmo tipo de pensamento e sentimento que os seres humanos têm? Os robôs executarão todas as tarefas dos seres humanos? Eles tornar-se-ão

indistinguíveis dos seres humanos? Essas questões deixaram o domínio acadêmico com a exibição de dois filmes recentes, *The bicentennial man* (O Homem Bicentenário) e *Artificial Intelligence* (Inteligência Artificial).

Muito tem sido escrito sobre estas questões. O que faço aqui é introduzir uma maneira diferente de abordá-los. Muitos leitores acharão meus argumentos muito estranhos. Quero deixar bem claro que eles não são baseados em pensamento místico ou religioso. O leitor sem preconceitos reconhecerá que meus argumentos são conceituais e não emocionais, e são dirigidos ao entendimento comum.

Tenho uma recomendação ao leitor interessado. Quando encarar novas idéias uma pessoa deve tomar uma atitude e três ações: 1. Não ter preconceito, independente do quão estranhas sejam as idéias e as informações. Por exemplo, se alguém tivesse me dito antes de 11/9/2001 que o WTC em Nova York tinha acabado de ser atingido por dois aviões comerciais, eu teria pensado: "Isso parece estranho; jamais aconteceu antes no mundo. Mas irei investigar." Isso também significa estar completamente aberto a receber a informação, deixando críticas para um momento futuro, depois de ter formado uma noção do todo das idéias. 2. Verificar que as novas idéias são consistentes isto é, se não há contradições lógicas entre elas. 3. Verificar que as novas idéias não contradizem o que pode ser observado no mundo, fora e dentro do observador. Por exemplo, se as novas idéias incluem o fato de que algum objeto deixado no ar parado não cai, elas contradizem o que todos podem observar. Se alguém me diz que eu não posso decidir qual será meu próximo pensamento, isto contradiz minha própria experiência interior. 4. Verificar que as novas idéias parecem ser atrativas, isto é, correspondem a um anseio por novos conhecimentos na direção indicada por elas. Esta condição é muito importante, porque é impossível construir uma teoria que seja absolutamente abrangente e explique tudo. Mas isso requer que o leitor vá até o fim, para ver se o todo corresponde a um desejo íntimo por novas explicações.

Um comentário especial sobre o item 3 acima. É muito importante separar fatos científicos de julgamentos científicos. Por exemplo, é um fato ver-se o sol movendo-se pelo céu em um dia claro. A conclusão de que a terra mantém-se no mesmo lugar e que o sol se move em torno dela, ou que o sol mantém-se no mesmo lugar e a terra move-se em torno de seu próprio eixo, são julgamentos (o último corresponde a teorias e experiências científicas bem fortes).

Se os quatro itens são preenchidos, então a pessoa deve lançar mão das novas idéias como hipótese de trabalho e não como questões de fé ou de dogma. É isto que eu espero dos leitores que terão a coragem de ler minhas idéias até o fim com uma mente sem preconceitos.

A Inteligência Artificial (IA) é baseada em computadores modernos. Portanto, no cap. 2 descrevo o que é um computador, do ponto de vista lógico, caracterizando suas formas de processar dados e o fato de tratar-se de um processamento sintático e não semântico. O pensar é o ponto central quando falamos sobre inteligência e é abordado no cap. 3. Ligado ao pensar examino a intuição, a percepção sensorial, o papel que o cérebro pode ter no processo de pensar, o entender e aprender e a questão do cérebro ser ou não um computador. O cap. 4 trata da inteligência, expondo o que é geralmente entendido acerca

deste termo e os vários tipos de inteligência, bem como a minha própria classificação. A questão "as máquinas podem ser inteligentes?" é tratada no cap. 5. Nele descrevo o teste de Turing, bem como suas extensões e os tipos tradicionais e objetivos de pesquisas sobre Inteligência Artificial, abordando novos tipos originais. O cap. 6 trata da questão da possibilidade das máquinas terem sentimentos. Para isso, comparo sentimento com pensamento, mostrando que o primeiro é sempre subjetivo e individual e que o segundo pode ser objetivo e universal. O problema das máquinas terem consciência é tratado no cap. 7 e o cap. 8 trata da questão central dos seres humanos – e seres vivos em geral – serem ou não máquinas. O cap. 9 critica dois filmes recentes de sucesso que descrevem robôs tendo sentimentos, Inteligência Artificial e O Homem Bicentenário, e suas possíveis influências na maneira como as pessoas consideram a si próprias. Finalmente, no cap. 10 apresento minhas conclusões, abordando as preocupações que tenho com a atual visão de muitos cientistas que pesquisam IA de que seres humanos são máquinas e que, portanto podemos introduzir nas máquinas todas as capacidades humanas.

2. O que é um Computador?

Os computadores modernos digitais são máquinas matemáticas, lógico-simbólicas, algorítmicas. Isto significa que o processamento e o efeito de qualquer instrução interpretada em linguagem de máquina (rigorosamente, um computador nunca executa uma instrução, ele a interpreta) pode ser matematicamente descrito, quer dizer, representa uma função matemática. Além disso, a matemática envolvida é restrita: só trabalha com símbolos tirados de um conjunto finito, discreto, para ao qual sempre pode-se atribuir um sistema numérico.

Um programa é uma seqüência de instruções, podendo sempre ser associado a uma função matemática que leva elementos de um conjunto de dados de entrada em elementos de em registro de dados de saída. Defino dados como a representação de símbolos quantificados ou quantificáveis. Chamo algo de quantificável quando, se esse algo passa por um processo de quantificação e é, então, feita uma representação de sua quantificação, não é possível distinguir essa representação do objeto original. Por exemplo, se uma figura é varrida por um scanner e introduzida no computador, que então a imprime e a figura resultante parece a mesma que a original, digo que a tal figura é quantificável, porque dentro do computador todo objeto é representado usando um sistema numérico, isto é, por meio de quantidades. Outros exemplos de dados são textos, sons gravados e animação.

Eu disse que computadores são máquinas algorítmicas. Um algoritmo é uma seqüência finita de ações matematicamente bem definidas que terminam sua execução para qualquer conjunto de valores dos dados inseridos. Um programa de computador pode ser uma seqüência de instruções bem definidas, isto é, válidas. Porém, se esse programa entra durante sua execução em uma série infinita de execuções de ações, certamente com repetições (pois a seqüência de ações é finita), na qual não há nenhuma entrada de novos dados, então não é um algoritmo. Portanto, nem todo programa é a descrição de um algoritmo.

Um programa, é assim, uma seqüência de regras matemáticas sobre como transformar, transportar e armazenar dados. Dentro do computador, os dados são representados como

cadeias de símbolos quantificados. Assim, as regras podem ser consideradas como uma sintaxe que é aplicada a essas cadeias. As próprias cadeias sempre seguem uma certa estrutura. Por exemplo, uma cadeia representando um endereço pode ser composta de três partes, logradouro e número, cidade e CEP. O CEP deve seguir um certo padrão – uma cadeia de cinco dígitos decimais, seguida de um sinal - e de mais uma cadeia com 3 dígitos. Como programas e dados seguem regras sintáticas, pode-se dizer que um computador é uma máquina sintática.

2.1 Dados e Informações

Defini o que são dados. Caracterizo informação como uma abstração que existe apenas em alguma mente humana e tem significado para aquela pessoa (para maiores detalhes, ler meu artigo "Dado, informação, conhecimento e competência" em [Setzer 2002] e no meu site). Isto não é uma definição: não é possível definir o que são "abstração", "mente" e "significado". Um exemplo pode ajudar a entender a diferença entre dados e informações.

Suponha que tenhamos uma tabela com duas colunas representando nomes de cidades e suas temperaturas na data do dia anterior, como essas que saem nos jornais. Há uma linha de cabeçalho com os títulos, textos nas células da primeira coluna, e números nas células da segunda coluna. Suponha que os títulos e os textos estejam escritos em algum idioma que use símbolos especiais para seu alfabeto, por exemplo, o chinês. Para uma pessoa que não conheça o chinês e seus ideogramas, esta tabela nada mais é do que dados. Se não houver linhas, talvez a pessoa nem mesmo reconhecerá ser isto uma tabela. Isto não impede a pessoa de formatá-la, alterando as fontes dos ideogramas – uma ação de processamento de dados. Reconhecendo ser uma tabela, mas não entendendo o que significa e sendo dada uma ordem alfabética para os ideogramas, essa pessoa pode ordenar as linhas de acordo com a coluna de texto ou a coluna de números – também ações de processamento de dados. Todas estas ações seguem regras estruturais exatas, isto é, regras sintáticas.

Agora suponha que a pessoa entenda chinês. Neste caso, ela reconhecerá quais cidades estão sendo descritas e se está frio, morno ou quente em cada uma delas. Esta tabela tem significado para essa pessoa; ela atribui semântica ao conteúdo da tabela. Em minha conceituação, pode-se dizer que a pessoa incorporou os dados, a representação da tabela, como informação. Desse modo, um mesmo dado pode ser apenas um dado para uma pessoa, mas pode representar informação para outra.

Dada essa definição de dados e a essa caracterização de informações, podemos dizer que computadores são máquinas que processam dados. Não são máquinas que processam informações, porque eles não têm nenhum entendimento do que processam. John Searle desenvolveu uma interessante experiência mental para ilustrar este ponto.

2.2 A sala chinesa de Searle

Searle [1991, p. 32] descreve uma sala com uma pessoa, o operador. Muitos cestos com ideogramas chineses estão na sala, assim como um livro de regras, escrito em inglês, de como combinar os ideogramas chineses. A pessoa recebe por um guichê de entrada uma seqüência de ideogramas; utilizando o livro de regras, combina esses ideogramas de entrada

e alguns que estão nos cestos, compondo uma nova seqüência, que é então passada para fora da sala através de um guichê de saída. O operador não sabe o que está fazendo; na verdade, ele está respondendo perguntas em chinês. Searle argumenta que há uma distinção essencial entre esse operador e uma pessoa que leia e entenda chinês e responda perguntas sem utilizar um livro de regras. O primeiro está apenas seguindo regras sintáticas, mas o último está associando semântica ao que está fazendo. Searle afirma que a segunda pessoa está fazendo mais do que a primeira, porque entende o que cada pergunta e resposta significa. Ele diz, corretamente, que computadores são simplesmente máquinas sintáticas, combinando símbolos seguindo regras predeterminadas. Assim sendo, um computador pode substituir o operador daquela sala. Mas seres humanos fazem mais, eles associam significado, semântica, ao que eles observam e pensam. Como ele diz: "Há algo mais em ter uma mente do que executar processos formais ou sintáticos." [p. 31.] Conseqüentemente, computadores nunca poderão pensar, porque pensar envolve semântica. Programas não são suficientes para atribuir mentes a computadores. Infelizmente, ele toma significado e semântica de uma forma ingênua, e não elabora sobre esses conceitos. Não acredito na sua premissa: ele diz que "cérebros geram mentes" [p. 39], quer dizer, mentes são meros resultados, conseqüências de nossos cérebros físicos. Veremos que, uma vez que abandonemos esse ponto de vista, é possível elaborar mais no que pode ser entendimento, significado e semântica. O ponto importante agora é que essa premissa não invalida o argumento de sua sala chinesa. De acordo com esse argumento, computadores nunca poderão pensar. Pensar é, portanto, uma atividade central para determinar se as máquinas serão capazes de fazer tudo o que um humano pode fazer, inclusive ter inteligência.

3. O pensar

A atual visão científica do mundo afirma que há apenas processos químicos e físicos no universo, e que eles acontecem somente devido a leis físicas. Vamos chamar esta visão de "materialista". Searle representa esta visão muito bem. Ele diz:

"Suponha que façamos a pergunta que mencionei no começo: 'Uma máquina poderia pensar?' Bem, em um determinado sentido, claro, pois somos todos máquinas. Nós podemos encarar o que há dentro de nossas cabeças como uma máquina de carne. E, é claro, todos nós podemos pensar. Então, em um sentido de 'máquina', isto é, aquele em que máquina é apenas um sistema físico, capaz de executar certos tipos de operações, nesse sentido, todos nós somos máquinas e podemos pensar. Então, trivialmente, há máquinas que podem pensar." [p. 35, meu grifo.]

Há aqui um grande problema lingüístico. Quando dizemos "máquina", queremos nos referir a um dispositivo físico/químico que foi projetado e construído por seres humanos (eventualmente, utilizando outras máquinas ou produtos delas). Mas seres humanos – a propósito, também as plantas e animais – não foram projetados e construído por seres humanos! (Manipulação de DNA não é projetar nem mesmo construir um ser vivo completo.) Portanto, considero absolutamente incorreto afirmar "os seres humanos são máquinas"; seria mais apropriado dentro dessa linha de pensamento considerar os seres humanos simplesmente como seres físicos ("sistemas", para Searle). No cap. 8 eu darei razões explícitas para não considerar seres humanos como máquinas ou seres puramente

físicos. De qualquer forma, o que importa aqui é que está claro, a partir de suas palavras e premissa de que cérebros geram mentes, que Searle é essencialmente um materialista.

Um outro materialista é Antônio Damásio. Seu principal argumento é que cérebro mais corpo é a mesma coisa que mente, quer dizer, seres humanos têm apenas cérebros e corpos e não mentes separadas.

"O que eu estou sugerindo é que a mente surge de atividades nos circuitos nervosos..." [1994, p. 226]. "Entretanto, a minha visão de mente verdadeiramente corporificada, não renuncia a seus níveis mais altos de operação, aqueles que constituem a sua alma e espírito. Na minha perspectiva, acontece apenas que a alma e o espírito, com toda a sua dignidade e escala humana, são agora estados complexos e únicos de um organismo." [p. 252.]

John L. Pollock é absolutamente claro nessa posição:

"Meu propósito geral neste livro é defender o conceito do ser humano como uma máquina inteligente. Especificamente, vou argumentar que estados mentais são estados físicos e pessoas são objetos físicos. [1989, p. 1]

John Haugeland escreve:

"O objetivo fundamental desta pesquisa [Inteligência Artificial] não é meramente imitar inteligência ou produzir uma impostura esperta [clever fake]. De jeito algum. 'AI' quer apenas o artigo genuíno: máquinas com mentes, no sentido completo e literal. Isso não é ficção científica, mas ciência real, baseada em um conceito teórico tão temerário [daring] quanto o seu desafio: o de que somos, no fundo, computadores. Essa idéia – a idéia de que pensar e computar são radicalmente a mesma coisa – é o tema deste livro." [1987, p. 2, seus grifos].

Aqui está o que diz o evolucionista Richard Dawkins:

"Somos máquinas de sobrevivência – robôs cegamente programados para preservar as moléculas egoístas conhecidas como genes. Isso é uma verdade que me deixa cheio de admiração." [1989, p. 23; ver também pp. 55, 75, 105]. "O argumento deste livro é que nós, e todos os demais animais, somos máquinas criadas pelos nossos genes." [p. 29.]

Estes são apenas cinco exemplos; conjeturo que a maioria dos cientistas atuais, em todos os campos, segue a visão materialista do mundo; contudo, há honrosas exceções. Por exemplo, ninguém menos do que Roger Penrose admite a existência de um mundo platônico não físico das idéias matemáticas [1991, pp. 97, 428].

Alguém pode ouvir muitos cientistas dizendo: "Eu acredito em Deus". Não considero essa uma posição que caracteriza esta pessoa como não-materialista. O que importa é a maneira como a pessoa pensa. Se toda a sua argumentação está baseada em conceitos físico-químicos, então eu a classifico como materialista. Nessa linha, eu também considero muitos, provavelmente a maior parte das pessoas religiosas, como sendo materialistas. Observe-se apenas quantos deles dão muita importância à sua vestimenta especial ou cortes

de cabelo, ou ao lugar físico de culto, aspectos puramente materiais. Qualquer fundamentalismo religioso, com a típica falta de respeito físico a pessoas de fora da sua fé, com casos extremos de matar os hereges, só pode estar baseado no mais grosseiro materialismo.

Uma das conseqüências da visão materialista do mundo é que o pensar é considerado como uma conseqüência de processos nervosos dentro do cérebro, sendo por determinado por eles. É importante reconhecer o fato que o único conhecimento que temos na atualidade dos processos de pensamento é de que certas áreas do cérebro são mais ativas do que outras quando certos tipos de pensamentos, sentimentos, percepções ou lembranças são exercitados. Não fazemos idéia de como as lembranças são exatamente arquivadas, ou de como produzimos $2 + 3$ ou lembramos desse resultado. Assim, não há bases científicas sólidas para considerar o pensar como um processo gerado pelo cérebro.

3.1 Pontos de vista sobre o problema mente-máquina

Como já vimos, Searle é contrário à idéia de que computadores terão algum dia as capacidades plenas da mente. Seu maior argumento é que compreender é uma característica essencial das mentes; sua alegoria da sala chinesa mostrou que processamentos complexos são feitos por computadores de uma forma puramente sintática. Sendo máquinas sintáticas, computadores nunca terão mentes, pois estas têm capacidades semânticas.

Há outros pontos de vistas contra o fato de a mente ser uma máquina. Um deles é representado por James Fetzer, que baseia suas considerações na semiótica de Peirce, a teoria dos símbolos [2001, pp. 117, 137, 156]. Ele não concorda com o conceito de que a mente é um dispositivo computacional [pp. 125, 176].

Outro importante autor que considera que as mentes não são máquinas é Roger Penrose. Seu principal argumento é que "nosso pensamento consciente parece ser algo não-algorítmico." [1991, p. 412; ver também p. 439].

3.2 Meu ponto de vista acerca do pensar

Meu ponto de vista é: o pensar não é um processo físico, isto é, não pode ser reduzido a um processo puramente físico-químico. Há características especiais no processo de pensar que me levam a esta conclusão. Muitas delas foram levantadas pelo pensador Rudolf Steiner, na sua obra-prima A Filosofia da Liberdade [2000], inicialmente publicada no final do século 19, como extensão de sua tese de doutorado. Steiner foi o primeiro a atribuir à atividade de pensar a mais fundamental importância da vida humana moderna. Naquele livro ele faz uma profunda análise desse processo. Por exemplo, ele chama atenção para o fato de que o pensar é auto-reflexivo: é possível pensar sobre o pensar [p. 33]. Todas as nossas outras atividades envolvem outro objeto, como por exemplo a digestão: digerimos a comida e não a digestão em si. Vemos um objeto e não a visão em si, ou o próprio processo de visão. Em geral, não pensamos sobre o processo de pensar, isto é, não estamos conscientes do nosso pensar. Isso vem do fato de que nossos pensamentos são geralmente direcionados aos objetos que percebemos com nossos sentidos, ou estamos associando conceitos [p. 35].

O pensar é auto-sustentável: não é necessário usar nenhuma outra atividade interna quando pensamos sobre o pensar:

"Todas as outras coisas e todos os outros acontecimentos existem sem mim; não sei se existem verdade ou fantasia enganosa, ou bem como sonho. Só de uma única coisa sei com segurança incondicional, visto que eu mesmo a levo à sua existência segura: meu pensar." [p. 38.]

De acordo com Steiner, foi essa característica que levou Descartes a formular seu famoso "cogito, ergo sum":

"Para qualquer pessoa que possui a faculdade de observar o pensar ..., essa é a observação mais importante que ele pode fazer, pois observa algo que ela mesmo engendra; não se vê diante de um objeto que lhe é estranho, mas sim diante de sua própria atividade. Sabe, portanto, como se origina o que observa. Discerne com clareza as relações e as conexões. Encontrou-se, assim, um firme ponto de apoio no qual se pode basear a compreensão e a explicação das outras coisas." [p. 37.] "O enunciado mais simples que posso emitir sobre uma coisa é que ela existe." [p. 38, grifo do autor.]

Baseado na autoreflexão e na autosustentabilidade do pensar, formulei como autodeterminação outra de suas características únicas: o fato de podermos decidir qual será nosso próximo pensamento e exercê-lo de fato. Cada pessoa deve chegar a essa conclusão por experiência própria: não é possível prová-lo, porque ninguém pode saber o que outra pessoa está pensando. Para isso, eu proponho a seguinte experiência mental: sente-se em um lugar quieto e feche-se os olhos; tente-se produzir uma calma interior, quer dizer, sem ser atraído por percepções sensoriais externas, ignorando problemas urgentes e imergindo interiormente em si mesmo. Há um sentimento especial relacionado a esse estado interior de calma. Uma vez sentindo esta calma interior, imagine-se um mostrador no qual números vermelhos podem ser vistos. Agora imagine-se que o número 100 é exibido lá, depois o 99, o 98, etc. até o 0. O exercício consiste em se concentrar imaginando estes números e evitando qualquer outro pensamento. Um contra-exemplo poderia ser imaginar os números até 97, e de repente lembrar-se que em 1997 algo muito importante aconteceu na própria vida: houve, por exemplo, uma mudança de emprego. Daí talvez venha o pensamento de como se sentia mal no emprego anterior e como é satisfatório o emprego atual: o salário aumentou, então decidiu-se comprar a casa atual, onde houve espaço bastante para instalar o escritório em casa, com alguns quadros bonitos, etc. Podemos observar que o objetivo inicial, concentrar-se apenas naqueles números, desapareceu de há muito. Uma característica interessante desse exercício é que se pode repeti-lo com frequência, notando-se como a concentração mental melhora e com ela a habilidade geral de controlar o pensamento.

Neste ponto, a razão principal para tentar este exercício é reconhecer ser possível controlar, pelo menos por alguns segundos, o próprio pensamento, isto é, determinar o próximo pensamento. Então tornamo-nos conscientes do fato de que no pensamento temos autodeterminação. Não há nenhuma máquina, nem mesmo uma abstrata, que tenha esta característica autodeterminante. Máquinas inexoravelmente seguem seus programas ou mecanismos.

Máquinas abstratas são máquinas formais, matemáticas, digitais. Todas podem ser reduzidas a (ou simuladas por) uma máquina de Turing (MT), uma máquina que em cada instante está em um determinado estado, dentre um número finito de estados. Ela utiliza uma fita infinita dividida em células. Cada célula pode estar em branco, ou conter um símbolo de um alfabeto finito. A máquina tem uma cabeça de leitura e de gravação, que está sempre sobre uma única célula da fita. Uma instrução da máquina é executada da seguinte maneira: estando em um determinado estado, a cabeça lê o símbolo na célula da fita que está sob ela; baseada neste símbolo e no seu estado atual, a máquina grava outro (ou o mesmo) símbolo do alfabeto na célula que está sob ela; em seguida, a máquina move a cabeça (ou a fita) para a direita e para a esquerda e finalmente muda para outro estado. A seguir, uma nova instrução é executada exatamente com a mesma seqüência, até que a máquina chegue a um estado especial designado como estado final, ou uma instrução não possa ser executada, isto é, no estado atual não existe instrução que reconheça o símbolo da fita. De acordo com a chamada tese de Church-Turing, qualquer procedimento eficaz – inclusive o cálculo de funções matemáticas – pode ser executado por meio da máquina Turing. Qualquer computador real pode ser simulado por meio de uma MT, isto é, uma MT poder ser programada para aceitar um programa na linguagem de máquina do computador, e alguns dados de entrada, ambos gravados na fita, produzindo exatamente os mesmos dados de saída que o computador geraria para aqueles dados de entrada. Além do mais, uma MT pode imitar qualquer outra MT; esta é a razão pela qual ela é chamada de máquina universal.

Obviamente uma MT, ou qualquer outra máquina digital, seja abstrata ou concreta, seguem seu programa, ou seja, não é autodeterminada. Uma pessoa pode objetar que, quando fazemos o exercício do mostrador de números, também estamos seguindo um programa. A resposta a esta questão é que não é possível dizer se temos um programa armazenado no cérebro –nem mesmo sabe-se como é "armazenado" no cérebro um número como o 2 –, e, o mais importante de tudo, nós não sentimos como se estivéssemos seguindo um programa predeterminado. Por exemplo, nossa sensação é de que decidimos executar o exercício de concentração mental e que em cada momento decidimos permanecer nele, determinando nosso próximo pensamento – pelo menos durante alguns momentos. Não há nada que nos force a seguir o modelo do exercício, porque não há absolutamente nenhuma necessidade de executá-lo.

Reconhecendo que somos capazes de determinar nosso próximo pensamento, devemos concluir que nosso pensamento não é executado por um sistema meramente físico, porque tal sistema teria que seguir seus princípios físicos, seguindo leis inexoráveis, levando ao determinismo ou, na melhor das hipóteses, a alguns efeitos aleatórios. Mas como vimos, podemos reconhecer que não estamos sempre determinados, mas podemos autodeterminar algumas de nossas atividades interiores; por outro lado, não temos a sensação de que nosso pensamento é sempre aleatório, pois se assim o fosse não seríamos capazes de escolher nosso próximo pensamento. Há então três possibilidades: ou o cérebro produz nossos pensamentos, mas não é uma máquina, ou o cérebro é uma máquina, mas não produz nossos pensamentos, ou o cérebro não é uma máquina e não produz nossos pensamentos. Na minha opinião, a última hipótese é a mais provável: considerando que nosso pensamento não é físico e não é gerado pelo cérebro, mas obviamente influencia a sua

atividade, o cérebro tem de ser influenciado por algum processo não-físico, portanto ele deve ser constituído de tal modo que as leis físicas devem em certos processos deixar lugar para a atuação do pensamento não-físico. Na seção 3.5, abordo essa questão do não-físico influenciar o físico, e exponho o porquê do cérebro ser necessário.

As três possibilidades contradizem a afirmação de Searle acerca do cérebro ser um sistema puramente físico.

Espero que o leitor tenha executado o exercício mental descrito, tendo-se convencido que no pensar podemos ser livres, pelo menos por alguns segundos. Esta sensação de liberdade vem precisamente do fato de reconhecermos nossa habilidade de determinar nosso próximo pensamento. Se nosso cérebro fosse uma máquina ele teria de ser sujeito a leis físicas. Mas as leis físicas são inexoráveis: elas não podem levar à liberdade.

A liberdade requer autoconsciência. Um bêbado não é livre, porque ele não está totalmente consciente do que está pensando e fazendo [Steiner 2000, p. 18]. Uma cuidadosa observação pode mostrar que animais têm consciência, mas não têm autoconsciência. Apenas seres humanos podem ser autoconscientes, devido à sua capacidade de pensar. Eles estão constantemente introduzindo novidades no mundo; animais e plantas apenas seguem seus "programas" interiores e condicionamentos do ambiente. Nenhuma abelha jamais parou para pensar que pudesse tentar uma forma diferente de colméia que não fosse hexagonal.

A essência de nosso presente questionamento é o que é o pensar. Vejamos como ele se manifesta.

3.3 Xadrez eletrônico e intuição

Em 1997, a máquina IBM Deep Blue (DB) ganhou um torneio de xadrez do campeão do mundo Kasparov (K). DB venceu dois jogos, K venceu 1 jogo, e houve três empates (um total de 3,5 contra 2,5 pontos). Muitas pessoas celebraram, dizendo que as máquinas haviam suplantado os seres humanos. Porém, se examinarmos as condições da partida, chegaremos a uma conclusão um tanto diferente. Para maiores detalhes, ver meu ensaio sobre xadrez e computador [Setzer 1999].

A primeira consideração é que xadrez é um jogo matemático. Os tabuleiros, a posição das figuras e as regras do jogo podem ser todas matematicamente descritas. Por exemplo, a descrição da posição das figuras usa um sistema matemático de coordenadas em duas dimensões.

DB era uma máquina altamente paralela, especializada em jogar xadrez, que podia testar 36 bilhões de movimentos no tempo atribuído de três minutos para cada um deles. Além disso, ela tinha jogos anteriores de Kasparov gravados na sua unidade de armazenamento, tentando fazer correspondência entre cada posição do tabuleiro com aqueles jogos, fazendo assim, em certas situações, movimentos adequados, que haviam sido planejados com antecedência. Usualmente, em jogos como o xadrez, se o próximo lance deve ser dado pelo computador, este para um dado estado do jogo, isto é, no caso, as posições das peças no

tabuleiro, todos os possíveis movimentos que a máquina pode fazer, levando cada um a um novo estado; para cada novo estado, são testados todos os possíveis movimentos que o oponente pode fazer, levando cada um a um novo estado; para cada um destes, são calculados os possíveis movimentos a serem produzidos pela máquina e assim por diante. Tais jogos podem ser descritos como uma "árvore" virtual com nós de onde saem ramos. Cada nó representa um estado do jogo. Num certo instante deste, o nó que faz o papel de raiz da árvore representa o estado atual e cada ramo que sai dele representa um possível movimento do jogo que um dos jogadores pode fazer. O nó no final de cada ramo representa o estado do jogo depois do movimento descrito pelo ramo, constituindo um novo nível da "árvore". Desse nó, surgem ramos, representando os possíveis movimentos do oponente em resposta ao movimento do primeiro jogador e assim por diante. Considerando-se, para simplificar, que em cada estado há um número fixo de n possíveis movimentos, essa árvore cresce exponencialmente, pois a partir daquele estado chega-se a n nós, no próximo nível haverá n^2 nós, no nível seguinte n^3 nós, etc. O problema é usar alguma estratégia para "podar" a "árvore", eliminando-se alguns ramos que representam alguns claros movimentos que levam a uma posição inferior, de forma que poucas combinações tenham que ser testadas, isto é, tenta-se reduzir o valor de n em cada nó. DB era tão poderosa que seus criadores decidiram não utilizar técnicas de "poda" e deixaram a máquina testar todos os movimentos possíveis. O raciocínio por trás dessa decisão era que a máquina era tão rápida, que seria melhor deixá-la testar cada combinação do que evitar ("podar") alguns movimentos que poderiam beneficiar alguém no futuro.

Quantos movimentos K testava mentalmente antes de mover uma peça? Talvez vinte, talvez cinqüenta. Como foi possível que uma máquina matemática, jogando um jogo matemático, sendo capaz de calcular busilhões de movimentos mais do que um ser humano, perder um jogo e obter apenas um empate nos três demais? A resposta é clara: K, como um grande mestre, não estava testando os possíveis movimentos. Ele não estava calculando. Estava usando sua intuição para rapidamente imaginar alguns bons movimentos e eventualmente testá-los mentalmente.

Intuição é uma centelha de pensamento, um insight – uma visão interior de uma situação que não é vista no mundo físico. Um dos aspectos interessantes da intuição é que seu processo não pode ser descrito. Ela corresponde a ter alguns pensamentos vindos de lugar nenhum. Neste sentido, a intuição é absolutamente anticientífica – no entendimento comum dessa palavra.

3.4 Percepção sensorial e pensamento

Todos podem notar que a percepção sensorial está sempre acompanhada de pensamento. Mas a conexão entre os dois é mais profunda do que a aparência pode mostrar. Eu gostaria de sugerir que o leitor fizesse uma experiência. Por favor, olhe para a porta de entrada de sua sala. O que percebe visualmente? Tente responder a essa pergunta antes de continuar até o próximo parágrafo (estou deixando algumas linhas em branco antes dele). Seria interessante escrever sua resposta antes de ler o que segue.

Quando eu faço palestras sobre o assunto deste artigo, aponto para a entrada da sala e peço à platéia para responder a essa pergunta. Invariavelmente, todos respondem que estão

percebendo uma porta. Repito a questão e pergunto se há alguém que não está percebendo uma porta. Ninguém levanta a mão. Então eu digo à platéia que infelizmente estão todos errados. Ninguém está percebendo uma porta; o que cada um vê são impulsos luminoso, provenientes de cores e diferenças de cores, talvez ainda a sensação de diferença de profundidade devido à visão estereoscópica. "Porta" é um conceito. Conceitos não podem ser vistos com órgãos sensoriais, porque eles não existem no mundo físico. Eles são percebidos com o pensamento. Foi Rudolf Steiner que chamou a atenção para o fato de o pensamento poder ser considerado como uma ponte, entre a nossa percepção interior (representação mental) de objetos externos no mundo físico e conceitos no mundo das idéias [2000, p. 71]. Mas estes não são o mundo físico. Isto é absolutamente claro no caso de conceitos matemáticos, como apontado por Roger Penrose [1989, p. 428]. Por exemplo, ninguém nunca viu uma circunferência perfeita (o local geométrico de todos os pontos equidistantes de um determinado ponto, o seu centro). Esse conceito é o mesmo para todas as pessoas; de fato, não depende de ninguém e é eterno. Ele reside no mundo platônico das idéias e é percebido por diferentes pessoas com a mesma objetividade que eles vêem um objeto no mundo físico. De fato, parece que a objetividade na percepção das idéias pode ser maior do que a percepção dos objetos no mundo físico porque, por exemplo, os órgãos sensoriais diferem de pessoa para pessoa (p. ex. meu dois olhos vêem cores levemente diferentes). Mas o conceito de uma circunferência perfeita é exatamente o mesmo para todos que adquiriram a habilidade de perceber conceitos matemáticos.

Muitas pessoas, certamente todas materialistas, diriam que conceitos não são realidades no mundo platônico das idéias, mas que estão "armazenados" ou são "gerados" em algum lugar no cérebro. Por exemplo, Ray Jackendoff coloca isso da seguinte maneira:

"... eu penso no significado de palavras como instâncias, em grande parte, em um subsistema particular de organização combinatória do cérebro ao qual chamo de estrutura conceitual." [1993, p. 54, seu grifo.]

Infelizmente para os materialistas, isso não é um fato científico, isto é uma especulação, porque eles não podem mostrar, como já referi, onde e como um simples conceito tal como "dois" é armazenado no cérebro. Há muitas evidências da existência daquele mundo platônico. Por exemplo, como puderam Darwin e Russel Wallace, que eram quase antípodas (este último viveu na Nova Zelândia), terem desenvolvido no mesmo espaço de tempo a idéia da seleção natural? Esta e outras "coincidências" podem ser explicadas pelo fato de que ambos estavam percebendo a mesma idéia no mundo dos conceitos. Rupert Sheldrake, reconhecendo tal fenômeno, usa a noção de "campo morfogenético", que penetra tudo, de átomos a organismos vivos [1987, pp. 12, 76]. Este campo seria, por exemplo, responsável por formas orgânicas, no que ele denominou "causa formativa" (formative causation) [pp. 13, 88, 116]. Aprecio seu esforço em desviar-se dos paradigmas tradicionais da ciência atual, mas não concordo com seu princípio básico que seu campo morfogenético é físico [p. 115], resultado da visão materialista do mundo expressa no livro citado.

Estou ciente do fato que os materialistas irão questionar que minha suposição para a existência de um mundo platônico de idéias "real" é também uma especulação. Felizmente, ninguém pode provar que o outro não está certo, do contrário os espiritualistas poderiam

provar a existência do mundo não-físico, ou os materialistas poderiam provar que o mundo não-físico não existe. Penso que isso envolve um mistério, ligado ao que eu chamo hipótese existencial fundamental: assumir ou a hipótese de que existe um mundo não-físico "real", ou a de que esse mundo não existe. Usei a palavra "felizmente" acima, porque nós estamos livres para escolher qualquer uma das duas; essa escolha parece-me ser a mais fundamental hipótese existencial que deveria ser feita por cada pessoa. Para mim, as evidências da primeira hipótese – muitas das quais são confirmadas pelas minhas próprias vivências interiores e cada um deveria examinar as suas próprias – são esmagadoras. Nesse sentido, é importante lembrar-se do que escrevi na introdução, considerando como uma pessoa deve encarar novas idéias.

De acordo com Steiner, o pensar pode ser considerado como um órgão para perceber aquilo que não é físico, o mundo platônico de conceitos ou idéias [2000, p. 95]. O pensar completa a percepção instantânea e parcial que temos dos objetos externos [p. 69], ligando-nos ao das Ding an sich de Kant, a coisa em si mesma, o seu noumenon, a essência do objeto percebido como oposta ao fenômeno observado. Kant escreveu que nós nunca podemos alcançar o Ding an sich de cada coisa, porque o nosso pensar é um processo mecânico, por isso é limitado:

"A crítica da razão pura ... não nos permite criar para nós mesmos um novo campo de objetos além dos que nos são apresentados como fenômenos e divagá-lo em palavras inteligíveis; não, ela nem mesmo nos permite um esforço para formar nem mesmo uma conceituação deles. ... Resta um modo de determinar o objeto como mero pensamento, que é realmente apenas uma forma lógica sem conteúdo que, contudo, parece-nos ser um modo da existência do objeto em si mesmo (noumenon), sem considerar a intuição, que é limitada pelos nossos sentidos." [Kant 1952, p. 107]

"... a razão especulativa não pode nunca ... ultrapassar os limites das experiências possíveis ... ela não deve tentar elevar-se sobre a esfera da experiência, além da qual nada permanece para nós a não ser o vazio fútil." [p. 209.]

Steiner desfez as limitações do pensar, chamando a atenção para o fato de que ele pode perceber a essência, o conceito de cada objeto, e que essa essência não está no mundo físico. Mas, para isso, é necessário admitir a hipótese de que há um mundo não-físico. Infelizmente, o materialismo impediu os cientistas e acadêmicos de formularem esta simples hipótese, que ampliaria enormemente a pesquisa científica, a interpretação da história, as ciências teóricas e aplicadas, e assim por diante. Muitos cientistas têm preconceitos – o que é contrário ao verdadeiro espírito científico. Outros temem que ao deixar a trilha materialista venham a perder sua objetividade e fiquem sujeitos à fé e a dogmas. Steiner mostrou que esse não é o caso. Seu monumental trabalho conceitual resultou em aplicações práticas que vêm sendo exercidas por mais de 85 anos em muitos aspectos da vida cotidiana (educação, terapia educacional, medicina, várias terapias, agricultura, arte, arquitetura, organização social, etc.). Vale a pena deixar de lado o preconceito materialista e tentar imergir em seu vasto trabalho e aplicações, de forma a melhor julgar a validade de suas idéias, principalmente em termos do que escrevi na introdução.

No mundo não-físico dos conceitos, alguns destes estão ligados a outros; nosso pensamento é também capaz de perceber esta conexão, através de associações mentais.

De volta à percepção sensorial, é importante notar que mesmo que nossos sentidos estejam transmitindo algo a nós, nós não percebemos nada se não formos capazes de associar a percepção a algum conceito. O fascinante livro sobre luz de Arthur Zajonc descreve o fato de que uma pessoa que nunca havia visto nada; uma vez operada, e com visão normal, não passa a ver os objetos [1995, pp. 3, 183]: esta pessoa tem que aprender a fazer a associação entre sua percepção interior e o conceito correspondente. Ele conta um caso publicado sobre uma pessoa chamada S.B., um homem de 50 anos de idade, cego desde os dez meses de idade, inteligente e independente, o qual trabalhava consertando botas. Em 9 de dezembro de 1958 e em 1º de janeiro de 1959 ele recebeu transplantes de córneas.

"Examinando-o cerca de um mês após suas operações, [os autores, os psicólogos] Gregory e Wallace perguntaram a ele sobre sua primeira experiência visual após a operação. S.B. respondeu dizendo que ele ouviu uma voz, a voz do cirurgião, vindo de diante dele e de um lado. Virando-se em direção ao som ele viu algo embaçado. Rostos, mesmo muito tempo após a operação, nunca foram fáceis de perceber, relatou S.B. A pesquisa de Gregory e Wallace com S.B. (e pesquisas similares anteriores e posteriores) deixaram claro que aprender a ver quando já se é adulto não é nada fácil.

Depois de sua alta do hospital, Gregory and Wallace levaram S.B. a um museu de ciência e tecnologia. S.B. tinha um interesse demorado em ferramentas e ficou claramente excitado com a probabilidade de ver o que, até então, ele só havia manuseado ou ouvido por meio de descrições. Eles levaram-no a um torno de fazer parafusos e pediram que ele dissesse o que era aquilo que estava à sua frente. Obviamente decepcionado, S.B. não pôde dizer nada. Reclamou que não podia ver o metal ser trabalhado. Então ele foi trazido para mais perto e deixaram-no tocar no torno. Ele correu seus dedos avidamente sobre o torno com seus olhos bem fechados. Então ele se distanciou um pouco e, abrindo seus olhos, declarou, 'Agora que eu o senti, posso vê-lo.'

No caso de S.B. o lento processo de aprender a ver continuou pelos próximos dois anos, até a sua morte." [p. 3]

Em minhas palestras sobre esse assunto, uso o conhecido exemplo da figura de um hexágono perfeito dividido pelas três diagonais em seis triângulos equiláteros. Mostro-a à platéia e pergunto o que todos estão vendo. Frequentemente as pessoas respondem "um hexágono", "seis triângulos", alguns reconhecem uma pirâmide vista de cima. Eu chamo atenção para o fato de que tudo isso são conceitos e que, uma vez feita a associação com o conceito por uma pessoa, ela pode facilmente ver a figura correspondente. Em geral, apenas aqueles que conhecem o "truque" dizem ver um cubo. Então, sobrepondo transparências, eu lhes mostro um dos possíveis cubos e então o outro. Peço-lhes que façam um exercício de olhar para a figura original do hexágono como as diagonais e mudar de um cubo para o outro. Fazendo este exercício, o leitor também notará um fenômeno bem peculiar. É necessário imaginar, por meio do pensamento, um dos cubos para poder "vê-lo"; em seguida, a pessoa deve fazer um esforço interior para imaginar o outro cubo e então poder vê-lo. Torna-se claro que há um processo mental de "chaveamento" entre um cubo e outro,

isto é, entre os conceitos de um cubo visto em duas diferentes posições. Um dos meus alunos ficou tão entusiasmado com isso que desenhou um grande hexágono e o colocou no teto sobre sua cama, de forma a poder repetir o exercício muitas vezes sempre que fosse deitar.

Zajonc também menciona o fato de que a percepção da perspectiva linear nas figuras dá uma impressão de realidade simplesmente devido a um efeito cultural (a perspectiva linear começou a ser estudada e amplamente utilizada no começo do século XV) [p. 64]. Eu uso o exemplo dos trilhos de uma estrada de ferro. Se um adulto desenha dois trilhos de trem, ele desenha duas linhas, levemente inclinadas em relação à vertical, não paralelas mas convergindo cada vez mais e mais; segmentos de linhas horizontais, diminuindo de comprimento, unem os trilhos, para dar a impressão de dormentes. A maior distância entre os trilhos fica geralmente na parte de baixo do desenho.

Por que a maioria dos adultos desenha trilhos dessa maneira, em perspectiva? Porque é assim que eles a vêem. É uma ilusão de óptica. A realidade é que os trilhos nunca se encontram; se alguém anda entre eles, vê que eles mantêm sempre a mesma distância. Para desenhar trilhos como linhas que convergem, é necessário abstrair-se da realidade, ou de sua imagem mental ideal e fixar-se na ilusão de óptica proveniente da percepção visual. Mas isso requer uma certa capacidade de abstração. Uma criança pequena que não seja precocemente intelectualizada (infelizmente, muito comum hoje em dia!), ou uma pessoa de uma tribo primitiva, não desenha trilhos convergentes, mas paralelos. Inversamente, muitas dessas pessoas não conseguem reconhecer, de um desenho em perspectiva, o que ele representa. Zajonc descreve uma experiência interessante de mostrar a pessoas primitivas na África o desenho de um elefante visto de cima [pp. 63-64]. Tanto adultos como crianças preferiam a figura da esquerda (elefante "dividido"), isto é, não aquela desenhada com a perspectiva adequada; esta era considerada como "saltando por aí perigosamente" ("jumping about dangerously"):

Até o século 15 era raro encontrar perspectiva linear em pinturas – veja-se por exemplo os ícones orientais ou as pinturas medievais com estradas, quartos, casas, alamedas com árvores, etc. A perspectiva linear requer a habilidade de associar a ilusão de óptica do desenho ao verdadeiro conceito da realidade sendo apresentada. Essa habilidade não é inata, ela é adquirida através da observação de figuras e desenhos utilizando a perspectiva, e também desenhando-os através da orientação de alguém que conheça a técnica. Em outras palavras, como já vimos com o cubo, isso requer desenvolver a habilidade de associar a percepção ao verdadeiro conceito.

Assim, vê-se que a percepção sensorial deve ser acompanhada por uma percepção mental dos conceitos verdadeiros, que são a essência do que está sendo percebido e isso é feito pelo nosso pensamento.

Nossa hipótese mais importante é que o pensar não é um processo físico – se fosse, seria impossível alcançar com ele o mundo não-físico das idéias. Mas, qual é, então, o papel do cérebro físico?

3.5 O papel do cérebro

O fato de pessoas que têm algum dano cerebral perderem certas capacidades mentais tem levado à conclusão de que o cérebro produz essas capacidades. A conhecida história de um jovem trabalhador que construía uma estrada de ferro no século 19 e teve sua cabeça atingida por uma barra de ferro, perdendo parte do cérebro, é talvez o caso mais famoso e estudado, descrito extensivamente em [Damasio 1994, p. 3]. Esse jovem alterou completamente seu comportamento social. Claramente, há alguma conexão entre as funções mentais e o cérebro.

Steiner usa uma interessante analogia para explicar essa conexão [1997, p. 139]. Suponha que uma pessoa está olhando seu rosto em um espelho. Por meio do espelho ela percebe a forma do mesmo. Sem uma imagem em um espelho, ninguém pode perceber como é sua própria face. Agora, quebre-se o espelho. Sua face ainda existe, mas agora não pode mais percebê-la. Similarmente, o cérebro funciona como um espelho para o pensamento. Devido a isto, percebemos nossos próprios pensamentos e podemos assim controlá-los e dirigi-los. Sem o cérebro, nós ainda pensamos, mas não estamos conscientes do que estamos pensando, e não podemos escolher o que pensar. A sabedoria da nossa linguagem natural revela um conhecimento antigo, intuitivo deste fato: "refletir" é sinônimo de "pensar".

De acordo com esse ponto de vista, a atividade cerebral é uma consequência e não uma causa. Se os cientistas fizessem essa hipótese, a pesquisa sobre a atividade cerebral seria enormemente expandida. De acordo com a minha hipótese, procurar pela origem do processo de pensamento no cérebro é como a da conhecida história de um bêbado procurando pelas suas chaves em uma área iluminada embaixo de um poste de luz e não aonde ele de fato as perdeu. Os cientistas estão usando os limitados postes de luz do materialismo e se recusam a olhar para outras áreas, construindo novos postes de luz e outros meios de investigação.

Agora a questão é: como pode uma atividade não-física como o pensar interagir com o mundo físico? Esta é a velha questão da mente vs. matéria.

Tenho duas explicações possíveis para esse efeito. Uma delas é que no cérebro deve haver vários elementos, talvez dentro dos neurônios, que estão em um equilíbrio instável. Em tal estado, uma quantidade infinitesimal de energia seria suficiente para produzir uma mudança de estado. Penrose formulou a hipótese de que os efeitos do quantum no nível atômico, em microtúbulos, podem ser responsáveis por este efeito [Kurzweil 1999, p. 117].

O físico Amit Goswami tenta explicar esta ação do não-físico sobre o físico utilizando a não-localidade da física quântica [ver, por exemplo, Goswami 1995]. Infelizmente, ele parece misturar os âmbitos físico e não-físico, utilizando para o último a mesma argumentação que ele usa para o primeiro. Mas talvez essa explicação fosse outra

possibilidade para o nosso problema. Há um outro campo onde a ação do não-físico sobre o físico talvez seja melhor entendida: o crescimento e a regeneração de tecidos vivos. Em uma determinada célula viva, há três possibilidades: ou ela permanece como está, servindo para a diferenciação do tecido; ou se subdivide (mitose ou meiose) ou morre (apoptose). Conjeturo que a decisão sobre qual destes caminhos tomar não gasta energia. O modelo não-físico seguido pelo tecido regula este processo. A existência de um modelo por trás de todos os seres vivos parece ser claro; observe-se, por exemplo, como um pinheiro regula o crescimento de seus galhos para preservar sua forma de cone. Outro exemplo é a simetria das nossas mãos, orelhas, etc., permanentemente preservada durante o crescimento e regeneração. Um caso extremo é a forma razoavelmente permanente da impressão digital, mesmo quando a pele é danificada ("Feridas tais como queimaduras superficiais, raspões, cortes e bolhas, que somente afetam a epiderme, não alteram a estrutura rígida e o modelo original é duplicado na nova pele." [E.Britannica 1966, Vol. 9, p. 277]) Para preservar a simetria, seria necessário a uma orelha enviar uma mensagem a outra dizendo o quanto cresceu em uma determinada direção e esperar que a outra faça quase o mesmo crescimento – o que não faz sentido. Uma explicação mais razoável é que o processo de crescimento de todas as partes dos seres humanos, das quais podemos reconhecer as formas, assim como o dos elementos simétricos do corpo, são regulados de fora, por seu modelo, quer dizer, por um conceito não-físico. Obviamente, este modelo não-físico interage com os componentes físicos, que também estão sujeitos às diferenças ambientais. Richard Lewontin chamou este terceiro fator de "desenvolvimento aleatório" (development noise) [2000, p. 36]; para mim o processo não é aleatório, mas controlado pelo modelo ativo, que é o mundo platônico das idéias.

3.6 A visão

O processo de visão dá mais indicações de que há processos não-físicos acontecendo durante a percepção e a cognição.

O olho divide o campo da visão em quatro partes na fóvea da retina, correspondendo aos quatro quadrantes; vamos chamá-los de a, b (lado direito da imagem), c e d (lado esquerdo). As partes a e b formadas por cada olho são combinadas no nervo óptico (o qual, segundo me recordo de ter lido, transmite sinais elétricos complexos, que não correspondem às imagens na forma elétrica), assim como c e d, cada par indo para um hemisfério cerebral, onde a e b (c e d) estão separados em regiões divididas pelo sulcus calcarinus. Além, disso, no córtex visual há uma separação de percepção visual, movimentos no espaço visual, e lembrança ótica que ativa diferentes áreas no hemisfério direito, bem como percepção de forma e de cor no hemisfério esquerdo [Rohen 2000, pp. 17, 19]. Como então uma pessoa vê apenas um campo de imagem, como uma imagem completa? Minha explicação é que a percepção por si só não é um processo físico. Em algum lugar ao longo do nervo óptico e no cérebro, processos não-físicos acontecem de modo que a sensação completa que uma pessoa tem de um objeto percebido é produzido por alguns componentes não-físicos.

3.7 Compreensão e o aprendizado

Supondo que o pensamento é um órgão não-físico para observar o mundo das idéias, é possível compreender o que significa compreensão. Tomemos o caso da percepção de um

certo objeto através de nossos sentidos. Nós compreendemos esse objeto se somos capazes de usar nosso pensamento e fazer a ponte entre a percepção e sua essência, seu "Ding an sich". Esse é o caso quando, por exemplo, olhamos um objeto à distância e não podemos distinguir o que ele é. Estamos conscientes de que nossa percepção não é boa o bastante para um entendimento completo. Chegamos mais perto e reconhecemos o que é o objeto, isto é, agora nossa percepção é clara o bastante para permitir façamos, por meio do pensamento, a associação entre a nossa percepção interior do objeto e sua essência.

Isto se aplica também para a percepção de fenômenos. Por exemplo, olhamos para uma árvore e, de repente, notamos que um de seus galhos está movendo-se, mas todos os outros galhos estão parados. Ficamos inquietos, porque reconhecemos que não há vento e, mesmo se houvesse, vários galhos iriam mover-se. De repente, vemos um pássaro voando da árvore; agora intuimos que o pássaro devia estar pousado no galho e, quando voou, fez com que o galho movesse. Esse exemplo também ilustra o fato de que a compreensão pode envolver a associação de várias percepções e pensamentos.

John Searle não foi capaz de caracterizar o significado da compreensão (cf. 2.2.). Conjeturo que, de um ponto de vista materialista, nunca será possível fazer esta caracterização.

É possível fazer, com o pensamento, associações incorretas entre percepções e conceitos que não correspondem à essência de um objeto, ou fazer associações mentais incorretas. Isto pode vir de percepções sensoriais imperfeitas, ou pensamento incorreto. O primeiro é geralmente claro: por exemplo, se vemos uma pessoa à distância e não podemos reconhecer se é um homem ou uma mulher, ficamos cientes de que nossos olhos não estão acurados o bastante para ver claramente à distância. Para uma pessoa com órgãos sensoriais saudáveis, eles são extremamente fiéis, e ela reconhece quando sua percepção não está clara. Do contrário, não poderíamos confiar em nossos sentidos, e seríamos esquizóides ou malucos.

O pensamento incorreto vem de não se examinar todas as possíveis associações e não escolher aquela com a evidência mais razoável. Esse foi o caso com o modelo planetário heliocêntrico. Todos vêem o sol e as estrelas movendo-se no céu. O julgamento de considerar a terra imóvel e o sol e as estrelas girando em volta dela, ou o sol e as estrelas imóveis e a terra girando em volta de seu eixo, é uma questão de julgamento, de pensamento correto, não de percepção. Esse exemplo ilustra um ponto muito importante: o pensamento correto depende de seu desenvolvimento. Por milênios os seres humanos não tiveram capacidade suficiente de abstração para se livrarem da forte impressão sensorial do sol e das estrelas movendo-se pelo céu. É importante observar que o modelo heliocêntrico já era aceito amplamente quando Newton estabeleceu, em 1687, uma teoria robusta, a da gravitação, para explicar o movimento dos planetas e suas órbitas elípticas em volta do sol. Mas isso aconteceu bem antes da primeira demonstração experimental desse fato com o pêndulo de Foucault em 1851. Até então, a aceitação do modelo heliocêntrico era uma questão de aceitar a teoria abstrata mais convincente. Uma fascinante consideração desse desenvolvimento da capacidade humana de pensar foi dada por Arthur Koestler [1964]. Ela mostra como a humanidade adquiriu a habilidade de chegar a conceitos corretos sobre o sistema planetário.

O aprendizado também pode ser entendido com esse modelo. Ele significa desenvolver a capacidade de perceber certos conceitos por meio do pensamento, fazendo as associações corretas entre os vários pensamentos que estão inerentemente conectados no mundo das idéias, ou fazer a ponte correta entre a percepção de um objeto ou fenômeno e seu conceito. Memorizar percepções e conceitos é obviamente também uma parte essencial do processo de aprendizado.

A propósito, uma pessoa nunca deveria dizer que um programa de computador "aprende". O que ele faz é calcular alguns parâmetros (como no caso das erroneamente chamadas "redes neurais"; a única similaridade entre elas e nossos neurônios é que estes também formam redes), ou armazenar alguns dados. Não há conhecimento científico de como nosso cérebro aprende algo; conjeturo que esse também é um processo não-físico. Isso não contradiz fatos científicos já conhecidos (mas certamente contradiz o julgamento da maioria dos cientistas que trabalham com cognição).

3.8 O cérebro é um computador?

Como eu já disse, não há base científica para afirmar que o cérebro funciona como um computador digital. Ao contrário, há muitas indicações de que ele não faz.

Aparelhos digitais apenas funcionam se os sinais que atravessam os vários circuitos estão sincronizados. Uma porta lógica não funciona se não houver sincronização dos pulsos elétricos que chegam a ela. Essa sincronia é obtida por um gerador central de pulso, o chamado "relógio" (clock); quando, por exemplo, é dito que um computador tem um relógio de 2 ghz, isto significa que seu gerador central de pulso gera dois bilhões de pulsos por segundo. Claramente, não há nenhum gerador central de pulso e sincronização de pulsos no cérebro. Recentemente, novos projetos de circuitos lógicos assíncronos, isto é, sem um relógio central, têm sido desenvolvidos (no momento, apenas pequenas partes de instrumentos utilizando circuitos digitais estão no mercado utilizando essa técnica):

"Sem um relógio para governar suas ações, um sistema assíncrono deve, em seu lugar, basear-se na coordenação dos circuitos locais. Esses circuitos trocam sinais de finalização para assegurar que as ações em cada estágio começam apenas quando os circuitos têm os dados de que precisam." [Sutherland 2002, p. 50]

Portanto, mesmo sem um gerador de pulso sincronizador, é necessário ter uma sincronia local para trocar sinais sincronizados, mas os neurônios não parecem seguir este padrão. Aparentemente, os neurônios podem disparar (isto é, produzir um sinal elétrico de saída nos seus axônios) ou não disparar nas mesmas circunstâncias ("O mesmo estímulo não produz sempre o mesmo resultado" [Penrose 1991, p. 396]). Então, como na maioria, senão na totalidade dos processos dos seres vivos, não há no cérebro um determinismo tal como nas máquinas digitais.

O determinismo é uma das características essenciais da potência dos computadores e os torna úteis. Seria um desastre, para um determinado programa e um determinado conjunto de dados de entrada, se o dado resultante fosse diferente cada vez que os primeiros forem processados. Se há processos não-deterministas, tais como aqueles que ocorrem em redes

de computadores (por exemplo, o trajeto que uma mensagem vai percorrer em uma rede complexa de computadores é em geral imprevisível), para cada máquina da rede tudo se passa como se houvesse determinismo.

O conexionismo (connectionism - ver, por exemplo, [Ramsey 1991]) introduziu a idéia de que o cérebro é um sistema altamente distribuído e é baseado nesse conceito que sua atividade é estabelecida entre os neurônios. Deixando de lado o fato de que as chamadas redes neurais não têm quase nada em comum com as redes formadas pelos neurônios (por exemplo, estas últimas não têm uma topologia fixa [Penrose 1991, p. 396], os parâmetros para combinar as entradas de dados de cada célula da rede neural são calculados através de algoritmos muito complexos, as entradas de dados têm de ser sincronizadas, etc.), há um forte argumento contra o modelo distribuído: nós não temos a impressão de que as nossas atividades mentais são separadas em várias partes [p. 398]. Ao contrário, como vimos no caso da visão, parece existir uma distribuição de funções pelo cérebro mas nossa percepção é de um todo. Atividades inconscientes ou semiconscientes podem ser feitas paralelamente. Assim, nós podemos fazer coisas diferentes com nossos membros, por exemplo nós podemos usar nossas mãos enquanto andamos. Mas o pensamento consciente está sempre concentrado em apenas um pensamento.

Assim, qualquer comparação entre o funcionamento do cérebro e uma máquina digital é absolutamente imprópria. Também considero isso perigoso, porque reduz a imagem que o ser humano faz de si mesmo à de uma máquina. O perigo vem do fato de não haver ética envolvendo nossa relação com as próprias máquinas, apenas como nós as usamos em relação a outras pessoas e ao mundo. Não há sentido em se ter compaixão para com uma máquina, por exemplo sentir pena de desligá-la. Infelizmente, as crianças têm sido condicionadas a terem tal sentimento; esse foi, por exemplo, o caso do terrível Tamagochi.

3.9 Crítica a Kurzweil

Ray Kurzweil é um dos expoentes da idéia de que seres humanos são máquinas, portanto as máquinas serão algum dia capazes de fazer qualquer coisa que um ser humano faz. Seu livro best-seller [1999] é cheio de profecias, baseadas na afirmação a seguir:

"O cérebro humano tem cerca de 100 bilhões de neurônios. Com uma média estimada de mil conexões entre cada neurônio e seus vizinhos, temos cerca de 100 trilhões de conexões, cada uma capaz de efetuar um cálculo simultâneo. É, sem dúvida um processamento paralelo maciço e uma chave para a potência do pensamento humano. Uma profunda fraqueza, contudo, é a excruciante baixa velocidade dos circuitos neurais, apenas 200 cálculos por segundo." [p. 103.]

Essa afirmação é absolutamente injustificada. Ele não diz que tipo de cálculos são feitos por cada conexão de neurônio e, como já mostramos, ele não pode nem dizer como os dados são armazenados no cérebro. Baseado no número acima, ele o multiplica pelas 100×10^{12} conexões existentes no cérebro, chegando à conclusão de que somos capazes de realizar 20×10^{15} "cálculos" por segundo. Ele nem sequer considera a possibilidade de que possa haver diferentes funções para diferentes conexões; para ele esta capacidade de

realizar cálculos é o fator mais importante. Ele usa o mesmo tipo de argumentação para chegar à conclusão de que a nossa memória tem 1015 bits.

No seu livro clássico, John von Neumann escreve: "um receptor padrão parece aceitar 14 impressões digitais distintas por segundo". Ele supõe que há "1010 células nervosas" cada uma trabalhando como "um receptor (interno ou externo)". Então, "assumindo além do mais que não há um verdadeiro esquecimento no sistema nervoso", e um tempo de vida normal de 60 anos ou cerca de 2×10^9 segundos, ele chega à conclusão que a capacidade da nossa memória é $2,8 \times 10^{20}$ bits [1958, p. 63].

É impressionante como pessoas tão brilhantes podem fazer esse tipo de cálculo, sem saber como nossa memória funciona, levando em conta que o nosso sistema nervoso não é uma máquina digital, etc. Contudo, vamos prosseguir com Kurzweil.

Extrapolando a crescente capacidade dos novos computadores ele chega à conclusão que em 2020 um computador de US\$ 1.000,00 terá a mesma capacidade de cálculo do cérebro humano, em 2023 a mesma capacidade de memória. Ele profetiza que em 2019 haverá "relatos de computadores passando o Teste Turing", e em 2029 terão de fato passado esse. Para criticar estas afirmações, vamos abordar a questão da inteligência e o Teste Turing.

4. Inteligência

Até agora temos falado sobre percepção e pensamento. Para responder à velha questão "as máquinas podem ser inteligentes?" temos que lidar com a questão do que é inteligência e o que significa considerar que uma máquina é inteligente.

4.1 O que é inteligência?

A resposta para essa velha questão depende do que se entende por "inteligência". Por exemplo, se alguém pensa que jogar uma boa partida de xadrez é demonstração de inteligência, então temos que atribuir inteligência aos computadores. Mas como nem toda pessoa inteligente joga xadrez, essa caracterização não é suficientemente ampla.

É importante perceber que não é possível formalmente definir o que é inteligência. A quarta edição do American Heritage Dictionary, tem as seguintes explicações para "inteligência":

- a. A capacidade de adquirir e aplicar conhecimento.
- b. A faculdade do pensar e da raciocinar.
- c. Poderes superiores da mente.

Percebe-se que essas caracterizações são um tanto vagas, a terceira é de quase nenhuma utilidade. O mesmo dicionário diz que o conhecimento, entre outras coisas, é "informação específica sobre algo" (ver meu artigo "Dado, informação, conhecimento e competência", no meu site, para a caracterização desses conceitos). Como já vimos na seção 2.1, vou fazer uma clara distinção entre dados e informações. Na minha conceituação, informação requer significado, compreensão. Computadores não processam informação, mas apenas dados.

Nesse sentido, computadores não podem ser inteligentes. Mas vamos estender esta caracterização de inteligência, cobrindo suas várias manifestações.

4.2 Tipos de inteligência

Howard Gardner foi o pioneiro em antecipar o que ele chamou de "visão pluralista da mente" [1995, p. 13] e classificar vários tipos de inteligência. Ele inicialmente reconhece sete tipos [pp. 15, 22]:

- a) Lingüística
- b) Lógico-matemática
- c) Espacial
- d) Musical
- e) Corporal-cinestésica
- f) Interpessoal
- g) Intrapessoal

Ele diz que normalmente a educação e as provas realizadas nas escolas enfatizam apenas as duas primeiras [p. 107], mas todas elas são importantes para uma vida saudável e frutífera. (c) tem a ver com "a capacidade de formar um modelo mental de um mundo espacial, e de ser capaz de manobrar e operar utilizando esse modelo", e é típica dos marinheiros, engenheiros, cirurgiões, escultores e pintores. (e) "é a capacidade de resolver problemas e ou de elaborar produtos utilizando o corpo inteiro ou partes do corpo" e é típica de dançarinos, atletas, cirurgiões, artesãos e artistas. (f) "é a capacidade de compreender outras pessoas: o que as motiva, como elas trabalham, como trabalhar cooperativamente com elas" e é típica de vendedores, políticos, professores, terapeutas e líderes religiosos bem sucedidos. (g) "é a capacidade de formar um modelo acurado e verídico de si mesmo e de utilizar esse modelo para operar efetivamente na vida".

Depois, ele expandiu os sete tipos para vinte [Goleman 1995, p. 51], por exemplo subdividindo (f) em liderança, manutenção de relacionamentos sociais e preservação de amigos, resolução de conflitos e ser capaz de fazer análise social.

Daniel Goleman concentrou-se na inteligência interpessoal de Gardner, chamando-a "inteligência emocional" [1995]. Ele diz que este tipo de inteligência é o mais importante na vida social e profissional [p. 165].

Agora vou dar a minha própria classificação.

4.3 Inteligências incorporada e criativa

Classifico inteligência em dois tipos: incorporada e criativa.

Se examinarmos nosso corpo podemos perceber uma infinita inteligência, uma infinita sabedoria. Na verdade, considero o corpo humano a coisa mais maravilhosa do mundo físico, sua maravilha primordial. Parece-me que é impossível imaginar algo fisicamente superior ao nosso corpo, considerado como um todo. Chamo isso de inteligência

incorporada. Mas nós não somos os únicos que possuem tal inteligência. Os animais também a têm, não apenas em seus corpos, mas também nas suas habilidades físicas inatas. (Provavelmente as únicas habilidades inatas que temos é chorar e mamar, talvez até rolar e engatinhar – todas as demais têm de ser conquistadas através de relações sociais, tais como andar, falar e pensar.)

As plantas também têm inteligência incorporada. A maneira como nascem, crescem e se reproduzem revelam uma enorme inteligência. Mas também considero que os minerais têm este tipo de inteligência. Se a Terra não fosse do jeito que é, a vida não seria possível. Por exemplo, se o nosso planeta fosse mais próximo ou mais distante do Sol, ou o seu eixo não tivesse a inclinação que tem em relação à eclíptica, nós não estaríamos aqui. Nós também não estaríamos aqui se a água não tivesse sua mais alta densidade em quatro graus centígrados, se a atmosfera não tivesse seu atual equilíbrio entre oxigênio e nitrogênio, se o solo não fosse como é, adequado para o mundo das plantas, se as constantes físicas não fossem como o são, etc.

As máquinas também têm muita inteligência incorporada: parte da inteligência de seus projetistas, manifestada através de sua forma e funcionalidade, assim como parte da inteligência de seus construtores. Em particular, um programador de computador incorporou parte de sua inteligência ao programa que desenvolveu. Em outras palavras, qualquer programa de computador revela uma inteligência incorporada.

Meu segundo tipo de inteligência é o que chamo inteligência criativa. Para caracterizá-la, tenho que abordar o que é criatividade. Para isso vou utilizar uma caracterização que ouvi em uma conferência do sociólogo italiano Domenico di Masi. Ele disse que a criatividade é composta de fantasia e concretividade (transliterando do seu original em italiano). Entendo fantasia como a habilidade de ter novas idéias; elas podem ser absurdas, "castelos no ar", ou podem corresponder à realidade ou a alguma verdade. Di Masi caracterizou concretividade como a habilidade de produzir algo concreto, ou fazer algo social ou pessoalmente útil. Com humor, ele disse que uma pessoa que fica fantasiando sem realizar nada de útil com suas idéias é um diletante. No sentido oposto, uma pessoa que fica produzindo coisas ou ações sem nenhuma fantasia é um burocrata.

Portanto, inteligência criativa é o tipo de inteligência que envolve ter novas idéias e fazer algo útil com elas.

Apenas seres humanos têm inteligência criativa. Nenhum outro ser vivo a tem. Os animais seguem seus instintos ou condicionamentos, que constituem como que um "programa". Seus "programas" podem mudar sob a influência do meio ambiente. Mas em um determinado momento, um animal está automaticamente seguindo seus impulsos resultantes dos instintos e condicionamentos. O ser humano é capaz de reconhecer uma força interior levando-o a fazer algo, mas ele pode pensar sobre as conseqüências de agir de acordo com aquela força, e simplesmente não realizar seus impulsos. Por exemplo, uma pessoa pode reconhecer que está um tanto gorda, decidindo fazer uma dieta com propósitos estéticos. Ela imagina-se mais magra, mais elegante. Seus instintos de fome vão levá-la a comer, mas ela pode refrear-se de fazê-lo, apenas para perder peso. Nenhum animal faz

dieta com finalidades estéticas. Observando o mundo, nós podemos ver que os seres humanos, não os animais, estão mudando-o (infelizmente, em geral para pior).

Computadores não podem fantasiar. Eles simplesmente geram novos dados a partir de dados previamente fornecidos ou calculados, seguindo algum programa, em um processo combinatório. Não considero que seres humanos têm novas idéias apenas como combinação de fatos ou idéias previamente conhecidos. Talvez o zíper seja um exemplo desse fenômeno. Se nos ativermos apenas às conquistas técnicas, provavelmente o transistor e o laser também foram idéias completamente novas, porque eles foram teoricamente planejados antes de suas construções. Situações sociais apresentam muitas idéias novas que não estão baseadas em experiência prévia.

Por outro lado, a concretividade requer bom senso. Esta é outra capacidade que os computadores não podem ter, pois ela transcende meras abstrações e teorias.

5. As máquinas podem ser inteligentes?

Vou abordar o que algumas pessoas disseram sobre este assunto, e então farei minha própria análise.

5.1 O Teste Turing e extensões

Alan Turing foi o primeiro a caracterizar em que situação um computador poderia ser chamado "inteligente". Em 1950, ele planejou o que se tornou conhecido como o famoso Teste Turing (TT) [1950]. Como ponto de partida, usou um jogo social envolvendo três pessoas que estão em salas diferentes: um homem (H), uma mulher (M) e um questionador (Q) que faz perguntas aos dois primeiros. Uma quarta pessoa pode ser incluída no jogo, transmitindo perguntas formuladas por Q direcionadas a H ou a M, e retornando a Q as respostas dadas; alternativamente, ele recomenda o uso de um "teletipo" (máquina que era usada para transmissão imediata de mensagens). O jogo consiste em Q descobrir quem é a mulher e quem é o homem. Além disso, "É objetivo de H no jogo tentar levar Q a uma identificação incorreta. O objetivo do jogo para M é ajudar o questionador" [p. 434] (portanto, para Turing as mulheres eram sempre honestas...).

No que ele chamou de "jogo da imitação", Turing substituiu H por um computador e M por uma pessoa de qualquer sexo. Em suas próprias palavras:

"Fazemos agora a pergunta, 'O que acontecerá quando uma máquina fizer o papel de H neste jogo?' Será que o questionador decidirá de forma errada com tanta frequência quando jogamos dessa forma como ele faz quando o jogo é com um homem e uma mulher? Estas questões substituem nossa original, 'As máquinas podem pensar?' ... O novo problema tem a vantagem de estabelecer uma linha divisória bastante precisa entre as capacidades físicas e intelectuais de um ser humano." [p. 434.]

Desse modo, Turing substitui a pergunta sobre a possibilidade das máquinas serem inteligentes pela questão "As máquinas podem pensar?" e então substitui esta última pelo seu "jogo de imitação", o TT. É interessante notar que, além do título do artigo, encontrei

nele apenas um único trecho onde ele menciona a palavra "inteligente" e um outro para "inteligência".

"Comportamento inteligente consiste presumivelmente em um desvio do comportamento completamente disciplinado envolvido em computação, mas apenas levemente, para um que não gera um comportamento aleatório ou repetições sem objetivo." [p. 459.]

Sua convicção era que a computação seria o bastante para um comportamento inteligente, porque ele estava certo que por meio dela as máquinas poderiam algum dia passar seu teste:

"Acredito que em cerca de 50 anos será possível programar um computador com uma capacidade de armazenamento de cerca de 10⁹ para fazê-lo jogar o jogo da imitação tão bem que um questionador mediano não terá mais de 70 por cento de chance de fazer uma identificação correta depois de 5 minutos de perguntas. Acredito que a questão original, 'Uma máquina pode pensar?' é demasiadamente sem sentido para merecer ser discutida. Contudo, acredito que no final deste século o uso das palavras e a opinião erudita geral terão sido tão alteradas que será possível a alguém falar sobre máquinas que pensam sem esperar ser rebatido." [p. 442.] (Eu estou aqui para rebater!)

Nem todas as suas profecias tornaram-se realidade. Na verdade, temos agora computadores com aquela capacidade de armazenamento. Mas não há computador próximo de passar em seu teste. Há um concurso do TT anual na Universidade de Manchester, e os melhores programas apresentados estão muitos distantes de passar no teste: com poucas perguntas é possível detectar qual participante é uma máquina. Pode-se até duvidar que tal programa chegará mesmo a existir. Kurzweil não tem absolutamente nenhuma base sólida para fazer sua profecia sobre o ano 2029 (cf. seção 3.9).

Turing teve uma intuição de que nosso pensamento é uma atividade interior muito especial, e que seria eventualmente impossível descrever seu processo cientificamente:

"Será que as máquinas não poderiam realizar algo que deveria ser descrito como pensar mas que é muito diferente do que um ser humano faz? Esta objeção é muito forte, mas ao menos podemos dizer que se, contudo, uma máquina puder ser construída para jogar o jogo da imitação satisfatoriamente, nós não precisamos nos preocupar com essa objeção." [p. 435.]

Suas esperanças não diziam respeito apenas ao seu teste:

"Devemos esperar que as máquinas eventualmente venham a competir com os seres humanos em todos os campos puramente intelectuais." [p. 460.]

Isto já é uma verdade para muitos campos, tais como jogar xadrez, reconhecer textos escritos e falados, provar certos tipos de teoremas, etc. Turing menciona que, para muitas pessoas, jogar xadrez seria o melhor ponto de partida. Mas ele não pára por aí:

"Também pode ser dito que o melhor é prover a máquina com os melhores órgãos sensoriais que o dinheiro pode comprar, e então ensiná-la a entender e falar inglês. Esse processo poderia seguir o ensino normal de uma criança. Coisas seriam apontadas e nomeadas, etc." [p. 460.]

É impressionante que Turing não reconheça que o que um computador faria em tais casos seria sintaticamente associar um grupo de letras a uma imagem armazenada do objeto nomeado. Quando muito, essa associação poderia ser feita por uma descrição de características comuns de imagens similares. Como vimos, de acordo com minha hipótese de trabalho, o que nós fazemos é associar um nome à essência de um objeto e não a uma imagem particular, grupos de imagens ou a uma descrição objetiva de características comuns. Outra objeção é que quando nos lembramos da imagem de um objeto já vivenciado, fazemos uma representação mental que nunca é tão clara quanto a percepção do objeto quando o observamos – exceto no caso de figuras geométricas, nas quais a representação em desenho é muito próxima à sua essência. Em geral, não nos prendemos aos detalhes, porque o que importa é a essência. O fato de que os computadores podem armazenar e trabalhar com a figura detalhada de qualquer objeto, enquanto nossa memória se concentra na essência do objeto, é uma indicação de que não somos máquinas. Retornarei a este assunto no cap. 8.

Uma parte essencial do aprendizado é aperfeiçoar a percepção do conceito, isto é, da essência. Conjeturo que se o procedimento de aprendizado de Turing fosse seguido, logo as máquinas não seriam capazes de reconhecer coisa alguma, porque teriam armazenado uma enorme quantidade de dados conflitantes. Muitas pessoas pensam que ao simplesmente armazenar todo texto escrito do mundo, um computador poderia adquirir "conhecimento" – e ainda mais. Estou absolutamente certo de que o resultado seria catastrófico. Já foi dito que não há nenhuma posição tão absurda que algum filósofo não a tenha tomado [Fetzer 2001, p. 8]. Por exemplo, como um computador reconciliaria minhas idéias com a atual visão científica do mundo? Talvez ele estatisticamente desconsiderasse a minha, em favor daquelas que aparecem mais freqüentemente. Novas idéias nunca seriam aceitas... Suponha que, contando da antigüidade até os dias de hoje, mais pessoas tenham escrito sobre o modelo planetário geocêntrico do que sobre o modelo heliocêntrico desde Copérnico. O computador então estatisticamente admitiria que o primeiro modelo é o correto.

Além do mais, dar nomes é apenas parte do processo de aprendizado de um idioma. E as ações, as qualidades, etc.? Isto nos leva à necessidade de estendermos o TT.

O TT é um teste comportamental, lingüístico e portanto apresenta limites estreitos para aquilo que alguém pode considerar como pensamento e inteligência.

Por outro lado, a sala chinesa de Searle (ver 2.2.) não é comportamental. Sem se olhar para dentro da sala, o comportamento do operador da mesma é supostamente o mesmo de uma pessoa que sabe chinês e que entende as perguntas. A diferença baseia-se no conceito de "compreensão", que é uma atividade interior.

James Fetzer chama a atenção para o fato de que o TT também é reducionista [2001, p. 19], porque lida apenas com a capacidade de manipular símbolos [p. 10]. Ele descreve a extensão de Stevan Harnad:

a) O Teste Turing Total (TTT) [p. 8] – TT mais comportamento não-verbal (capacidades robóticas):

"Harnad introduz o TTT como um teste de comportamento tanto simbólico como não-simbólico, onde símbolos podem ser estabelecidos por meio de comportamentos verbais e não-verbais exibidos pelo sistema. Se dois sistemas exibem comportamentos similares em identificar, ordenar, selecionar e rotular objetos e propriedades de coisas, isso provê uma forte evidência de que tais sistemas entendem [mean] a mesma coisa pelas marcas [marks] que usam." [p. 9.]

Pode-se ver que há uma inferência no "entendem" (o original mean, tem também a ver com "querer dizer"). Considero que a Sala Chinesa de Searle também pode ser estendida para o comportamento robótico. O robô ainda estaria seguindo regras escritas de sintaxe, sem entender o que se passa. Assim, as objeções de Searle podem também ser estendidas ao robô. Por causa desta e das demais objeções, Harnad propôs ainda uma outra extensão:

b) O Teste Turing Total (TTTT): TTT mais a indistinguibilidade corpórea.

Não há evidência de que um computador jamais passará no TT, muito menos o TTT, portanto, imaginando que fosse possível construir uma máquina que seria absolutamente indistinguível de um ser humano parece levar a discussão filosófica um tanto longe no reino da ficção científica. Mas mesmo se isto fosse possível, como Fetzer coloca,

"Se duas máquinas fossem indistinguíveis segundo o TTTT, isto não mostraria que elas têm mente. E mesmo se um robô e um ser humano fossem distinguíveis segundo o TTTT, isso não mostraria que eles não a teriam." [p. 14, grifo do autor.]

Pode-se ver aqui que o problema de considerar seres humanos como sistemas puramente físicos. Não há teoria satisfatória que possa garantir que computadores terão comportamento inteligente.

5.2 IA: Inteligência Artificial ou Imbecilidade Automática?

O campo da Inteligência Artificial (IA) é normalmente dividido em duas categorias: IA forte e fraca.

John Searle caracterizou a IA forte como segue:

"A visão que prevalece na filosofia, psicologia e inteligência artificial é a que enfatiza as analogias entre o funcionamento do cérebro humano e o funcionamento de computadores digitais. De acordo com a visão mais extrema, o cérebro é apenas um computador digital e a mente é um programa de computador. Podemos resumir esse ponto de vista – chamo-a de

'inteligência artificial forte' ou 'IA forte' – dizendo que a mente está para o cérebro assim como o programa está para o hardware do computador." [1991, p. 28].

Esse ponto de vista é representado, por exemplo, por Alan Newell, "que afirma que descobrimos ... que a inteligência é só uma questão de manipulação física de símbolos." [p. 29.] Newell e Herbert Simon, em uma famosa declaração, disseram em 1976 que um sistema físico de símbolos é necessário e suficiente para o que eles chamam de "ação geral inteligente" [Fetzer 2001, pp. 43, 74, 156]

Para James Fetzer,

"Muitos estudantes de inteligência artificial tendem a dividir-se em dois campos amplos mas heterogêneos. Um campo mantém a tese 'forte' de que a IA trata de como nós pensamos. A outra mantém a tese 'fraca' de que a IA trata de como devemos pensar. E há bases para acreditar que a visão predominante hoje entre os pesquisadores é que a tese forte está correta." [p. 74.]

"De acordo com (o que é conhecido como) a tese da IA forte, computadores na verdade possuem características mentais [mentality] quando estão executando programas. ... Por comparação, de acordo com o que é conhecido como a tese da IA fraca, os computadores são simplesmente ferramentas, dispositivos ou instrumentos que são ou podem ser úteis, ajudar ou ter valor no estudo das características da mente [mentality] mas que não possuem mentes, mesmo quando estão executando programas." [p. 159, grifos do autor.]

O ponto de vista de John Pollock lida com a construção de máquinas que têm capacidades humanas:

"A tese da IA forte é a de que podemos construir uma pessoa (uma coisa que literalmente pensa, sente e é consciente) por meio da construção de um sistema físico dotado de 'inteligência artificial' apropriada." [1989, p. ix]

De acordo com as idéias expostas antes, não concordo com essas duas categorias ou teses. A primeira foi suficientemente refutada por Searle e Fetzer. Como vimos, seguindo minha hipótese de trabalho, ela não faz sentido porque inteligência e pensamento não são processos físicos. Para a segunda, podemos aplicar uma de minhas "leis": "Um programa de computador que simula comportamento humano é uma demonstração que os seres humanos não 'funcionam' dessa maneira." (Vejam-se outras "leis" e aforismos no meu site.) Quer dizer, em minha opinião os computadores podem ser válidos como contra-exemplos, mas não como base para estudarmos inteligência ou pensamentos, porque estes últimos não são processos digitais ou algorítmicos.

Adiciono a essas duas categorias outras duas: IA prática e IA humilde. A IA prática lida com a simulação de ações complexas humanas, por exemplo para a substituição do trabalho humano por máquinas. Concordo totalmente com esse ponto de vista, na medida em que o trabalho a ser substituído degrade o ser humano e que um trabalho mais digno seja providenciado para as pessoas substituídas. Por exemplo, sou totalmente contra qualquer uso de computadores na educação de crianças pequenas, por muitas razões [Setzer 1988,

1989 e 2002; ver também artigos em meu site]. A IA prática aplica simplesmente minha noção de inteligência incorporada. A IA humilde é apenas uma coleção de técnicas e algoritmos interessantes da ciência da computação, tal como simulação de jogos, reconhecimento de padrões, programação em lógica, sistemas "especialistas", etc. foi com esse último espírito que eu organizei o programa para a primeira disciplina sobre IA para o bacharelado em ciência de computação na minha universidade, na década de 1970.

A questão das máquinas terem inteligência, isto é, se pode existir uma inteligência artificial, obviamente depende do conceito adotado de inteligência. Vamos primeiramente considerar meus dois tipos de inteligência. Obviamente, computadores, como qualquer máquina, têm uma inteligência incorporada. Mas eles não podem ter inteligência criativa. Em relação a esta, computadores são idiotas absolutos: eles só podem fazer o que são forçados a fazer por seus programas; são a última palavra em burocracia de acordo com a caracterização de di Masi (ver a seção 4.3). Aqui devemos mais apropriadamente falar sobre imbecilidade automatizada. Considerando as múltiplas inteligências de Howard Gardner (cf. 4.2), está claro que a única inteligência incorporada bem desenvolvida que os computadores podem ter é a lógico-matemática, e mesmo esta deve ser restrita à matemática algorítmica e discreta. Alguns outros tipos de matemática podem ser simulados até certo ponto, mas sempre por meio dessa última. Em particular, como veremos no próximo capítulo, toda verdadeira inteligência que requer sentimentos, como a interpessoal (a emocional de Goleman) e a musical (relacionada às artes), não pode ser incorporada em um computador. O mesmo aplica-se às inteligências que requeiram autoconsciência (ver o cap. 7). Portanto, conjeturo que uma máquina jamais passará o TTT (5.1).

6. Uma máquina pode ter sentimentos?

Esta questão tornou-se relevante, porque deixou em larga escala os meios acadêmicos, com os filmes discutidos no cap. 9. Muitos cientistas expressaram sua visão de que as máquinas podem ter – ou que até mesmo já têm – sentimentos. Por exemplo, John McCarthy, o inventor da expressão "Inteligência Artificial", e o introdutor na ciência da computação do campo da "semântica" formal de linguagens de computação, escreveu o seguinte:

"Podemos afirmar que máquinas simples como termostatos têm opiniões [beliefs] e ter opiniões parece ser a característica da maioria das máquinas capazes de resolver problemas." [1979, Introdução.]

Ele caracteriza "opiniões", no caso, da seguinte forma:

" 'A sala está muito fria', 'A sala está muito quente' e 'A sala está OK' – as opiniões sendo atribuídas a estados do termostato da maneira óbvia." [Em sua seção 4.1; ver também Searle 1991, p. 30.]

Utilizei esse exemplo porque para um ser humano essas três frases representam sentimentos, ou melhor o tipo especial de sentimento normalmente chamado de sensação. John Haugeland apresenta uma interessante e detalhada taxonomia dos sentimentos, sendo as sensações a primeira categoria [1987, p 232]. Ele é cuidadoso na questão de atribuir sentimentos às máquinas:

"É surpreendentemente difícil avaliar a relevância dessas questões [os vários tipos de sentimentos] na Inteligência Artificial. Até mesmo a sensação, que deveria ser de alguma forma o caso mais fácil, é de uma perplexidade profunda. Não se pode negar que as máquinas podem 'ter sensação' do que as rodeia, se tudo o que isso significa é discriminação – dar respostas simbólicas em diferentes circunstâncias. Olhos elétricos, termômetros digitais, sensores de tato, etc., são comumente utilizados como órgãos de entrada [input] em toda parte, de brinquedos eletrônicos aos robôs industriais. Mas é difícil imaginar que esses sistemas sentem na verdade qualquer coisa quando reagem a estímulos impingidos a eles. Apesar de o problema ser geral, a intuição é mais clara no caso da dor: muitos sistemas complexos podem detectar danos internos ou mal-funcionamento e até mesmo adotar medidas corretivas; mas será que eles sentem dor? Parece incrível, mas o que exatamente está faltando? Quanto mais penso sobre essas questões, menos sou fico convencido de que sequer sei o que isso significa (o que não quer dizer que acho isso sem sentido)." [p. 235.]

Vamos então partir para a questão de o que é "sentimento".

6.1 O que significa ter sentimentos?

Porque Haugeland – e provavelmente todos – têm tanta dificuldade em caracterizar o que significa "sentimento"? Vamos comparar essa atividade interior com o pensamento. Se eu penso em um conceito, por exemplo o de uma circunferência como o lugar geométrico de todos os pontos equidistantes de um dado ponto (seu centro), esse conceito e a figura geométrica correspondente são absolutamente claros em minha mente. Se olho a entrada de minha sala, vejo algumas cores acompanhadas por algum sentido visual de profundidade e reconheço a percepção global como sendo uma "porta". Esse conceito vem à minha mente também de uma forma absolutamente clara – se eu não fosse capaz de associar essas percepções a esse conceito eu seria ou me tornaria um esquizóide. Por outro lado, o mínimo que posso dizer sobre meus sentimentos é que eles são nebulosos. Por exemplo, neste momento em que estou escrevendo estou um tanto entusiasmado pela formulação que estou dando a este assunto; estou excitado. Mas como posso perceber esses sentimentos? Eles não são claros. Para começar, note-se como qualquer adulto mentalmente saudável reconhece que o pensar tem haver com a cabeça e com o cérebro. Por outro lado, a que parte de nosso corpo associamos com nossos sentimentos? Estou falando aqui sobre sentimentos gerais, como alegria, tristeza, compaixão pelo sofrimento de outra pessoa, etc., e não sobre dor em algum órgão, como dor de cabeça. Minha excitação no momento em que estou escrevendo estas linhas é um sentimento geral; parece-me que tem haver com meu tórax, meus pulmões e coração. Rudolf Steiner associou o pensamento com a consciência desperta, e o sentimento a uma consciência de sonho [1988, palestra de 27 de agosto, p. 76]. De fato é possível estar completamente consciente do próprio pensamento e controlá-lo – ao menos por um certo tempo, como demonstrado através dos exercícios que propus na seção 3.3. Mas sentimentos não são tão claros: uma pessoa pode, por exemplo, confundir sentimentos de frustração, depressão e tristeza. Mais ainda, é impossível controlá-los: se como algo, mesmo que ou não goste do sabor, eu não posso imediatamente mudar este sentimento – leva tempo para aprender a gostar de algo do qual no momento eu não gosto, e o resultado

de um treinamento nessa direção não é garantido. O que podemos controlar são nossas ações baseadas em nossos sentimentos, mas não os sentimentos em si.

Claramente, sentimento significa uma reação interna. Pode ser acompanhado por alguma mudança fisiológica, como corar, empalidecer, sorrir, ter aceleração cardíaca, etc. Mas certamente uma expressão facial não é o sentimento, é uma de suas conseqüências. Provavelmente muitas pessoas diriam que os sentimentos são percebidos pelo cérebro; de fato, é possível detectar o aumento da atividade neuronal em algumas partes do cérebro quando uma pessoa tem certos tipos de sentimentos. Mas o conhecimento científico presente não nos leva a uma clara afirmação de que o sentimento é gerado pelo cérebro; como o pensar, certas atividades cerebrais podem ser conseqüência de sentimentos, e não a causa deles (ver o cap. 3). Bower e Parsons, estabelecendo a teoria de que o cerebelo tem apenas uma função de apoio às atividades de coordenação motora, memória, percepção, atenção, etc., como recentemente proposto, escrevem:

"Uma hipótese como a nossa leva a uma lembrança útil para pesquisa futura: a presença de atividade numa área do cérebro não significa necessariamente que ela está diretamente envolvida em um comportamento ou processo psicológico particular. Por analogia, a maior parte do maquinário sob o capô de um carro está lá para dar apoio à função do motor. Poder-se-iam gerar todos os tipos de hipóteses sobre o papel que o radiador tem na propulsão – por exemplo, correlacionando aumentos de temperatura com quilômetros por hora, ou observando que um carro deixa de funcionar se o radiador é removido. Mas o radiador não é o motor." [2003, p. 47.]

Só falta os cientistas terem a coragem de admitir que todo o cérebro tem apenas a função de apoio a todas as atividades mentais...

É muito importante reconhecer que a atividade interior descrita acima é de um tipo absolutamente diferente de uma reação física, tal como um pedaço de metal submetido à dilatação devido ao aumento da temperatura. Um pedaço de ferro simplesmente não sente nada. Da mesma forma, um termostato – construído com um pedaço de metal que se dilata ou contrai quando submetido a diferentes temperaturas, ou com um par termoelétrico que gera corrente quando suas extremidades têm diferentes temperaturas, ou com um resistor que varia sua resistência com a temperatura, etc. – também não sente nada. Por isso, McCarthy está totalmente errado ao dizer que termostatos têm opiniões. Ele está distorcendo os conceitos de opinião e de sentimento. Além disso, a interpretação de belief como "opinião", implica em entendimento, semântica. Como vimos, máquinas não podem ter semântica; em particular, computadores são máquinas puramente sintáticas.

Há situações em que sentimentos não são devidos a mudanças físicas – pelo menos, não é possível detectar que tipo de mudança é e onde está ocorrendo. Uma situação típica ocorre quando alguém fica vermelho de vergonha. A causa aqui é moral, e não física, tal como seria no caso de se sentir dor devido a um ferimento. Se alguém segue as causas para as mudanças físicas, depara-se com algo que não é físico: ficar vermelho é causado por dilatação nos vasos sanguíneos devido, por exemplo a algum hormônio, que foi produzido por alguma glândula, excitada por algum impulso elétrico, que foi gerado em algum lugar

do cérebro – mas porque esta parte do cérebro gera este impulso? Minha conjectura é que reações morais não são reações físicas.

No caso de sentimentos claramente gerados por mudanças físicas, como a dor, é importante lembrar de que não há idéia precisa do porquê sentimos dor e como os analgésicos exatamente funcionam.

Posso então formular a hipótese que, de forma similar ao pensamento (cf. 3.2.), o sentimento não é um processo físico. Conseqüentemente uma máquina, um sistema puramente físico, nunca terá sentimentos. Esta argumentação metafísica é similar à utilizada para o pensar. Agora darei outra argumentação, puramente lógica, aristotélica, baseada em uma observação que qualquer pessoa pode fazer de si mesma.

6.2 A Subjetividade e individualidade do sentimento

Quando se pensa em conceitos (não em imagens) como "circunferência" e "porta", é possível descrevê-los a outras pessoas. No caso de conceitos matemáticos, uma descrição pode ser exata. Uma outra pessoa é então capaz de pensar o mesmo conceito no qual a primeira, que o descreveu, pensou. Mas esse não é o caso em relação a sentimentos. Sentimentos típicos são os de alegria e tristeza. Posso afirmar que estou alegre, talvez mesmo descrever porque eu esteja tendo esse sentimento, mas ambos não são precisos. Além do mais, a mesma razão pode levar uma pessoa que me ouve a ficar triste; de qualquer modo, ela não será capaz de sentir meus sentimentos. Por exemplo, compaixão é a habilidade social de sofrer quando se vê outra pessoa sofrendo. Mas cada um sente seu próprio sofrimento.

Obviamente, cada um tem que exercitar seu próprio pensamento. Mas o pensar pode ser direcionado para algo universal. Os sentimentos não são universais (não confundir com o conceito de um sentimento), eles têm de ser sentidos por cada um. Steiner coloca esse fato da seguinte maneira:

"Faz parte, pois, da natureza peculiar do pensar ser uma atividade voltada ao objeto observado e não à personalidade pensante. Isso se mostra também na maneira como expressamos ligüisticamente pensamentos sobre objetos diferentes de nós e como falamos sobre nossos sentimentos ou atos volitivos. Quando vejo um objeto e o identifico como sendo uma mesa, digo em condições normais: isto é uma mesa e não: eu penso sobre uma mesa. Quando se trata de um sentimento, digo provavelmente: eu gosto desta mesa. No primeiro caso, não importa dizer que eu estou me relacionando com a mesa; no segundo, é justamente esse relacionamento que interessa." [2000, p. 35, grifos do autor.]

"O pensar é o elemento por meio do qual participamos do universo geral; o sentir, é o meio pelo qual nos retraímos em nosso mundo próprio. Nosso pensar nos une ao mundo. Nosso sentir nos reconduz a nós próprios, fazendo de nós seres individuais. Se fôssemos apenas seres pensantes e dotados de percepção, a nossa vida transcorreria em uma indiferença total. Se apenas nos reconhecêssemos como eu, nosso eu nos seria completamente indiferente. Apenas porque, além de reconhecer a nós mesmos, sentimos também nosso ser, somos entes individuais, cuja existência não se esgota em estabelecer relações conceituais

entre as coisas, mas possui um também um valor particular em si mesma." [p. 80, grifos do autor.]

Assim, o pensar pode ser absolutamente objetivo. O ato de pensar é pessoal, mas o objeto do pensar pode ser universal: todos podem ver a mesa que vejo e pensar a respeito dela. Mas o sentimento é sempre subjetivo. Eu posso gostar da mesa que estou vendo, outra pessoa pode não gostar. Mas ambos teremos que concordar que é uma mesa. Pensando sobre aquela mesa, estou sendo universal. Ao sentir simpatia ou antipatia por aquela mesa, e como o sinto, estou sendo absolutamente individual.

A conclusão é que os sentimentos são totalmente subjetivos e individuais.

Os computadores não são individuais e subjetivos. Não há individualidade em um computador: outro computador pode emular suas instruções em linguagem de máquina, e interpretar um certo programa exatamente da mesma forma como o primeiro. Computadores, com a devida capacidade de armazenamento e velocidade, assim como as Máquinas de Turing, são máquinas universais (ver 3.2.). Máquinas analógicas são certamente pouco diferentes entre si. Por exemplo, se eu ajusto o termostato de minha geladeira vazia em uma certa posição, outra geladeira vazia da mesma marca e modelo, com seu termostato ajustado na mesma posição, produzirá uma temperatura interna um pouco diferente. Mas não é por isso que devemos atribuir individualidade a tais máquinas. Duas delas, de mesmo fabricante e modelo, têm o mesmo projeto, são construídas da mesma maneira e é possível descrever cada parte da máquina, assim como a função da máquina como um todo. É possível prever com certa aproximação como uma máquina reagirá em cada condição; tudo isso dá a ela uma característica universal. Como vimos no cap. 3, cada máquina tem um projetista. Um ser vivo não tem um projetista humano e é impossível descrever completamente o funcionamento de uma parte dele, porque o todo influencia cada parte. Quando se trata de um ser vivo, se uma parte é tirada do todo, ela não funcionará como antes; por exemplo, uma célula sob o microscópio não é a mesma coisa que era no corpo de onde foi retirada. É interessante notar que é impossível prever o quanto uma planta vai crescer [Lewontin 2000, p. 20]. Cada planta tem uma espécie de individualidade, um animal um grau ainda maior de individualidade (pois tem sentimentos!) e o mais alto grau, incomparável ao de qualquer animal, é o do ser humano (pois pode ser auto-consciente e ter liberdade!).

Subjetividade também não pode ser atribuída a nenhuma máquina. Quando um animal ou um ser humano têm uma percepção, há uma reação interna que não é apenas uma transformação física. Algo mais acontece, que não pode ser precisamente descrito. As reações de uma máquina podem sempre ser precisamente descritas (exatamente, do ponto de vista lógico, no caso de computadores funcionando bem).

Assim, qualquer máquina é uma construção objetiva e universal. Vimos que os sentimentos são subjetivos e individuais. Portanto, concluo que as máquinas nunca terão sentimentos. Apenas animais e seres humanos podem ter sentimentos.

6.3 Pensar, Sentir e querer

Como abordei o pensar e o sentir, pode ser interessante completar o quadro com a terceira atividade interna presente nos seres humanos: a vontade. As ações são sempre conseqüência de desejos conscientes ou inconscientes. Os movimentos de nossos membros são um tipo de ação; quando falamos, temos que mover nosso queixo, lábios e língua. Mas quando nos concentramos em nosso pensar, como no exemplo que dei na seção 3.2., também estamos exercitando nossa vontade.

Como vimos (6.1), não podemos dizer onde nossos sentimentos residem em nós. Mas o querer é ainda mais obscuro. Geralmente movemos nossos membros inconscientemente. Imagine-se uma pianista pensando conscientemente em cada movimento de seus dedos, mão e braço, que precisa fazer para tocar cada nota; suas posições, os músculos envolvidos, com que velocidade ela tem que realizar cada movimento, etc. Ela não conseguiria tocar nada. Mas mesmo simples movimentos como os necessários para andar são usualmente feitos de forma inconsciente.

Obviamente, quando sentimos fome e queremos comer, esse impulso tem algo a ver com nosso estômago. Mas como podemos localizar alguns impulsos internos, por exemplo o de seguir um certo estudo, desejar um certo tipo de vida? A avó de minha esposa teve um forte impulso de deixar a Alemanha em 1936, e sua família embarcou no último navio da Holanda para o Brasil em 1939 antes do estouro da guerra e assim salvou-se do holocausto. Seu esposo, assim como muitas pessoas, nem sequer pensaram que a situação para os judeus tornar-se-ia tão perigosa como de alguma forma ela sentiu – mas ela convenceu-o a abandonar a cidade de Hameln, na Alemanha. Quase todos os judeus não sentiram este mesmo impulso e permaneceram na Europa, tendo sido assassinados, como toda a família de meu avô na Polônia (ele já havia emigrado nos anos vinte; um seu irmão esteve no Brasil, não se adaptou e voltou, para ser assassinado em Auschwitz).

Mencionei que Steiner associou o pensar com um estado de consciência desperta e sentimento a uma consciência de sonho (ver 6.1). Ele associa o querer com uma consciência de sono profundo [1988, palestra de 27/8, p. 76]. De fato, podemos tornar-nos conscientes de nosso querer se pensarmos sobre ele. Mas a vontade em si provém de um inconsciente profundo.

As considerações de Steiner sobre essas atividades internas têm amplas aplicações, por exemplo na educação – o principal objetivo da citada série de palestras, dadas aos professores da primeira escola Waldorf.

É interessante observar que a vontade está indiretamente conectada à liberdade. Querer o que se quer não é uma tautologia: nós podemos desejar querer algo, e educar nosso querer. Por exemplo, suponha que uma pessoa seja fumante e reconheça que fumar não é bom para à sua saúde e dos que estão por perto. Além disso, suponha que ele reconheça que não é justo incomodar os não-fumantes, que em geral são muito sensíveis à fina fumaça produzida pelos cigarros (a propósito, considero esse incômodo causado aos não-fumantes um mal social muito pior do que os possíveis males físicos, pois mina a moral do fumante). Conseqüentemente, ele decide parar de fumar. Por um tempo, ele terá um grande impulso para fumar. Mas ele refreia-se de fazê-lo e tenta esquecer o impulso. Sua vontade é não ter aquele impulso, aquela vontade. Com o tempo a sua vontade de fumar desaparecerá e ele

poderá até vir a achar fumar algo desagradável. Ele terá então libertado a si mesmo do impulso que vem talvez de suas condições fisiológica e psicológica. Steiner diz:

"Livre é quem consegue querer aquilo que ele intui como verdadeiro. Quem faz algo diferente daquilo que quer, precisa ser impelido a tais ações por motivos que são dele. Ele é, portanto, não-livre." [2000, p. 139, grifo do autor.]

Assim, há dois tipos de vontade: a que vem do nosso inconsciente e outra que vem de uma decisão consciente ("... o que ele intui como verdadeiro"). Decisões conscientes originam-se de nossos pensamentos. De fato, é da maior importância reconhecer que não somos livres em nossos sentimentos, nem em nossos impulsos de vontade que vêm do nosso inconsciente. Não é possível controlar um sentimento. Por exemplo, ou gostamos de uma comida, ou não gostamos dela, ou somos indiferentes a ela. Por outro lado, é impossível forçar uma vontade que vem do mais profundo inconsciente. Por exemplo, quando sentimos sede temos vontade de beber. Não é possível evitar isto. O que podemos fazer é nos tornarmos conscientes de nossos sentimentos e vontades e evitar algumas ações resultantes deles. Assim, podemos forçar-nos a comer o que não gostamos ou não beber se o líquido disponível, por exemplo, não é saudável. Mas evitamos ou não essas ações após termos pensado sobre nossos sentimentos e vontades. Ao contrário, podemos escolher e produzir nosso próximo pensamento, como já expus na seção 3.2. Assim, somos livres para pensar o que decidirmos, ou o que intuimos "como verdadeiro". Depois de termos, em completa liberdade, escolhido o que pensar, podemos decidir fazer algo no mundo baseado em tal pensamento, isto é, criamos uma vontade – mas uma vontade que veio do nosso consciente, e não de nosso inconsciente. Isso é o que penso que Steiner queria dizer com sua primeira frase. Também é necessário entender o que ele queria expressar com "motivos que são dele". Penso que ele não está se referindo ao nosso corpo, ou às nossas memórias, sentimentos ou vontades inconscientes. Ele está se referindo ao que poderia ser brevemente descrito como nosso "eu superior" ou "personalidade" (self), nossa verdadeira essência – nosso constituinte não-físico que torna possível alcançarmos o mundo das idéias platônicas, sermos indivíduos além das influências de nossos genes e do ambiente, sermos autoconscientes e livres, podermos exercitar um amor altruísta, sermos criativos (ver o cap. 8), etc.

Muitas pessoas não reconhecem que pensar, sentir e querer são coisas separadas, atividades interiores diferentes. Isto vem do fato de vida normal eles sempre virem mesclados. Quando pensamos em algo, imediatamente reagimos com nossos sentimentos. Se sentimos algo como fome, pensamentos surgem imediatamente conectados a ela, como nossa última refeição. Um impulso de vontade imediatamente nos faz pensar nas conseqüências de nossas ações se formos seguir aquele impulso e assim por diante.

Como já vimos, o querer está conectado às ações: desejamos fazer algo ou tornarmo-nos algo – e temos que agir de forma a alcançar esses desejos. Seres humanos podem controlar suas ações antes de realizá-las porque eles podem imaginar as conseqüências antes de executá-las. Os animais não o podem; eles são impelidos pelos seus desejos, instintos e pelo condicionamento, vindo do exterior, que os moldou. Se agimos com o senso de dever, ou devido ao medo, não somos livres. Steiner chamou a atenção para o fato de que apenas agimos em liberdade se agimos devido ao amor pela ação em si [2000, p. 114]. Uma

máquina nunca será capaz de agir dessa forma. As máquinas agem a partir de sua construção física ou inexoravelmente dirigidas por seus programas. Como já dissemos, as máquinas não podem ser livres (ver 3.3). Obviamente, se a noção de vontade é distorcida, como McCarthy fez com o sentir, então é possível atribuir vontade às máquinas.

7. As máquinas podem ter consciência?

Fazer as máquinas tornarem-se conscientes é considerado um dos mais difíceis problemas da Inteligência Artificial.

É necessário distinguir dois diferentes tipos: consciência e autoconsciência. Animais podem ser conscientes: se um animal se fere, ele se torna consciente de sua dor e reage de acordo. Mas apenas seres humanos podem ser autoconscientes. Uma cuidadosa observação levará a essa diferença. A autoconsciência requer pensar. Nós apenas podemos ser conscientes quando estamos totalmente despertos e pensamos a respeito do que percebemos, pensamos, sentimos ou desejamos. Os animais não podem pensar. Se pudessem seriam criativos como os seres humanos. Como já mencionei, nenhuma abelha tenta uma forma diferente do hexágono para o seu favo de mel. Os animais apenas seguem seus instintos e condicionamentos, e agem de acordo. Devido às suas habilidades de pensar, os seres humanos podem refletir sobre as conseqüências de suas ações futuras, e controlar suas ações. Como já mencionei (ver 3.2) um bêbado poder estar consciente, mas certamente não está autoconsciente – ele não pode controlar seu pensamento e ações até mesmo se quiser fazê-lo. Ele age impulsivamente.

Assim, a consciência de animais e de seres humanos depende de sentimentos e a autoconsciência humana depende de pensamento consciente. Como já mencionei, máquinas não podem ter sentimentos e podem apenas simular um tipo muito restrito de pensar: o pensar lógico-simbólico. Assim, eu concluo que as máquinas nunca serão conscientes e muito menos autoconscientes.

É interessante notar que em geral pode-se ler sobre máquinas e consciência, e muito raramente sobre autoconsciência. Talvez isso derive do fato de a maioria dos cientistas consideram os homens simples animais – ou, ainda pior, máquinas.

8. O ser humano é uma máquina?

Como expus no cap. 3, é linguisticamente incorreto dizer que seres humanos são máquinas, porque o conceito de máquina não se aplica a algo que não foi projetado ou construído por seres humanos ou por máquinas. Mas usemos essa denominação popular incorreta, em lugar da mais apropriada "sistema físico".

Há muitas evidências de que seres humanos não são máquinas. Já mencionei algumas delas, tais como o fato de que seres humanos podem autodeterminar seu próximo pensamento. Fetzer argumenta contra a mente ser uma máquina utilizando o fato de haver outros tipos de pensar além do lógico-simbólico, tais como sonhos e divagações sonhadoras [daydreams], o exercício da imaginação e da conjectura [2001, p. 105], e mostra que os símbolos lógicos são uma pequena parte dos símbolos que temos que usar, em termos do semiólogo Peirce

[p. 60]. Ele também concorda com Searle que as mentes têm semântica, e os computadores não [p. 114]. Para mim, o fato de sentirmos e termos vontade é também uma evidência de que não somos máquinas. Outra forte indicação é o fato de termos consciência e autoconsciência, como explicado no último capítulo.

Em particular, as evidências de que não somos máquinas digitais são esmagadoras, como vimos na seção 3.8. Aqui falarei um pouco mais sobre isso, agora a respeito da nossa memória. Se ela é digital, porque lembramos de algo que vemos de uma maneira não tão clara quanto a nossa percepção original? Se nossa memória fosse digital, não haveria razão para o esquecimento – ou perda – de detalhes. Há também um argumento evolucionista nessa direção. Certamente as pessoas que pensam que seres humanos são máquinas também acreditam na evolução darwinista. Mas se fôssemos máquinas, não haveria razão evolucionista para não armazenarmos – pelo menos por algum tempo – todos os detalhes percebidos pelos nossos sentidos, similarmente à capacidade que os computadores têm de armazenar imagens, sons, etc. Parece-me que o armazenamento e lembrança de detalhes certamente aumentaria as chances de sobrevivência e dominação. Isto significa, então, que na perspectiva darwinista nossa memória imperfeita não faz sentido. Isto significa que o conceito darwinista de evolução está errado, ou que não somos máquinas – ou ambos.

Além disso, como é possível "armazenarmos" algo, esquecê-lo e, de repente, sem "consultar" nossa memória, lembrarmos daquilo? Isso não é uma questão de tempo de acesso. Uma máquina ou tem acesso a algum dado ou não o tem, e esse estado somente pode ser alterado por uma ação prevista, programada. Acessos podem ser interrompidos tanto devido a efeitos aleatórios ou de propósito, direcionados pelo programa. Esse não é nosso caso. Frequentemente fazemos um esforço para lembrarmos de algo e não conseguimos – mas certamente memorizamos, em nosso inconsciente, cada experiência que temos. Algumas pessoas poderiam dizer que nosso inconsciente tem um "funcionamento" independente e que ele executa a "busca" por nós. Mais aqui retornamos novamente à questão da consciência e da inconsciência. Certamente todas as máquinas são inconscientes, como já expliquei na última seção. A reação de um termostato não é devido à consciência.

Finalmente, em termos de nossa memória, aparentemente ela é infinita; não há máquina concreta com memória infinita.

A capacidade de aprendizado é para mim também uma indicação de que não somos máquinas. Como disse antes, os computadores não aprendem, eles armazenam dados, por meio de alguma entrada ou como resultado do processamento de dados. Se soubéssemos como aprendemos, os estudos médicos no Brasil não levariam seis anos. O fato deste texto estar sendo lido mostra que o leitor aprendeu a ler. Mas note que durante a leitura ele não segue todo o processo pelo qual teve que passar para aprender a ler. De alguma forma, permanece apenas a técnica, um resultado final do processo de aprendizado. E isso não é uma questão de algum parâmetro calculado ter sido armazenado, como no caso das (erroneamente chamadas) rede neurais.

Nós partilhamos com todos os seres humanos uma extraordinária capacidade de crescimento e regeneração de tecidos e órgãos. Como expliquei na seção 3.5, uma clara

observação mostra que ambos processos seguem modelos. Ora, modelos não são físicos, são idéias. O modelo não-físico está permanentemente agindo sobre os seres vivos para preservar e evoluir sua forma; portanto, eles não podem ser sistemas puramente físicos.

O fato de podermos pensar em conceitos matemáticos é uma indicação de que alcançamos o mundo platônico das idéias (ver o cap. 3). Como disse Spinoza, "Se duas coisas não têm nada em comum entre si, uma não pode ser a causa da outra." (Ética, primeira parte, Prop. 3). Portanto, nós devemos ter em nós algo da mesma qualidade do mundo platônico, não-físico, das idéias para que possamos alcançá-lo. Isto é, nós não somos sistemas puramente físicos.

Respeito aqueles que são materialistas (cf. cap. 3), contanto que sejam conseqüentes. É muito importante perceber que se somos apenas sistemas físicos, não temos liberdade. Tais sistemas são sujeitos apenas às leis físicas, mas de leis físicas ou aleatoriedade é impossível ter um organismo livre. Sem liberdade, não há dignidade humana ou responsabilidade. Este foi um grande dilema para Einstein, que considerou os seres humanos como sistemas puramente físicos, determinísticos e que não possuíam liberdade [Jammer 2000, pp. 62, 71, 173], assim não poderia atribuir-lhes responsabilidade [pp. 71, 105]. Mas ele teve que mudar de idéia quando se conscientizou dos horrores perpetrados pelos nazistas. Acabou até mesmo por responsabilizar todo o povo alemão [p. 71]. Assim, se um materialista assume que os seres humanos podem ser livres, ter dignidade humana e responsabilidade, podem exercer um amor altruísta, ou que há propósito na vida, ele deve mudar de lado e se tornar um espiritualista, isto é, ele deve assumir a hipótese de que também existem processos não-físicos. Estes processos acontecem em um mundo não-físico, e podem influenciar o universo físico.

Muitos cientistas presumem que deve ter havido algo não-físico atuando no começo do universo material. De fato, de ponto de vista materialista, a origem da matéria ou da energia e as fronteiras do universo não fazem sentido. De um ponto de vista materialista, nem mesmo a matéria faz sentido: porque não subdividi-la indefinidamente? (Por exemplo, hoje o elétron é considerado não-estruturado.) Infelizmente, o materialismo influenciou tanto os cientistas que muitos, se não a maioria, daqueles que admitem alguma "criação" original da matéria não podem admitir que ainda haja um universo não-físico agindo "por detrás" do físico. Um exemplo disso foi Einstein [Jammer 2000, p. 97].

É muito importante perceber que presumir a hipótese espiritualista não leva ao misticismo ou a uma "má ciência". Se o espírito básico da ciência moderna é mantido, o que acontece é que essa hipótese expande o espaço de pesquisa. Por exemplo, poder-se-ia assumir que algumas mutações evolucionistas são não-aleatórias, sendo direcionadas por constituintes não físicos de certos seres vivos. Dessa forma ter-se-ia um superconjunto próprio da evolução materialista darwinista. Outro exemplo é supor que o pensar não é produzido pelo cérebro (ver 3.2). Isto expandiria enormemente o campo de pesquisa sobre cognição; em lugar de examinar a atividade dos neurônios como origem do pensamento, saeria investigado, por exemplo, como o pensar influencia a atividade neuronal, sem descartar a possibilidade de, em certas circunstâncias, os neurônios poderem agir independentemente da atividade não-física de pensar, induzindo alguns tipos de pensamentos.

Talvez eu deva ser mais específico. De um ponto de vista espiritualista, as diferenças entre os vários reinos da natureza são facilmente justificadas. Os minerais são sistemas puramente físicos; vamos chamar seu corpo físico de "um membro do primeiro tipo". As plantas obviamente têm um corpo físico, um membro do primeiro tipo. Mas de acordo com a hipótese que estamos formulando, as plantas não são apenas sistemas físicos. Vamos chamar seu constituinte não-físico de "um membro individual (isto é, cada planta tem seu próprio) de um segundo tipo". Esse constituinte, agindo sobre o primeiro (o corpo físico), é responsável pela forma orgânica, regeneração dos tecidos, crescimento, reprodução, etc. – em suma, todos os processos vitais. Os animais mostram mais do que processos vitais – eles têm algo mais do que as plantas. Vamos chamá-lo de "membro individual (não-físico) de um terceiro tipo". É responsável pelo movimento, órgãos ocos, respiração, consciência, sentimentos, instintos, desejo, etc. Obviamente, os animais têm todo tipo de manifestação de vida, portanto têm um membro do segundo tipo além do corpo físico. Os seres humanos têm diferenças essenciais dos animais, devido a um outro constituinte não-físico, além dos dois inferiores: vamos chamá-lo de "um membro individual de um quarto tipo". Ele é responsável por nossa posição ereta, pelo pensar, pelo falar, pela autoconsciência, pelo amor altruísta, pela liberdade, pelos ideais, etc. Ela também é responsável pelo que poderíamos chamar de "individualidade superior", "eu superior" ou "eu", algo maior do que, individualmente, apenas nossa forma física, costumes, instintos, preconceitos, temperamento, lembranças, etc. O membro do quarto tipo é da mesma "substância" não-física que os pensamentos – daí podermos alcançar por meio do nosso pensar o mundo platônico das idéias, e animais e plantas não o podem (ver a citação de Spinoza acima). Esta é a razão pela qual eles não são criativos no sentido humano. A presença atuante de cada membro não-físico modifica os membros inferiores, incluindo o corpo físico. Por isso a forma das plantas difere tanto das formas dos minerais, a dos animais das plantas e a dos seres humanos das dos animais. Na verdade, quando observamos os quatro reinos à procura de diferenças, deveríamos concentrar-nos em espécies típicas de cada reino, e não em espécies de transição. Quando um ser vivo morre, seus membros não-físicos "abandonam" o corpo físico, isto é, uma conexão entre eles é cortada. O corpo físico se torna totalmente sujeito às leis da natureza, seguindo-se o decaimento. Nesse sentido a vida é uma luta permanente contra o mundo físico, contra a morte. Quando animais e seres humanos estão em um estado de sono profundo, houve uma certa separação entre os dois primeiros membros e os superiores. Processos vitais continuam, mas não há consciência. De acordo com essa hipótese, a morte e o sono não são apenas processos físicos. Conjeturo que uma visão materialista do mundo nunca nos fará entender o que realmente significam sono e morte. A propósito, um antigo ditado é "a morte é a irmã mais velha do sono".

Por favor, leitor, note que o último parágrafo contém conceitos. Não é baseado em religião, fé ou misticismo. É apenas uma forma de expandir nossos conceitos comuns para abraçar conceitos não-físicos por meio de hipóteses conceituais. Mas sua admissão teria tremendas conseqüências para a pesquisa científica, para as relações humanas (e também dos seres humanos para com os outros três reinos da natureza) e a moralidade.

Para alguém que admite que há um "mundo" não-físico e que plantas, animais e seres humanos têm constituintes não-físicos, considerar que seres humanos (ou animais, ou plantas) são máquinas não faz sentido, porque as máquinas são simplesmente sistemas físicos.

Para aqueles que, lendo minha descrição extremamente breve dos membros não-físicos humanos, reconheceram-nos, devo dizer que não usei alguns dos nomes atualmente atribuídos pela literatura especializada porque queria evitar conotações desnecessárias. Para aqueles que estão interessados em saber mais e conceitos mais profundos sobre este assunto, posso recomendar a leitura de meu artigo a esse respeito, que tem sido usado como texto de um curso introdutório ministrado por mim e minha esposa nos últimos anos [Setzer 2000].

9. Os filmes O Homem Bicentenário e Inteligência Artificial

Esses filmes recentes, dirigidos por C. Columbus (1999) e por S. Spielberg (2001), respectivamente, mostram robôs que agem como seres humanos, interpretados por atores reais. Vou criticá-los, porque em minha opinião eles exercem influência perniciosa sobre o público.

Em resumo, O Homem Bicentenário (HB) conta a história de um robô que dura cerca de 200 anos. Seu criador e seus descendentes mudam o robô, de forma que ele passe a ter mais e mais características humanas. Robin Williams faz o papel do robô. No começo, ele usa um tipo de armadura, que o faz parecer uma máquina; durante a noite ele conecta um plug na tomada elétrica para recarregar suas baterias. Gradualmente, sua aparência se torna mais e mais humana, até um ponto que ele age como um ser humano normal, com pensamentos, sentimentos e desejos, ou seja, o robô teria passado no TTT (cf. 5.1). Não fica claro se ele teria passado o TTTT. No final a mulher pela qual o robô está apaixonado vai morrer, ele decide que não pode ficar sem ela e também deve "morrer".

Inteligência Artificial (I.A.) começa com a história de uma família cujo filho está em coma. O pai trabalha para uma fábrica de robôs, tão similares aos seres humanos que há um teste especial para distingui-los: um dispositivo muito similar ao leitor de código de barras é apontado para a testa de uma pessoa, emitindo um tipo de raio laser. Se é possível ver algum circuito sob a pele, a "pessoa" é um robô. Isto é, todos esses robôs também passam no TTT, mas certamente não passam no TTTT. A fábrica decide que falta aos robôs uma característica essencial: eles não podem ter sentimentos. Realizam pesquisas e programam o primeiro robô que tem sentimentos, sendo até capaz de amar: um garoto-robô. Ele é dado como teste à família cujo filho está em coma, como um filho substituto. Eles recebem o garoto-robô, que inicialmente comporta-se de uma maneira totalmente fria. Num certo momento a mulher diz uma série de palavras-chave na frente do robô, que a partir de então passa a ter a habilidade de ter sentimentos. Ele se torna muito apegado à mulher, mas em um determinado ponto seu filho real retorna curado do hospital. Com o tempo, há desacordos entre o garoto real e o robô, e a mulher tem que se livrar deste último. Ela o faz sentindo extrema dor. O robô sente saudades da mulher. Ele está a bordo de um tipo de nave aquática que cai na cidade de Nova York; o robô entra em um tipo de estado de sonolência. Mil anos passam-se; vê-se agora novamente o robô dentro da nave, no mesmo lugar, mas dentro do mar. Devido ao efeito estufa, os oceanos cobriram grande parte das cidades. Os seres humanos desapareceram. O garoto-robô "acorda", e ainda sente saudades da mulher. É recebido por pessoas entranhas, altas e magras, com uma aparência

extraterrestre. Dizem ao robô que podem reproduzir a mulher, por apenas um dia, se ele tiver uma parte dela. Ele encontra um cabelo da mulher e dele cabelo os ETs a reproduzem.

9.1 Absurdos

Eu nunca assisto a filmes de ficção, pois seus absurdos irritam-me em demasia. Porém, os membros de um grupo de estudo semanal do qual participo pediram-me para assistir I.A. e comentá-lo, portanto eu fui vê-lo. Depois disseram-me que devia também assistir ao HB e novamente concordei.

Vamos primeiro ver alguns absurdos triviais. No HB, o robô inicialmente recarrega a si mesmo, mas depois não mais. Talvez ele tenha adquirido a habilidade de digerir a comida. Em I.A., o robô claramente não tem tal habilidade, porque em uma cena ele decide imitar o garoto real e começa a engolir macarrão, obstruindo seus sistemas. Mas ninguém o vê recarregando suas baterias, ou retirando energia de algum lugar. Além do mais, ele permanece parado por 1000 anos, não é corroído e "acorda" tendo retido sua energia por todo este tempo.

Mas há absurdos mais profundos. Por exemplo, em I.A., do cabelo de uma mulher, seu corpo completo e suas memória, temperamento, etc., são reproduzidos 1000 após sua morte. O DNA não determina por si só o desenvolvimento de um organismo [Lewontin 2000, Setzer 2000a], muito menos as memórias e o temperamento de um ser humano. Porém, em minha opinião o absurdo mais sério é que robôs serão capazes de ter sentimentos. Já demonstrei que máquinas nunca terão sentimentos.

9.2 Mensagens

Eu disse que esses filmes têm efeitos perniciosos sobre o público. O que quis dizer é que eles dão uma imagem incorreta e absurda das máquinas, influenciando a maneira como as pessoas consideram o mundo, as máquinas e os seres humanos em particular. Coletei as seguintes mensagens transmitidas pelo filmes; quando um item se aplica a ambos os filmes, não faço nenhuma indicação, caso contrário menciono a qual filme se aplica.

- 1) Robôs poderão comportar-se física e mentalmente como seres humanos (em nossos termos, passarão o TTT, cf.5.1).
- 2) Será possível fazer com que as máquinas tenham sentimentos humanos.
- 3) Robôs podem ter ideais, no sentido humano.
- 4) Os seres humanos amarão robôs, da mesma forma que se amam uns aos outros.
- 5) Robôs com sentimentos são uma tragédia para "si mesmos", porque eles não "morrem"
- 6) Robôs não precisarão de energia para funcionar.
- 7) Haverá máquinas que durarão infinitamente, sem corrosão ou defeitos.
- 8) A humanidade desaparecerá, e apenas os robôs permanecerão na Terra (A.I.).
- 9) Do DNA de um cabelo, é possível recriar um ser humano completo, incluindo suas memórias e temperamento (A.I.).

Em ambos filmes, o fato de todas as habilidades humanas poderem ser introduzidas nas máquinas pode dar aos leigos a impressão de que não somente as máquinas se tornarão

humanas, mas que também nós devemos ser máquinas (do contrário teríamos alguma característica não-maquinal que não seria passível de ser inserida em uma máquina).

O leitor pode pensar que minhas preocupações são exageradas, porque as pessoas não misturam fantasia com realidade. Contudo, há alguns problemas a respeito desses filmes. Essa argumentação aplica-se a adultos; as crianças vivem (ou deveriam viver) em grande parte na sua fantasia, portanto elas não são capazes de distinguir precisamente a fantasia da realidade. Além disso, os leigos, isto é, pessoas que têm pouca formação científica, ou até mesmo as que conhecem computadores como simples usuários, não podem fazer um julgamento correto da mensagem transmitida pelos filmes. Isto é agravado pelo fato de que muitos cientistas realmente pensam que aquelas cenas de ficção-científica tornar-se-ão realidade. I.A. foi apresentado como avant-première no MIT Media Lab. Um artigo escrito por um jornalista científico que estava naquela seção descreveu as opiniões de alguns cientistas sobre o filme. Ele escreveu:

"Para Kurzweil, o enredo do filme – um robô-criança com sentimentos, que tenta conquistar o amor dos pais adotivos – não tem nada de fantástico. 'Em 2030 não haverá nenhuma distinção clara entre nós e os robôs. As emoções, em especial o amor, são as coisas mais profundas e complexas de que somos capazes. Mas, em 25 anos, saberemos tudo sobre o cérebro humano e seremos capazes de reproduzi-lo com perfeição em máquinas. Elas serão capazes de fazer todas as coisas que fazemos, inclusive amar.' " [Burgierman 2001, p. 50].

Isso é consistente com o seu livro [Kurzweil 1999]. Podemos assim ver que previsões científicas impróprias presentes naqueles filmes estão em total acordo com o ponto de vista de muitos cientistas. De fato, aqueles filmes vieram ajudar alguns cientistas a passar para o público sua visão de que somos máquinas.

9.3 Influências nas visões do mundo

Minha maior preocupação com aqueles filmes é que eles apresentam uma idéia – e imagens – de que toda função e característica humana pode ser inserida em uma máquina. Ainda pior, as máquinas têm características superiores, tais como velocidade, força e memória exata. Portanto, se inserirmos nossas características nas máquinas, elas serão superiores a nós. Isto foi expresso por uma frase do físico Steve Hawking publicada nos jornais, quando ele disse algo como "nós deveríamos melhorar o ser humano geneticamente, do contrário ele será suplantado pelas máquinas". Mesmo que ele não tenha dito isso – poderia ser conversa de jornalista –, não tem importância: a frase encaixa-se perfeitamente na visão de mundo adotada por muitos, provavelmente a maioria, dos cientistas. Um dos grandes problemas é que, sendo materialistas, eles não podem especificar o que entendem por "melhora": deveríamos ter três metros de altura, nunca ter de dormir, nunca contrariar a lei? Talvez atingiremos a perfeição se agirmos como robôs.

Em segundo lugar, se pudermos inserir cada característica humana nas máquinas, isso significa que nós mesmos somos máquinas.

Temo que se os seres humanos considerarem-se a si mesmos como máquinas, um futuro muito negro está a espera de nós e de nossos descendentes. Vou fazer uma digressão a esse respeito nas conclusões.

10. Conclusões

Tentei mostrar neste artigo que a visão dominante no campo da AI é que seremos capazes de (1) inserir nas máquinas todas as nossas funções e características, porque (2) nós somos simplesmente máquinas ou, mais precisamente, sistemas meramente físicos. Dei muitas razões para concluirmos que este dois pontos de vista são falsos. Para isso, por um lado usei alguns fatos que podem ser aceitos por qualquer um e, por outro, usei uma hipótese de que todos os seres humanos têm constituintes não-físicos. Dei várias indicações de como se pode considerar ambas abordagens como adequadas.

Devo confessor que gosto de lutar por causas justas – contra moinhos de vento, para algumas pessoas; de meu ponto de vista, no entanto, são contra monstros reais. A primeira dessas lutas começou publicamente em 1972, contra a TV. Em seguida vieram os computadores na educação elementar, depois os joguinhos eletrônicos e, agora, a IA. Ou, melhor dizendo, as IAs forte e fraca, porque eu não tenho nada contra o que denominei acima de IAs prática e humilde (cf. 5.2). Minha batalha contra as atuais tendências filosóficas em IA é devida ao fato de elas consideram seres humanos como sistemas meramente físicos ou, popularmente, máquinas. Considero essa visão extremamente perigosa, porque se consistente ela tem que negar a liberdade, a responsabilidade e a dignidade humanas, assim como a possibilidade de se exercer um amor altruísta. Se este último é admitido, é em geral considerado como uma característica programada em nós pela evolução, como afirmado pelo próprio Darwin, e elaborado em termos genéticos especulativos por Richard Dawkins [1989, p. 23]. Considero o verdadeiro amor altruísta como um ato praticado em total consciência e liberdade, de modo que ele é inconsistente com a visão darwiniana de que somos apenas animais ou, de acordo com as IAs forte ou fraca, ainda pior, isto é, somos apenas sistemas físicos. Como disse, um sistema físico é completamente sujeito às leis da física, assim não pode ter liberdade ou dignidade e exercer amor altruísta.

A idéia de que os seres humanos são animais produziu tremendas catástrofes humanas. Essa foi claramente a visão de Hitler do mundo, porque ele tratou pessoas como animais, transportando-as em trens de gado, enjaulando-as, matando-as com gás, submetendo-as a extremo sofrimento, fazendo experiências de laboratório com elas, etc. Uma de suas ideologias era que a sobrevivência do mais capaz deveria se aplicar aos seres humanos: apenas uma nação no mundo deveria dominá-los [Haffner 1990, p. 79]. Depois que a invasão da União Soviética falhou durante o inverno de 1941-42, ele percebeu que não venceria a guerra. Decidiu então destruir o próprio povo alemão, permitindo assim o aparecimento de outra nação que deveria dominar o mundo [pp. 118, 142-157, ver também Toland 1992, pp. 707, 849]. Outras catástrofes de mesma linha foram perpetradas pelos turcos contra os armênios, por Stalin, Mao e Pol Pot contra seu próprio povo, em Hiroshima e Nagasaki e assim por diante. O século 20 foi realmente o que chamo de "o século da barbárie" (tenho uma leve esperança de que o 21 não o ultrapasse nesta triste sina). Todas estas tragédias foram, na minha opinião, baseadas na idéia de que os seres

humanos – ou pelo menos parte da humanidade, os inimigos daqueles de outro grupo étnico, de fé ou de ideologia – são animais.

O que acontecerá se as pessoas abraçarem amplamente a atual idéia, defendida por muitos, se não a maioria dos cientistas – especialmente no campo da IA –, de sermos simplesmente máquinas? Temo que o futuro seja muito pior do que as catástrofes citadas no parágrafo anterior. Na verdade, o que estamos vendo no momento em todo lugar em termos de decadência social e individual pode muito bem ser, em boa parte, a consequência daquela visão de mundo. Podemos ver um crescente desrespeito geral pelas pessoas (com honrosas exceções, ver abaixo), um aumento da angústia psicológica nos indivíduos, e assim por diante. No número de julho 2002, a revista americana Newsweek trouxe um artigo dizendo Que 70 milhões de americanos sofrem de insônia (ver o jornal O Estado de São Paulo, 22/7/02, p. A8). As religiões, que foram largamente deixadas de lado porque não seguiram nosso desenvolvimento, principalmente nos últimos dois séculos, foram substituídas pela fé na ciência e na tecnologia. Falando em termos gerais, as religiões tradicionais contribuíram em geral para impulsos morais positivos nas relações sociais – pelo menos entre seus membros; por outro lado, a ciência e a tecnologia atual não lidam com a moral e a ética. Elas lidam com teorias, objetos e máquinas; elas vêem os seres vivos como máquinas. Um segundo fator, a tradição, também ajudou a manter a sociedade razoavelmente estável, mas no começo do século 20 as tradições começam a desaparecer e perderam sua força coesiva social. Um terceiro fator, sensibilidade social intuitiva (uma forte característica do brasileiro, mas que infelizmente está diminuindo), parece também estar desaparecendo. Hoje em dia os seres humanos são abandonados, deixados por sua própria conta para decidir o que fazer consigo mesmos. Esse é o passo difícil que temos de dar, no sentido de adquirir nossa própria liberdade. Infelizmente estamos falhando miseravelmente, principalmente porque não estamos reconhecendo o que um ser vivo realmente é, o que um ser humano e o seu desenvolvimento deveriam de fato significar. Vou darei um exemplo dessa situação.

A visão científica atual sobre as doenças é que todas elas devem ser erradicadas. Mas outra visão do ser humano pode contar uma história completamente diferente: as doenças são necessárias para um verdadeiro desenvolvimento individual. Isto não significa que devemos ir por aí inoculando qualquer possível doença em cada indivíduo que encontrarmos. Mas significa que devemos entender que muitas doenças podem ser um processo pedagógico, adquiridas por cada indivíduo quando ele delas precisa. A sabedoria da linguagem natural mostra isso muito bem: alguém não diz "um resfriado pegou-me", mas sim "eu peguei um resfriado" – isto é, a pessoa tinha a predisposição, talvez a necessidade, de pegá-lo. Os médicos deveriam ajudar seus pacientes a vencer suas doenças passando positivamente pelos processos correspondentes. Em lugar disto, eles apenas dão remédios e fazem cirurgias como se os pacientes fossem máquinas com algumas partes defeituosas, precisando ser consertadas. Tenho a impressão de que os seres humanos modernos sem essas doenças necessárias não seriam mais seres humanos – eles seriam autômatos.

Eu disse que há honrosas exceções a respeito ao aumento do desrespeito ao ser humano. Uma delas é a campanha contra o fumo: que direito tem um indivíduo de perturbar uma pessoa que não fuma? (A fumaça dos cigarros, sendo muito fina, é altamente irritante para os não-fumantes.) Os direitos humanos, o respeito às pessoas deficientes, anti-racismo e

anti-sexismo e são outras características muito positivas de nossos tempos. Em termos de todos os seres vivos, o movimento ecológico é também muito positivo. Tudo isso mostra que os seres humanos também evoluíram em direções positivas, indicando uma alta sensibilidade, ao menos em algumas direções, em relação a natureza e em relação aos outros seres humanos.

Infelizmente, não desenvolvemos novas e maiores habilidades sociais positivas, baseadas na consciência, autoconsciência, liberdade, individualidade, sensibilidade social, responsabilidade e ação. Por exemplo, em muitos lugares, como supermercados e restaurantes, se é forçado a ouvir música de fundo enlatada. Pouquíssimas pessoas parecem perceber que essa música prejudica a liberdade individual de forma similar à do fumo. Nenhuma daquelas habilidades será corretamente desenvolvida se, para começar, temos uma visão errada e parcial da humanidade e do que os demais reinos da natureza são. Esse conceito errado é que os seres humanos são máquinas, e que produziremos máquinas bem melhores do que os seres humanos.

Há uma grande razão para a idéia de que somos apenas máquinas, que foi explicada no cap. 3: o materialismo. O materialismo foi desenvolvido durante os dois últimos séculos, como uma necessidade para a humanidade. Ele tornou possível a nós nos imergirmos na matéria até um ponto que não seria possível sem ele. Sem essa imersão não teríamos desenvolvido nossa capacidade de sermos livres e autoconscientes. Mas penso que agora é a hora de conscientemente superar esta visão de mundo, sem perder tudo que desenvolvemos, do contrário continuaremos a ver a miséria social crescendo continuamente. Temos que desenvolver novas maneiras de organizar a economia, desviando do infeliz princípio que tem governado o capitalismo desde o século 18: a idéia de Adam Smith de que se satisfizemos nossas ambições e egoísmos – em um sentido meramente material – uma misteriosa "mão invisível" regularia a sociedade e todos seriam felizes. Mas uma nova organização social, baseada no materialismo, não pode produzir as mudanças sociais que urgentemente precisamos, porque ela ignora a essência do ser humano: o fato de que ele tem constituintes não-físicos (cf. 8). Como consequência, ela deve ignorar a possibilidade de exercitar o amor altruísta o qual é, obviamente, socialmente construtivo, enquanto o egoísmo é destrutivo.

Infelizmente, há forças que querem evitar o reconhecimento de que o materialismo tem de ser superado. As IAs forte e fraca são parte de suas manifestações. Espero ter mostrado que elas estão erradas. O homem não é sistema meramente físico; nossas atividades de pensar, sentir e ter vontade não se originam em nossas partes físicas. Portanto, será impossível introduzir uma mente humana real nas máquinas; estudar e desenvolver máquinas nunca revelará nossa essência real ao contrário, isto desvia nossa atenção. O grau de deterioração dos recursos naturais, inclusive o ar, a água e o solo agrícola (que paradoxo: nossa era essencialmente materialista destrói a matéria), o crescimento da instabilidade social e econômica e a miséria que qualquer um pode observar torna absolutamente urgente que mudemos algo. Penso que isso tem que começar pela mudança radical da visão que os cada pessoa têm de si mesma e dos outros seres humanos. Essa visão tem contribuído para a eliminação de nossa dignidade humana e da responsabilidade social. Filmes como os mencionados vão nas mesmas direções, agora em termos populares.

Outros cientistas também chegaram a conclusão que nós estamos em perigo, mas por outras razões. Por exemplo, Bill Joy teme a maneira como a IA (especialmente a robótica), a engenharia genética e a nanotecnologia estão sendo desenvolvidas: podem ser introduzidas máquinas autoreprodutoras que destruirão o mundo [2000]. Eu não concordo com ele; não acho que essas áreas serão capazes de chegar a esse ponto. Mas suas preocupações são as de um cientista que está preocupado com o fato de a ciência e a tecnologia terem saído do controle. Também não concordo com a sua solução, isto é, deve ser declarada uma moratória para as pesquisas ligadas a tais campos. Meu ponto de vista é que os meios devem estar identificados com os fins. Não é por meio da restrição à liberdade que conseguiremos mais liberdade. Acho que a solução reside na decisão e na ação individuais de cada cientista e técnico – eles devem individualmente decidir o que devem investigar e produzir. Espero que essas linhas tenham ajudado àqueles que estão procurando por uma ciência mais responsável, a tornarem-se mais conscientes de que as IAs forte e fraca não são os campos que devem ser investigados para o desenvolvimento da humanidade. Ao contrário, se exercidos, esses campos apenas contribuirão para acelerar e aumentar nossa miséria. Nossos maiores problemas não são problemas materiais. Apenas resolvendo nosso maior problema, derivado do que eu caracterizei como a "hipótese fundamental existencial" (cf. 3.4), isto é, a maneira como consideramos nós mesmos e o mundo, seremos capazes de reverter os crescentes decaimentos social, individual e da natureza.

Referências

Bower, J.M. e L.M Parsons. Rethinking the "Lesser Brain". Scientific American 289 (2), Aug. 2003, pp. 41–47.

Burgierman, D.R. Inteligência Artificial. In Super Interessante. No. 166. São Paulo: Editora Abril, julho de 2001, pp. 49-54.

Damasio, A. Descartes' Error - Emotion, Reason, and the Human Brain. New York: Grosset/Putnam, 1994.

Dawkins, R. O Gene Egoísta, trad. A. P. Oliveira. Lisboa: Gradiva, 1989.

Encyclopaedia Britannica. Chicago: Encyclopaedia Britannica, 1966.

Fetzer, J. H. Computers and Cognition: Why Minds are not Machines. Dordrecht: Kluwer Academic Publishers, 2001.

Gardner, H. Inteligências Múltiplas: A Teoria na Prática. Trad. M.A.V. Veronese. Porto Alegre: Ed. Artes Médicas Sul, 1995.

Goleman, D. Inteligência Emocional: A Teoria Revolucionária que Redefine o que é Ser Inteligente, trad. M. Santarrita. Rio de Janeiro: Ed. Objetiva, 1995.

Goswami, A. Death and the Quantum: A New Science of Survival and Reincarnation. 1995. Available at <http://www.swcp.com/~hswift/swc/Essays/death.html>.

- Haffner, S. Anmerkungen zu Hitler. Frankfurt a.M.: Fischer Taschenbuch Verlag, 1990.
- Haugeland, J. Artificial Intelligence: The Very Idea. Cambridge: MIT Press, 1987.
- Jammer, M. Einstein e a Religião: Física e Teologia, trad. V. Ribeiro. Rio de Janeiro: Contraponto Editora, 2000.
- Jackendorff, R. Languages of the Mind: Essays on Mental Representation. Cambridge: MIT Press, 1993.
- Joy. B. Why the future doesn't need us. Wired 8.04, April 2000. Available at <http://www.wired.com/wired/archive/8.04/joy.html>.
- Kant, I. The Critique of Pure Reason. In Great Books of the Western World Vol. 42, R.M.Hutchins (ed.). Chicago: Encyclopaedia Britannica, 1952, pp. ix-209.
- Koestler, A. The Sleepwalkers - A History of Man's Changing Vision of the Universe. Harmondsworth: Penguin Books, 1964.
- Kurzweil, R. The Age of Spiritual Machines – When Computers Exceed Human Intelligence. New York: Penguin Books, 1999.
- Lewontin, R. The Triple Helix - Gene, Organism, and Environment. Cambridge, MA: Harvard University Press, 2000.
- McCarthy, J. Ascribing mental qualities to machines. 1979. Available at <http://www-formal.stanford.edu/jmc/ascribing.html>.
- Neumann, J. v. The Computer and the Brain. New Haven: Yale University Press, 1958.
- Penrose, R. The Emperor's New Mind – Concerning Computers, Minds and the Laws of Physics. New York: Penguin, 1991.
- Pollock, J. L. How to Build a Person: A Prolegomenon. Cambridge: MIT Press, 1989.
- Ramsey, W., S. Stich and D. Rumelhart (eds). Philosophy and Connectionist Theory. Hillsdale: Lawrence Erlbaum Ass., 1991.
- Rohsen, J. W. Morphologie des menschlichen Organismus. Stuttgart: Verlag Freies Geistesleben, 2000.
- Searle, J. R. Minds, Brains and Science - the 1984 Reith Lectures. London: Penguin Books, 1991.

Setzer, V.W. O computador no ensino: nova vida ou destruição? in E.O.C. Chaves e V.W.Setzer, O Uso de Computadores em Escolas, São Paulo: Ed. Scipione,1988, pp. 69-123.

Setzer, V. W. Computers in Education. Edinburgh: Floris Books, 1989.

Setzer, V. W. Reflections on electronic chess. In The Southern Cross Review. No. 2, Nov.-Dec. 1999, revista eletrônica disponível em <http://www.southerncrossreview.org>. Também em meu site.

Setzer, V. W. Uma introdução antroposófica à constituição humana. 2000. Disponível em <http://www.sab.org.br/antrop/const1.htm>.

Setzer, V. W. Desmistificação da onda do DNA. Disponível em meu site, 2000a.

Setzer, V. W. Meios Eletrônicos e Educação: uma Visão Alternativa. São Paulo, Ed. Escrituras, Série Ensaio Transversais No. 10, 2a. ed. 2002.

Sheldrake, R. A New Science of Life - The Hypothesis of Formative Causation. Los Angeles: J.P.Tarcher, 1987.

Steiner, R. A Arte da Educação I – o estudo geral do homem, uma base para a pedagogia (GA 293). 14 palestras proferidas em Stuttgart, 21/8-5/9/1919. Trad. R. Lanz. São Paulo: Ed. Antroposófica, 1988.

Steiner, R. De Jesus a Cristo (GA 131). 11 palestras proferidas em Karlsruhe, 4-14/10/1911. Trad. R. Lanz. São Paulo: Ed. Antroposófica, 1997

Steiner, R. A Filosofia da Liberdade (GA [Catálogo Geral] 4). Trad. M. da Veiga. São Paulo: Ed. Antroposófica, 2000.

Sutherland, I. V. and Jo Ebergen. Computers without clocks. In Scientific American, Vol. 287, No. 2, Aug. 2002, pp. 46-53.

Toland, J. Adolf Hitler. New York: Anchor Books, 1992.

Turing, A. M. Computing machinery and intelligence. In Mind - a Quarterly Review of Psychology and Philosophy, Vol. LIX No. 236, Oct. 1950, pp. 433-460, disponível em <http://www.abelard.org/turpap/turpap.htm>. Também in E. Feigenbaum and J. Feldman, eds., Computers and Thought. New York: McGraw-Hill, 1963, pp. 11-35.

Zajonc, A. Catching the Light - The Entwined History of Light and Mind. New York: Bantam, 1995.

Agradecimento

Agradeço a Frank Thomas Smith, editor da excelente revista eletrônica Southern Cross Review, por comentários em relação ao conteúdo da versão original em inglês.

* Departamento de Ciência da Computação, Universidade de São Paulo, Brasil

vwsetzer@ime.usp.br - www.ime.usp.br/~vwsetzer

Escrito em 11/2002 – esta é uma tradução do original em inglês, feita por Morgana Ribeiro, por iniciativa de Anderson Paulino, da versão 3.1 de 2/6/2003; última revisão de V.W.Setzer em 27/4/05

Disponível em: <<http://www.ime.usp.br/~vwsetzer/IAtrad.html>>. Acesso em: 02 mai. 2007.