

# COMPUTAÇÃO DE MAQUINARIA E INTELIGÊNCIA

Por AM Turing

## 1. O Jogo da Imitação

Eu proponho considerar a questão "As máquinas podem pensar?" Isso deve começar com definições do significado dos termos "máquina" e "pensar". As definições podem ser enquadradas de modo a refletir tanto quanto possível o uso normal das palavras, mas essa atitude é perigosa, se o significado das palavras "máquina" e "pensar" for encontrado examinando-se como elas são comumente usadas. É difícil escapar à conclusão de que o significado e a resposta à pergunta "As máquinas podem pensar?" deve ser procurado em uma pesquisa estatística, como uma pesquisa Gallup. Mas isso é um absurdo. Em vez de tentar tal definição, substituirei a pergunta por outra, que está intimamente relacionada a ela e expressa-se em palavras relativamente inequívocas.

A nova forma do problema pode ser descrita em termos de um jogo que chamamos de "jogo de imitação". Ele é jogado com três pessoas, um homem (A), uma mulher (B) e um interrogador (C) que pode ser de qualquer sexo. O interrogador fica em um quarto à frente dos outros dois. O objetivo do jogo para o interrogador é determinar qual dos outros dois é o homem e qual é a mulher. Ele os conhece pelos rótulos X e Y, e no final do jogo ele diz que "X é A e Y é B" ou "X é B e Y é A." O interrogador pode fazer perguntas a A e B assim:

C: X vai me dizer o comprimento do cabelo dele?

Agora suponha que X seja realmente A, então A deve responder. É o objeto de A no jogo tentar fazer com que C faça a identificação errada. Sua resposta pode, portanto, ser:

"Meu cabelo está arrepiado e os fios mais compridos medem cerca de quinze centímetros."

Para que os tons de voz não ajudem o interrogador, as respostas devem ser escritas ou, melhor ainda, datilografadas. O arranjo ideal é ter uma teleprinter comunicando entre as duas salas. Alternativamente, a pergunta e as respostas podem ser repetidas por um intermediário. O objetivo do jogo para o terceiro jogador (B) é ajudar o interrogador. A melhor estratégia para ela é, provavelmente, dar respostas verdadeiras. Ela pode acrescentar coisas como "Eu sou a mulher, não dê ouvidos a ele!" às suas respostas, mas nada valerá como o homem pode fazer observações semelhantes.

Agora fazemos a pergunta: "O que acontecerá quando uma máquina fizer o papel de A neste jogo?" Irá o interrogador decidir erroneamente quantas vezes o jogo for jogado assim como acontece quando o jogo é jogado entre um homem e uma mulher? Essas perguntas substituem nosso original, "As máquinas podem pensar?"

## 2. Crítica do Novo Problema

Além de perguntar: "Qual é a resposta para essa nova forma de pergunta", pode-se perguntar: "Essa nova pergunta é digna de ser investigada?" Esta última questão investigamos sem mais delongas, cortando assim uma regressão infinita.

O novo problema tem a vantagem de traçar uma linha bastante nítida entre as capacidades físicas e intelectuais de um homem. Nenhum engenheiro ou químico afirma ser capaz de produzir um material que seja indistinguível da pele humana. É possível que em algum momento isso possa ser feito, mas mesmo supondo que esta invenção esteja disponível, deveríamos sentir que havia pouco sentido em tentar tornar uma "máquina pensante" mais humana, vestindo-a com essa carne artificial. A forma em que definimos o problema reflete esse fato na condição que impede o interrogador de ver ou tocar os outros competidores, ou de ouvir suas vozes. Algumas outras vantagens do critério proposto podem ser mostradas por perguntas e respostas de espécimes. Portanto:

Q: Por favor, escreva-me um soneto sobre o tema da Ponte Forth.

A: Não conte comigo para isso. Eu nunca pude escrever poesia.

Q: Adicione 34957 a 70764.

A: (Pausa cerca de 30 segundos e depois dar como resposta) 105621.

Q: Você joga xadrez?

A: sim.

Q: Eu tenho K no meu K1 e não há outras peças. Você tem apenas K em K6 e R em R1. É o seu movimento. O que você toca?

A: (Após uma pausa de 15 segundos) R-R8 mate.

O método de perguntas e respostas parece adequado para introduzir quase qualquer um dos campos do esforço humano que desejamos incluir. Nós não queremos penalizar a máquina por sua incapacidade de brilhar em competições de beleza, nem penalizar um homem por perder em uma corrida contra um avião. As condições do nosso jogo tornam essas deficiências irrelevantes. As "testemunhas" podem gabar-se, se considerarem aconselhável, o quanto quiserem sobre seus encantos, força ou heroísmo, mas o interrogador não pode exigir demonstrações práticas.

O jogo talvez seja criticado pelo fato de que as probabilidades são pesadas demais contra a máquina. Se o homem tentasse fingir ser a máquina, ele faria uma exibição muito pobre. Ele seria doado imediatamente pela lentidão e imprecisão na aritmética. Podem as máquinas não realizar algo que deveria ser descrito como pensamento, mas que é muito diferente do que um homem faz? Essa objeção é muito forte, mas pelo menos podemos dizer que, se, no entanto, uma máquina pode ser construída para jogar o jogo da imitação satisfatoriamente, não precisamos ser incomodados por essa objeção.

Pode-se dizer que, ao jogar o "jogo da imitação", a melhor estratégia para a máquina talvez seja outra coisa que não a imitação do comportamento de um homem. Isso pode ser, mas acho improvável que haja algum grande efeito desse tipo. Em qualquer caso, não há intenção de investigar aqui a teoria do jogo, e será assumido que a melhor estratégia é tentar fornecer respostas que seriam naturalmente dadas por um homem.

### 3. As Máquinas Preocupadas no Jogo

A questão que colocamos em 1 não será bem definida até que tenhamos especificado o que queremos dizer com a palavra "máquina". É natural que devamos permitir que todo tipo de técnica de engenharia seja usado em nossas máquinas. Também desejamos permitir a possibilidade de que um engenheiro ou equipe de engenheiros possa construir uma máquina que funcione, mas cuja maneira de operar não possa ser satisfatoriamente descrita por seus construtores, porque eles aplicaram um método que é amplamente experimental. Finalmente, desejamos excluir das máquinas os homens nascidos da maneira usual. É difícil enquadrar as definições para satisfazer essas três condições. Poderíamos, por exemplo, insistir que a equipe de engenheiros deveria ser de um só sexo, mas isso não seria realmente satisfatório, pois é provavelmente possível criar um indivíduo completo a partir de uma única célula da pele (digamos) de um homem. Fazê-lo seria uma façanha de técnica biológica que merece o mais alto elogio, mas não estaríamos inclinados a considerá-lo como um caso de "construção de uma máquina pensante". Isso nos leva a abandonar a exigência de que todo tipo de técnica deva ser permitido. Estamos mais dispostos a fazê-lo, em vista do fato de que o interesse atual pelas "máquinas pensantes" foi despertado por um tipo particular de máquina, geralmente chamado de "computador eletrônico" ou "computador digital". Seguindo esta sugestão, permitimos que apenas computadores digitais participem do nosso jogo.

Esta restrição parece à primeira vista ser muito drástica. Tentarei mostrar que não é assim na realidade. Para fazer isso, é necessária uma breve descrição da natureza e das propriedades desses computadores.

Pode-se dizer também que essa identificação de máquinas com computadores digitais, como nosso critério de "pensar", só será insatisfatória se (ao contrário do que eu acredito), os computadores digitais não conseguirem dar uma boa demonstração no jogo.

Já existe um número de computadores digitais em funcionamento, e pode ser perguntado: "Por que não experimentar o experimento imediatamente? Seria fácil satisfazer as condições do jogo. Vários interrogadores poderiam ser usados e as estatísticas compiladas para mostrar quantas vezes a identificação correta foi dada ". A resposta curta é que não estamos perguntando se todos os computadores digitais se sairiam bem no jogo nem se os computadores atualmente disponíveis se sairiam bem, mas se há computadores imagináveis que funcionariam bem. Mas esta é apenas a resposta curta. Vamos ver essa questão em uma luz diferente depois.

#### 4. Computadores Digitais

A idéia por trás dos computadores digitais pode ser explicada dizendo que essas máquinas são destinadas a realizar quaisquer operações que possam ser feitas por um computador humano. Supõe-se que o computador humano esteja seguindo regras fixas; ele não tem autoridade para se desviar deles em qualquer detalhe. Podemos supor que essas regras sejam fornecidas em um livro, que é alterado sempre que ele é colocado em um novo emprego. Ele também tem um suprimento ilimitado de papel no qual ele faz seus cálculos. Ele também pode fazer suas multiplicações e acréscimos em uma "máquina de mesa", mas isso não é importante.

Se usarmos a explicação acima como uma definição, estaremos em perigo de circularidade de argumento. Nós evitamos isso, dando um esboço dos meios pelos quais o

efeito desejado é alcançado. Um computador digital geralmente pode ser considerado como constituído de três partes:

- (i) Armazenar.
- (ii) unidade executiva.
- (iii) controle.

A loja é uma loja de informações e corresponde ao papel do computador humano, seja este o papel no qual ele faz seus cálculos ou aquele em que seu livro de regras é impresso. Na medida em que o computador humano faz cálculos em sua conta, uma parte da loja corresponderá à sua memória.

A unidade executiva é a parte que executa as várias operações individuais envolvidas em um cálculo. O que essas operações individuais irão variar de máquina para máquina. Geralmente operações bastante longas podem ser feitas como "Multiplicar 3540675445 por 7076345687", mas em algumas máquinas apenas as mais simples, como "Anotar 0", são possíveis.

Mencionamos que o "livro de regras" fornecido ao computador é substituído na máquina por uma parte da loja. É então chamado de "tabela de instruções". É dever do controle ver que essas instruções são obedecidas corretamente e na ordem correta. O controle é construído de tal forma que isso necessariamente acontece.

A informação na loja é geralmente dividida em pacotes de tamanho moderadamente pequeno. Em uma máquina, por exemplo, um pacote pode consistir em dez dígitos decimais. Os números são atribuídos às partes da loja em que os vários pacotes de informações são armazenados, de alguma maneira sistemática. Uma instrução típica poderia dizer

"Adicione o número armazenado na posição 6809 àquele em 4302 e coloque o resultado de volta na última posição de armazenamento."

Escusado será dizer que não ocorreria na máquina expressa em inglês. Seria mais provável que fosse codificado em uma forma como 6809430217. Aqui, 17 diz qual das várias operações possíveis deve ser executada nos dois números. Neste caso, a operação é a descrita acima, "Adicionar o número..." Será notado que a instrução ocupa 10 dígitos e, assim, forma um pacote de informações, muito convenientemente. O controle normalmente tomará as instruções a serem obedecidas na ordem das posições nas quais elas estão armazenadas, mas ocasionalmente uma instrução como

"Agora obedeça a instrução armazenada na posição 5606 e continue a partir daí"

pode ser encontrado, ou novamente

"Se a posição 4505 contiver 0, obedeça a instrução armazenada em 6707, caso contrário, continue em frente."

As instruções destes últimos tipos são muito importantes porque permitem que uma sequência de operações seja substituída repetidamente até que alguma condição seja satisfeita, mas ao fazê-lo obedecer, não instruções novas em cada repetição, mas as mesmas sobre e de novo. Para tomar uma analogia doméstica. Suponha que mamãe queira que Tommy chame o sapateiro todas as manhãs a caminho da escola para ver se seus sapatos estão prontos, ela pode perguntar-lhe de novo todas as manhãs. Alternativamente, ela pode colocar um aviso de uma vez por todas no corredor que ele verá quando ele sair para a escola e que lhe diz para chamar os sapatos, e também para destruir o aviso quando ele voltar, se ele tiver os sapatos com ele. .

O leitor deve aceitá-lo como um fato de que os computadores digitais podem ser construídos e, de fato, construídos de acordo com os princípios que descrevemos, e que podem de fato imitar as ações de um computador humano muito de perto.

O livro de regras que descrevemos como sendo o nosso computador humano é, naturalmente, uma ficção conveniente. Os verdadeiros computadores humanos realmente

lembram o que eles têm que fazer. Se alguém quiser fazer uma máquina imitar o comportamento do computador humano em alguma operação complexa, é preciso perguntar a ele como é feito e depois traduzir a resposta na forma de uma tabela de instruções. A construção de tabelas de instruções é geralmente descrita como "programação". "Programar uma máquina para executar a operação A" significa colocar a tabela de instruções apropriada na máquina, de modo que ela funcione.

Uma variante interessante da idéia de um computador digital é um "computador digital com um elemento aleatório". Estes têm instruções envolvendo o lançamento de um dado ou algum processo eletrônico equivalente; Uma dessas instruções pode, por exemplo, ser: "Jogue o dado e coloque o número resultante na loja 1000". Às vezes, tal máquina é descrita como tendo livre arbítrio (embora eu mesmo não usasse essa frase). Normalmente, não é possível determinar, observando uma máquina se ela possui um elemento aleatório, pois um efeito similar pode ser produzido por dispositivos como fazer as escolhas dependem dos dígitos do decimal para.

A maioria dos computadores digitais reais tem apenas uma loja finita. Não há dificuldade teórica na ideia de um computador com um armazenamento ilimitado. Claro que apenas uma parte finita pode ter sido usada a qualquer momento. Da mesma forma, apenas uma quantidade finita pode ter sido construída, mas podemos imaginar mais e mais sendo adicionados conforme necessário. Tais computadores têm interesse teórico especial e serão chamados de computadores de capacidade infinitiva.

A ideia de um computador digital é antiga. Charles Babbage, professor de matemática da Lucasian em Cambridge de 1828 a 1839, planejou uma máquina desse tipo, chamada de Analytical Engine, mas nunca foi concluída. Embora Babbage tivesse todas as idéias essenciais, sua máquina não era, na época, uma perspectiva tão atraente. A velocidade que estaria disponível seria definitivamente mais rápida que um computador humano, mas algo como 1 000 vezes mais lenta que a máquina de Manchester, ela mesma uma das



máquinas mais lentas e modernas. O armazenamento era puramente mecânico, usando rodas e cartões.

O fato de que o Mecanismo Analítico de Babbage seria inteiramente mecânico nos ajudará a nos livrar de uma superstição. A importância é frequentemente atribuída ao fato de que os computadores digitais modernos são elétricos e que o sistema nervoso também é elétrico. Como a máquina de Babbage não era elétrica, e como todos os computadores digitais são equivalentes, vemos que esse uso de eletricidade não pode ter importância teórica. É claro que a eletricidade geralmente vem onde a sinalização rápida está em causa, de modo que não é surpreendente que a encontremos nessas duas conexões. No sistema nervoso, os fenômenos químicos são pelo menos tão importantes quanto os elétricos. Em certos computadores, o sistema de armazenamento é principalmente acústico. A característica de usar eletricidade é, portanto, vista como apenas uma semelhança muito superficial.

## 5. Universalidade de Computadores Digitais

Os computadores digitais considerados na última seção podem ser classificados entre as "máquinas de estado discreto". Estas são as máquinas que se movem por saltos ou cliques repentinos de um estado bem definido para outro. Esses estados são suficientemente diferentes para que a possibilidade de confusão entre eles seja ignorada. Estritamente falando, não existem tais máquinas. Tudo realmente se move continuamente. Mas existem muitos tipos de máquinas que podem ser pensadas como sendo máquinas de estado discreto. Por exemplo, ao considerar os interruptores para um sistema de iluminação, é uma ficção conveniente que cada interruptor deve estar definitivamente ligado ou definitivamente desligado. Deve haver posições intermediárias, mas, para a maioria dos propósitos, podemos esquecê-las. Como um exemplo de uma máquina de estado discreto, poderíamos considerar uma roda que clica em volta de 120 por segundo, mas pode ser parada

por qualquer um que possa ser operada de fora; Além disso, uma lâmpada é acender em uma das posições da roda. Esta máquina pode ser descrita abstratamente da seguinte forma. O estado interno da máquina (que é descrito pela posição da roda) pode ser  $q_1$ ,  $q_2$  ou  $q_3$ . Existe um sinal de entrada  $i_0$  ou  $i_1$  (posição de] sempre). O estado interno a qualquer momento é determinado pelo último estado e sinal de entrada de acordo com a tabela.

(TABLE DELETED)

Os sinais de saída, a única indicação visível externamente do estado interno (a luz) são descritos pela tabela

Estado  $q_1$   $q_2$   $q_3$

saída  $o_0$   $o_0$   $o_1$

Este exemplo é típico de máquinas de estado discreto. Eles podem ser descritos por essas tabelas, desde que tenham apenas um número finito de estados possíveis.

Parecerá que, dado o estado inicial da máquina e os sinais de entrada, é sempre possível prever todos os estados futuros. Isso lembra a visão de Laplace de que a partir do estado completo do universo em um momento, conforme descrito pelas posições e velocidades de todas as partículas, deve ser possível prever todos os estados futuros. A previsão que estamos considerando, no entanto, está mais próxima da praticidade do que aquela considerada por Laplace. O sistema do "universo como um todo" é tal que erros bem pequenos nas condições iniciais podem ter um efeito esmagador em um momento posterior. O deslocamento de um único elétron por um bilionésimo de centímetro em um momento pode fazer a diferença entre um homem ser morto por uma avalanche um ano depois, ou escapar. É uma propriedade essencial dos sistemas mecânicos que chamamos de "máquinas

de estado discreto" que esse fenômeno não ocorre. Mesmo quando consideramos as máquinas físicas reais em vez das máquinas idealizadas, o conhecimento razoavelmente preciso do estado em um momento produz conhecimento razoavelmente preciso em qualquer número de etapas posteriores.

Como mencionamos, os computadores digitais estão dentro da classe de máquinas de estado discreto. Mas o número de estados dos quais tal máquina é capaz é geralmente muito grande. Por exemplo, o número da máquina que agora trabalha em Manchester é de cerca de 2 165.000 , ou seja, cerca de 10 50.000. Compare isso com nosso exemplo da roda de clique descrita acima, que tinha três estados. Não é difícil ver por que o número de estados deveria ser tão imenso. O computador inclui uma loja correspondente ao papel usado por um computador humano. Deve ser possível escrever na loja qualquer uma das combinações de símbolos que possam ter sido escritos no papel. Por simplicidade, suponha que apenas dígitos de 0 a 9 sejam usados como símbolos. Variações no manuscrito são ignoradas. Suponha que o computador receba 100 folhas de papel, cada uma contendo 50 linhas, cada uma com espaço para 30 dígitos. Então o número de estados é  $10^{100 \times 50 \times 30}$  ou seja, 10 150.000. Isso é sobre o número de estados de três máquinas de Manchester juntas. O logaritmo da base dois do número de estados é geralmente chamado de "capacidade de armazenamento" da máquina. Assim, a máquina de Manchester tem uma capacidade de armazenamento de cerca de 165.000 e a máquina de rodas do nosso exemplo, cerca de 1.6. Se duas máquinas forem juntas, suas capacidades devem ser adicionadas para obter a capacidade da máquina resultante. Isso leva à possibilidade de declarações como "A máquina de Manchester contém 64 trilhas magnéticas, cada uma com capacidade para 2560, oito tubos eletrônicos com uma capacidade de 1280. O armazenamento diverso é de cerca de 300, perfazendo um total de 174.380".

Dada a tabela correspondente a uma máquina de estados discretos, é possível prever o que ela fará. Não há razão para que esse cálculo não seja realizado por meio de um

computador digital. Desde que pudesse ser realizado com bastante rapidez, o computador digital poderia imitar o comportamento de qualquer máquina de estado discreto. O jogo de imitação poderia então ser jogado com a máquina em questão (como B) e o imitando computador digital (como A) e o interrogador seria incapaz de distingui-los. É claro que o computador digital deve ter uma capacidade de armazenamento adequada, além de funcionar com velocidade suficiente. Além disso, deve ser programado novamente para cada nova máquina que se deseja imitar.

Essa propriedade especial dos computadores digitais, que podem imitar qualquer máquina de estado discreto, é descrita dizendo que são máquinas universais. A existência de máquinas com essa propriedade tem a importante consequência de que, considerando as diferenças de velocidade, não é necessário projetar várias máquinas novas para realizar vários processos computacionais. Todos eles podem ser feitos com um computador digital, adequadamente programado para cada caso. Pode-se ver que, como consequência disso, todos os computadores digitais são, em certo sentido, equivalentes.

Podemos agora considerar novamente o ponto levantado no final do §3. Foi sugerido experimentalmente que a pergunta "As máquinas podem pensar?" deve ser substituído por "Existem computadores digitais imagináveis que fariam bem no jogo de imitação?" Se quisermos, podemos tornar isto superficialmente mais geral e perguntar "Existem máquinas de estado discreto que se saiam bem?" Mas, em vista da propriedade da universalidade, vemos que qualquer uma dessas questões é equivalente a isso: "Vamos fixar nossa atenção em um computador digital específico". C. É verdade que, ao modificar esse computador, ter um armazenamento adequado, aumentando sua velocidade adequadamente? de ação, e fornecendo-lhe um programa apropriado, C pode ser feito para jogar satisfatoriamente a parte de A no jogo de imitação,

## 6. Opiniões contrárias sobre a questão principal

Podemos agora considerar que o terreno foi limpo e estamos prontos para prosseguir com o debate sobre a nossa questão: "As máquinas podem pensar?" e a variante citada no final da última seção. Não podemos abandonar completamente a forma original do problema, pois as opiniões serão diferentes quanto à adequação da substituição e devemos pelo menos ouvir o que deve ser dito nessa conexão.

Isso simplificará as coisas para o leitor se eu explicar primeiro minhas próprias crenças sobre o assunto. Considere primeiro a forma mais precisa da questão. Acredito que daqui a cerca de cinquenta anos será possível, programar computadores, com uma capacidade de armazenamento de cerca de 10<sup>9</sup>, para fazê-los jogar tão bem o jogo de imitação que um interrogador médio não terá mais de 70% de chance de fazer a identificação correta após cinco minutos de questionamento. A pergunta original, "As máquinas podem pensar?" Eu acredito ser muito sem sentido para merecer discussão. Não obstante, acredito que, no final do século, o uso das palavras e da opinião geral e educada terá mudado tanto que poderemos falar de máquinas pensando sem esperar que sejam contraditas. Acredito ainda que nenhum propósito útil é servido pela ocultação dessas crenças. A visão popular de que os cientistas procedem inexoravelmente de um fato bem estabelecido para um fato bem estabelecido, nunca sendo influenciados por qualquer conjectura aperfeiçoada, está bastante equivocada. Desde que seja esclarecido quais são os fatos comprovados e quais são as conjecturas, nenhum dano pode resultar. Conjecturas são de grande importância, uma vez que sugerem linhas úteis de pesquisa.

Eu agora procuro considerar opiniões opostas às minhas.

#### (1) A objeção teológica

Pensar é uma função da alma imortal do homem. Deus deu uma alma imortal a todo homem e mulher, mas não a qualquer outro animal ou a máquinas. Por isso, nenhum animal ou máquina pode pensar.

Não posso aceitar nada disso, mas tentarei responder em termos teológicos. Eu deveria achar o argumento mais convincente se os animais fossem classificados com homens, pois há uma diferença maior, em minha opinião, entre o típico animado e o inanimado do que entre o homem e os outros animais. O caráter arbitrário da visão ortodoxa se torna mais claro se considerarmos como ela pode parecer para um membro de alguma outra comunidade religiosa. Como os cristãos consideram a visão muçulmana de que as mulheres não têm alma? Mas deixemos este ponto de lado e voltemos ao argumento principal. Parece-me que o argumento citado acima implica uma séria restrição da onipotência do Todo-Poderoso. Admite-se que há certas coisas que Ele não pode fazer, como fazer um igual a dois, mas não devemos acreditar que Ele tem liberdade para conferir uma alma a um elefante se Ele achar adequado? Poderíamos esperar que Ele só exercesse esse poder em conjunção com uma mutação que proporcionasse ao elefante um cérebro apropriadamente melhorado para atender às necessidades desse tipo [. Um argumento de forma exatamente similar pode ser feito para o caso de máquinas. Pode parecer diferente porque é mais difícil "engolir". Mas isso realmente significa apenas que achamos que seria menos provável que Ele considerasse as circunstâncias adequadas para conferir uma alma. As circunstâncias em questão são discutidas no restante deste artigo. Na tentativa de construir tais máquinas, não devemos estar usurpando irremediavelmente Seu poder de criar almas, mais do que estamos na procriação de crianças:

No entanto, isso é mera especulação. Não estou muito impressionado com os argumentos teológicos, sejam quais forem os que possam ser usados para apoiar. Tais argumentos foram freqüentemente encontrados insatisfatórios no passado. Na época de Galileu, argumentou-se que os textos "E o sol parou ... e se preparou para não descer por um dia inteiro" (Josué x. 13) e "Ele lançou os fundamentos da terra, que não deve se mover a qualquer momento" (Salmo 5) foram uma refutação adequada da teoria copernicana. Com

o nosso conhecimento atual, tal argumento parece fútil. Quando esse conhecimento não estava disponível, causou uma impressão bem diferente.

## (2) A objeção "Heads in the Sand"

As consequências do pensamento das máquinas seriam terríveis demais. Vamos esperar e acreditar que eles não podem fazer isso. "

Esse argumento raramente é expresso de forma tão abertamente quanto na forma acima. Mas isso afeta a maioria de nós que pensa sobre isso. Nós gostamos de acreditar que o homem é de alguma maneira sutil superior ao resto da criação. É melhor que se mostre que ele é necessariamente superior, pois não há perigo de ele perder sua posição de comando. A popularidade do argumento teológico está claramente conectada com esse sentimento. É provável que seja bastante forte em pessoas intelectuais, já que eles valorizam o poder de pensar mais do que outros, e estão mais inclinados a basear sua crença na superioridade do Homem nesse poder.

Não creio que esse argumento seja suficientemente substancial para exigir refutação. O consolo seria mais apropriado: talvez isso devesse ser buscado na transmigração das almas.

## (3) A objeção matemática

Existem vários resultados da lógica matemática que podem ser usados para mostrar que existem limitações aos poderes das máquinas de estados discretos. O mais conhecido desses resultados é conhecido como o teorema de Gödel (1931) e mostra que em qualquer sistema lógico suficientemente poderoso podem ser formuladas declarações que não podem ser provadas nem contestadas dentro do sistema, a menos que o próprio sistema seja inconsistente. Há outros, em alguns aspectos similares, resultados devidos a Church (1936), Kleene (1935), Rosser e Turing (1937). O último resultado é o mais conveniente a ser considerado, já que se refere diretamente a máquinas, enquanto os outros só podem

ser usados em um argumento comparativamente indireto: por exemplo, se Gödel ' O teorema deve ser usado. Além disso, precisamos de alguns meios para descrever sistemas lógicos em termos de máquinas e máquinas em termos de sistemas lógicos. O resultado em questão refere-se a um tipo de máquina que é essencialmente um computador digital com uma capacidade infinita. Afirma que existem certas coisas que tal máquina não pode fazer. Se for montado para dar respostas a perguntas como no jogo de imitação, haverá algumas perguntas para as quais ele dará uma resposta errada ou não responderá, por mais tempo que seja permitido para uma resposta. Pode haver, é claro, muitas dessas questões, e perguntas que não podem ser respondidas por uma máquina podem ser satisfatoriamente respondidas por outra. É claro que estamos supondo que as questões são do tipo que responde " Sim "ou" Não "é apropriado, em vez de perguntas como" O que você acha do Picasso? "As perguntas que sabemos que as máquinas devem falhar são desse tipo," Considere a máquina especificada da seguinte maneira. . . . Será que esta máquina responderá "Sim" a alguma pergunta? "Os pontos devem ser substituídos por uma descrição de alguma máquina em um formulário padrão, que poderia ser algo como o usado no § 5. Quando a máquina descrita tiver um certo comparativamente simples Em relação à máquina que está sob interrogação, pode-se mostrar que a resposta está errada ou não está próxima. Esse é o resultado matemático: argumenta-se que prova uma deficiência de máquinas à qual o intelecto humano não está sujeito. em vez de perguntas como "O que você acha de Picasso?" As perguntas que nós sabemos que as máquinas devem falhar são desse tipo, "considere a máquina especificada da seguinte maneira ... Essa máquina responderá alguma vez 'Sim' a alguma pergunta?" Os pontos devem ser substituídos por uma descrição de alguma máquina em um formulário padrão, que pode ser algo como o usado em §5. Quando a máquina descrita tem uma certa relação comparativamente simples com a máquina que está sob interrogação, pode-se mostrar que a resposta está errada



ou não está próxima. Este é o resultado matemático: argumenta-se que prova uma deficiência de máquinas à qual o intelecto humano não está sujeito. em vez de perguntas como "O que você acha de Picasso?" As perguntas que nós sabemos que as máquinas devem falhar são desse tipo, "considere a máquina especificada da seguinte maneira ... Essa máquina responderá alguma vez 'Sim' a alguma pergunta?" Os pontos devem ser substituídos por uma descrição de alguma máquina em um formulário padrão, que pode ser algo como o usado em §5. Quando a máquina descrita tem uma certa relação comparativamente simples com a máquina que está sob interrogação, pode-se mostrar que a resposta está errada ou não está próxima. Este é o resultado matemático: argumenta-se que prova uma deficiência de máquinas à qual o intelecto humano não está sujeito. As perguntas que nós sabemos que as máquinas devem falhar são desse tipo, "considere a máquina especificada da seguinte maneira ... Essa máquina responderá alguma vez 'Sim' a alguma pergunta?" Os pontos devem ser substituídos por uma descrição de alguma máquina em um formulário padrão, que pode ser algo como o usado em §5. Quando a máquina descrita tem uma certa relação comparativamente simples com a máquina que está sob interrogação, pode-se mostrar que a resposta está errada ou não está próxima. Este é o resultado matemático: argumenta-se que prova uma deficiência de máquinas à qual o intelecto humano não está sujeito. Os pontos devem ser substituídos por uma descrição de alguma máquina em um formulário padrão, que pode ser algo como o

usado em §5. Quando a máquina descrita tem uma certa relação comparativamente simples com a máquina que está sob interrogação, pode-se mostrar que a resposta está errada ou não está próxima. Este é o resultado matemático: argumenta-se que prova uma deficiência de máquinas à qual o intelecto humano não está sujeito. Os pontos devem ser substituídos por uma descrição de alguma máquina em um formulário padrão, que pode ser algo como o usado em §5. Quando a máquina descrita tem uma certa relação comparativamente simples com a máquina que está sob interrogação, pode-se mostrar que a resposta está errada ou não está próxima. Este é o resultado matemático: argumenta-se que prova uma deficiência de máquinas à qual o intelecto humano não está sujeito.

A resposta curta a este argumento é que, embora seja estabelecido que há limitações para os Poderes. Se qualquer máquina particular, foi apenas declarado, sem qualquer tipo de prova, que tais limitações não se aplicam ao intelecto humano. Mas não creio que essa visão possa ser descartada com tanta leviandade. Sempre que uma dessas máquinas faz a pergunta crítica apropriada, e dá uma resposta definitiva, sabemos que essa resposta deve estar errada, e isso nos dá um certo sentimento de superioridade. Isso é sentimento ilusório? É sem dúvida genuína, mas não creio que deva ser dada muita importância a ela. Com demasiada frequência, nós respondemos erroneamente a perguntas para nos justificarmos em ficar muito satisfeitos com essa evidência de falibilidade da parte das máquinas. Mais distante, nossa superioridade só pode ser sentida em tal ocasião em relação à única máquina sobre a qual marcamos nosso pequeno triunfo. Não haveria nenhuma questão de triunfar simultaneamente sobre todas as máquinas. Em suma, então, pode haver homens mais espertos do que qualquer máquina, mas, novamente, pode haver outras máquinas mais inteligentes, e assim por diante.

Aqueles que defendem o argumento matemático, na minha opinião, prefeririam aceitar o jogo da imitação como base para discussão. Aqueles que acreditam nas duas objeções anteriores provavelmente não estariam interessados em nenhum critério.

#### (4) O Argumento da Consciência

Este argumento é muito bem expresso na Lister Oration do Professor Jefferson em 1949, da qual cito. "Não até que uma máquina possa escrever um soneto ou compor um concerto por causa de pensamentos e emoções sentidas, e não pela queda accidental de símbolos, poderíamos concordar que máquina é igual a cérebro - isto é, não apenas escrevê-lo, mas saber que ele escreveu Nenhum mecanismo pode sentir (e não meramente artificialmente sinalizar, um artifício fácil) prazer em seus sucessos, pesar quando suas válvulas se fundem, ser aquecido por lisonjas, ser feito miserável por seus erros, ser encantado com sexo, ficar zangado ou deprimido quando não pode conseguir o que quer. "

Este argumento parece ser uma negação da validade do nosso teste. De acordo com a forma mais extrema dessa visão, a única maneira pela qual alguém poderia ter certeza de que a máquina pensa é ser a máquina e sentir-se pensando. Poder-se-ia então descrever esses sentimentos para o mundo, mas é claro que ninguém teria justificativa para tomar qualquer aviso. Da mesma forma, de acordo com essa visão, a única maneira de saber que um homem pensa é ser esse homem em particular. Na verdade, é o ponto de vista do solipsista. Pode ser a visão mais lógica, mas dificulta a comunicação de idéias. A é capaz de acreditar "A pensa, mas B não", enquanto B acredita que "B pensa, mas A não." em vez de discutir continuamente sobre esse ponto, é comum ter a convenção educada que todos pensam.

Tenho certeza de que o professor Jefferson não deseja adotar o ponto de vista extremo e solipsista. Provavelmente ele estaria disposto a aceitar o jogo de imitação como teste. O jogo (com o jogador B omitido) é freqüentemente usado na prática sob o nome de viva voce para descobrir se alguém realmente entende alguma coisa ou se "aprendeu de maneira paroquial". Vamos ouvir uma parte de tal viva voce :

Interrogador: Na primeira linha do seu soneto, que diz "Devo comparar-te a um dia de verão", o "dia de primavera" não seria tão bom ou melhor?

Testemunha: não escaneia.

Interrogador: Que tal "um dia de inverno"? Isso escaneia bem.

Testemunha: Sim, mas ninguém quer ser comparado a um dia de inverno.

Interrogador: Você diria que o Sr. Pickwick lembrou você do Natal?

Testemunha: De certa forma.

Interrogador: No entanto, o Natal é um dia de inverno, e eu não acho que o Sr. Pickwick se importaria com a comparação.

Testemunha: Eu não acho que você é sério. Em um dia de inverno, significa um dia típico de inverno, em vez de um dia especial como o Natal.

E assim por diante, o que o Professor Jefferson diria se a máquina de escrever sonetos fosse capaz de responder assim no viva voce ? Não sei se ele consideraria a máquina como "meramente sinalizando artificialmente" essas respostas, mas se as respostas fossem tão satisfatórias e sustentadas quanto na passagem acima, não creio que ele as despreveria como "um artifício fácil". Esta frase é, penso eu, destinada a cobrir tais dispositivos como a inclusão na máquina de um registro de alguém lendo um soneto, com a mudança apropriada para ligá-lo de tempos em tempos.

Em suma, penso que a maioria daqueles que defendem o argumento da consciência poderiam ser persuadidos a abandoná-lo, em vez de serem forçados a entrar na posição de solipsista. Eles provavelmente estarão dispostos a aceitar nosso teste.

Não quero dar a impressão de que penso que não há mistério sobre a consciência. Existe, por exemplo, um paradoxo ligado a qualquer tentativa de localizá-lo. Mas eu

não acho que esses mistérios necessariamente precisem ser resolvidos antes que possamos responder à questão com a qual estamos preocupados neste artigo.

#### (5) Argumentos de várias deficiências

Esses argumentos assumem a forma: "Eu garanto que você pode fazer as máquinas fazerem todas as coisas que você mencionou, mas você nunca será capaz de fazer uma para fazer X". Numerosos recursos X são sugeridos nesta conexão Eu ofereço uma seleção:

Seja gentil, engenhoso, bonito, simpático, tenha iniciativa, tenha um senso de humor, diga certo e errado, cometa erros, se apaixone, desfrute de morangos e creme, faça alguém se apaixonar por ele, aprenda com a experiência, use palavras propriamente, seja o sujeito de seu próprio pensamento, tenha tanta diversidade de comportamento quanto um homem, faça algo realmente novo.

Geralmente, não há suporte para essas instruções. Eu acredito que eles são principalmente fundados no princípio da indução científica. Um homem viu milhares de máquinas em sua vida. Do que ele vê deles, ele tira uma série de conclusões gerais. Eles são feios, cada um é projetado para um propósito muito limitado, quando necessário para um propósito minuciosamente diferente eles são inúteis, a variedade de comportamento de qualquer um deles é muito pequena, etc., etc. Naturalmente ele conclui que estas são propriedades necessárias de máquinas em geral. Muitas dessas limitações estão associadas à capacidade de armazenamento muito pequena da maioria das máquinas. (Eu estou supondo que a idéia de capacidade de armazenamento é estendida de alguma forma para cobrir máquinas que não sejam máquinas de estado discreto. A definição exata não importa, já que nenhuma precisão matemática é reivindicada na presente discussão. Alguns anos atrás, quando muito pouco tinha sido ouvido sobre computadores digitais, era possível obter muita incredulidade sobre eles, se alguém mencionasse suas propriedades sem descrever sua

construção. Isso foi presumivelmente devido a uma aplicação semelhante do princípio da indução científica. Essas aplicações do princípio são, em grande parte, inconscientes. Quando uma criança queimada teme o fogo e mostra que ele teme evitá-lo, deve dizer que ele estava aplicando a indução científica. (Eu poderia, claro, também descrever seu comportamento de muitas outras maneiras.) As obras e os costumes da humanidade não parecem ser um material muito adequado ao qual se possa aplicar a indução científica. Uma parte muito grande do espaço-tempo deve ser investigada, se resultados confiáveis forem obtidos. Caso contrário, podemos (como faz a maioria dos ingleses de crianças) decidir que todos falam inglês e que é bobagem aprender francês.

Há, no entanto, observações especiais a serem feitas sobre muitas das deficiências mencionadas. A incapacidade de desfrutar de morangos e creme pode ter atingido o leitor como frívolo. Possivelmente uma máquina poderia ser feita para apreciar este delicioso prato, mas qualquer tentativa de fazer isso seria idiota. O que é importante sobre essa deficiência é que ela contribui para algumas das outras deficiências, por exemplo, para a dificuldade do mesmo tipo de amizade que ocorre entre homem e máquina, entre homem branco e homem branco, ou entre homem negro e homem negro.

A alegação de que "máquinas não podem cometer erros" parece curiosa. Alguém é tentado a replicar: "Eles são o pior para isso?" Mas vamos adotar uma atitude mais simpática e tentar ver o que realmente significa. Eu acho que essa crítica pode ser explicada em termos do jogo de imitação. Alega-se que o interrogador poderia distinguir a máquina do homem simplesmente definindo-os uma série de problemas na aritmética. A máquina seria desmascarada devido à sua precisão mortal. A resposta para isso é simples. A máquina (programada para jogar o jogo) não tentaria dar as respostas certas para os problemas aritméticos. Ele deliberadamente apresentaria erros de uma maneira calculada para confundir o interrogador. Uma falha mecânica provavelmente se mostraria através de uma de-

cisão inadequada quanto ao tipo de erro a ser cometido na aritmética. Mesmo essa interpretação da crítica não é suficientemente simpática. Mas não podemos arcar com o espaço para ir mais além. Parece-me que essa crítica depende de uma confusão entre dois tipos de erro. Podemos chamá-los de "erros de funcionamento" e "erros de conclusão". Erros de funcionamento são devidos a alguma falha mecânica ou elétrica que faz com que a máquina se comporte diferente do que foi projetado para fazer. Nas discussões filosóficas, gosta-se de ignorar a possibilidade de tais erros; por isso, estamos discutindo "máquinas abstratas". Essas máquinas abstratas são ficções matemáticas, e não objetos físicos. Por definição, eles são incapazes de erros de funcionamento. Nesse sentido, podemos dizer verdadeiramente que "as máquinas nunca podem cometer erros". Erros de conclusão só podem surgir quando algum significado é anexado aos sinais de saída da máquina. A máquina pode, por exemplo, digitar equações matemáticas ou frases em inglês. Quando uma proposição falsa é digitada, dizemos que a máquina cometeu um erro de conclusão. Não há razão alguma para dizer que uma máquina não pode cometer esse tipo de erro. Pode não fazer nada, mas digitar repetidamente "O = eu". Para dar um exemplo menos perverso, pode haver algum método para tirar conclusões por indução científica. Devemos esperar que tal método leve ocasionalmente a resultados errados.

A alegação de que uma máquina não pode ser objeto de seu próprio pensamento pode, naturalmente, ser respondida apenas se puder ser demonstrado que a máquina tem algum pensamento com algum assunto. No entanto, "o assunto das operações de uma máquina" parece significar algo, pelo menos para as pessoas que lidam com isso. Se, por exemplo, a máquina estivesse tentando encontrar uma solução da equação  $x^2 - 40x - 11 = 0$ , seria tentado descrever essa equação como parte do assunto da máquina naquele momento. Nesse tipo de sentido, uma máquina, sem dúvida, pode ser seu próprio assunto. Pode ser usado para ajudar na criação de seus próprios programas ou para prever o efeito de alterações em sua própria estrutura. Ao observar os resultados de seu próprio

comportamento, ele pode modificar seus próprios programas de modo a atingir uma finalidade mais efetiva. Estas são possibilidades do futuro próximo, em vez de sonhos utópicos.

A crítica de que uma máquina não pode ter muita diversidade de comportamento é apenas uma maneira de dizer que ela não pode ter muita capacidade de armazenamento. Até muito recentemente, uma capacidade de armazenamento de até mil dígitos era muito rara.

As críticas que estamos considerando aqui são frequentemente formas disfarçadas do argumento da consciência. Normalmente, se alguém sustenta que uma máquina pode fazer uma dessas coisas, e descreve o tipo de método que a máquina poderia usar, não se fará muito de um impressão. Acredita-se que o método da telha (seja ela qual for, pois deve ser mecânico) é realmente bastante básico. Compare os parênteses na declaração de Jefferson citada na página 22.

#### (6) Objeção de Lady Lovelace

Nossas informações mais detalhadas do Mecanismo Analítico de Babbage vêm de um livro de memórias de Lady Lovelace (1842). Nele ela declara: "O Mecanismo Analítico não tem pretensões de originar nada. Ele pode fazer o que quisermos para encomendá-lo." Esta afirmação é citada por Hartree (1949), que acrescenta: "Isto não implica que possa não ser possível construir equipamentos eletrônicos que "pensem por si", ou nos quais, em termos biológicos, poder-se-ia estabelecer um reflexo condicionado, que serviria de base para o "aprendizado". Se isso é possível, em princípio ou não, é uma questão estimulante e excitante, sugerida por alguns desses desenvolvimentos recentes. Mas não parecia que as máquinas construídas ou projetadas na época tivessem essa propriedade".

Eu estou em completo acordo com Hartree sobre isso. Será notado que ele não afirma que as máquinas em questão não obtiveram a propriedade, mas sim que a evidência disponível para Lady Lovelace não a encorajou a acreditar que a possuíam. É bem possível



que as máquinas em questão tenham, de certo modo, adquirido essa propriedade. Suponha que alguma máquina de estado discreto tenha a propriedade. O Mecanismo Analítico era um computador digital universal, de modo que, se sua capacidade de armazenamento e velocidade fossem adequadas, poderia ser feita uma programação adequada para imitar a máquina em questão. Provavelmente este argumento não ocorreu à condessa ou a Babbage. Em qualquer caso, não havia obrigação de reivindicar tudo o que poderia ser reivindicado.

Toda essa questão será considerada novamente sob o título de máquinas de aprendizagem.

Uma variante da objeção de Lady Lovelace afirma que uma máquina "nunca pode fazer algo realmente novo". Isto pode ser defendido por um momento com a serra: "Não há nada de novo sob o sol". Quem pode ter certeza de que o "trabalho original" que ele realizou não foi simplesmente o crescimento da semente plantada nele pelo ensino, ou o efeito de seguir princípios gerais bem conhecidos. Uma melhor variante da objeção diz que uma máquina nunca pode "nos pegar de surpresa". Essa afirmação é um desafio mais direto e pode ser atendida diretamente. Máquinas me pegam de surpresa com grande frequência. Isso ocorre principalmente porque não faço cálculo suficiente para decidir o que esperar que eles façam, ou melhor, porque, embora eu faça um cálculo, faço isso de maneira apressada e desleixada, assumindo riscos. Talvez eu diga a mim mesmo: "Suponho que a Voltagem aqui deveria ser a mesma de lá: de qualquer maneira, vamos supor que é." Naturalmente, muitas vezes estou errado, e o resultado é uma surpresa para mim, pois no momento em que o experimento é feito, essas suposições foram esquecidas. Essas admissões me deixam aberto para palestras sobre o assunto de minhas maneiras cruéis, mas não duvido da minha credibilidade quando testemunho as surpresas que sinto.

Não espero que esta resposta silencie meu crítico. Ele provavelmente dirá que as surpresas são devidas a algum ato mental criativo da minha parte e não refletem nenhum

crédito na máquina. Isso nos leva de volta ao argumento da consciência e longe da ideia de surpresa. É uma linha de argumentação que devemos considerar fechada, mas talvez valha a pena observar que a apreciação de algo tão surpreendente exige tanto de um "ato mental criativo" se o evento surpreendente se origina de um homem, um livro, uma máquina ou qualquer coisa. outro.

A visão de que as máquinas não podem dar origem a surpresas deve-se, acredito, a uma falácia à qual filósofos e matemáticos estão particularmente sujeitos. Esta é a suposição de que, assim que um fato é apresentado a uma mente, todas as conseqüências desse fato surgem na mente simultaneamente com ele. É uma suposição muito útil em muitas circunstâncias, mas é fácil esquecer que é falsa. Uma conseqüência natural de fazê-lo é que, então, pressupõe-se que não há virtude no mero resultado das conseqüências de dados e princípios gerais.

#### (7) Argumento da Continuidade no Sistema Nervoso

O sistema nervoso certamente não é uma máquina de estado discreto. Um pequeno erro na informação sobre o tamanho de um impulso nervoso que incide sobre um neurônio pode fazer uma grande diferença no tamanho do impulso de saída. Pode-se argumentar que, sendo assim, não se pode esperar ser capaz de imitar o comportamento do sistema nervoso com um sistema de estado discreto.

É verdade que uma máquina de estado discreto deve ser diferente de uma máquina contínua. Mas se aderirmos às condições do jogo de imitação, o interrogador não poderá tirar vantagem dessa diferença. A situação pode ser esclarecida se considerarmos a máquina contínua mais simples. Um analisador diferencial fará muito bem. (Um analisador diferencial é um certo tipo de máquina que não é do tipo de estado discreto usado para alguns tipos de cálculo.) Alguns deles fornecem suas respostas em uma forma digitada e, portanto,

são adequados para participar do jogo. Não seria possível para um computador digital prever exatamente quais respostas o analisador diferencial daria a um problema, mas seria perfeitamente capaz de dar o tipo certo de resposta. Por exemplo, se solicitado a dar o valor de (na verdade, cerca de 3,1416) seria razoável escolher aleatoriamente entre os valores 3,12, 3,13, 3,14, 3,15, 3,16 com as probabilidades de 0,05, 0,15, 0,55, 0,19, 0,06 (digamos). Nestas circunstâncias, seria muito difícil para o interrogador distinguir o analisador diferencial do computador digital.

#### (8) O argumento da informalidade do comportamento

Não é possível produzir um conjunto de regras que pretendem descrever o que um homem deve fazer em todas as circunstâncias possíveis. Pode-se, por exemplo, ter uma regra de que uma é parar quando se vê um semáforo vermelho, e ir se alguém vir um verde, mas e se por alguma falha ambos aparecerem juntos? Pode-se talvez decidir que é mais seguro parar. Mas algumas dificuldades posteriores podem surgir dessa decisão mais tarde. Tentar fornecer regras de conduta para cobrir todas as eventualidades, mesmo as decorrentes dos semáforos, parece ser impossível. Com tudo isso eu concordo.

A partir disso, argumenta-se que não podemos ser máquinas. Tentarei reproduzir o argumento, mas temo que dificilmente eu faça justiça. Parece rodar algo assim. "se cada homem tivesse um conjunto definido de regras de conduta pelas quais ele regulasse sua vida, ele não seria melhor do que uma máquina. Mas não existem tais regras, então os homens não podem ser máquinas." O meio não distribuído é gritante. Eu não acho que o argumento seja sempre assim, mas acredito que este é o argumento usado, no entanto. No entanto, pode haver uma certa confusão entre "regras de conduta" e "leis de comportamento" para nublificar o problema. Por "regras de conduta" quero dizer preceitos como "Pare se você vir luzes vermelhas", sobre as quais alguém pode agir, e das quais uma pode ser consciente. Por "

Podemos demonstrar com mais força que qualquer declaração desse tipo seria injustificada. Pois suponhamos que poderíamos ter certeza de encontrar tais leis se elas existissem. Então, dada uma máquina de estado discreto, certamente seria possível descobrir, por observação, o suficiente para prever seu comportamento futuro, e isso dentro de um prazo razoável, digamos, de mil anos. Mas isso não parece ser o caso. Instalei no computador de Manchester um pequeno programa usando apenas 1.000 unidades de armazenamento, em que a máquina fornecida com um número de dezesseis dígitos responde a outra em dois segundos. Eu desafiaria qualquer um a aprender com essas respostas o suficiente sobre o programa para poder prever qualquer resposta a valores não testados.

#### (9) O argumento da percepção extra-sensorial

Presumo que o leitor esteja familiarizado com a ideia de percepção extra-sensorial e o significado dos quatro itens dela, isto é, telepatia, clarividência, precognição e psicocinese. Esses fenômenos perturbadores parecem negar todas as nossas idéias científicas usuais. Como gostaríamos de os desacreditar! Infelizmente, a evidência estatística, pelo menos por telepatia, é esmagadora. É muito difícil reorganizar as idéias de alguém de modo a encaixar esses novos fatos. Uma vez aceitos, não parece um grande passo acreditar em fantasmas e fantasmas. A idéia de que nossos corpos se movem simplesmente de acordo com as leis conhecidas da física, junto com alguns outros ainda não descobertos, mas um pouco semelhantes, seria um dos primeiros a ir.

Este argumento é para minha mente bastante forte. Pode-se dizer, em resposta, que muitas teorias científicas parecem permanecer viáveis na prática, apesar de colidirem com ESP; que, na verdade, é possível se dar muito bem se alguém se esquece disso. Este é um conforto bastante frio, e tememos que o pensamento seja apenas o tipo de fenômeno em que o PES pode ser especialmente relevante.

Um argumento mais específico baseado no PES poderia ser o seguinte: "Vamos jogar o jogo da imitação, usando como testemunhas um homem que é bom como um receptor telepático, e um computador digital. O interrogador pode fazer perguntas como: cartão na minha mão direita pertence a? O homem, por telepatia ou clarividência, dá a resposta correta 130 vezes em 400. A máquina só pode adivinhar aleatoriamente e talvez acerte a razão, de modo que o interrogador faça a identificação correta. Existe uma possibilidade interessante que se abre aqui. Suponha que o computador digital contenha um gerador de números aleatórios. Então, será natural usar isso para decidir qual resposta dar. Mas então o gerador de números aleatórios estará sujeito aos poderes psicocinéticos do interrogador. Talvez essa psicocinese possa fazer com que a máquina adivinhe com mais frequência do que seria esperado em um cálculo de probabilidade, de modo que o interrogador ainda possa ser incapaz de fazer a identificação correta. Por outro lado, ele pode ser capaz de adivinhar sem questionamentos, por clarividência. Com ESP, tudo pode acontecer.

Se a telepatia for admitida, será necessário aumentar o nosso teste. A situação poderia ser considerada análoga àquela que ocorreria se o interrogador estivesse falando sozinho e um dos competidores estivesse escutando com o ouvido na parede. Colocar os competidores em uma "sala à prova de telepatia" satisfaria todos os requisitos.

## 7. Máquinas de Aprendizagem

O leitor terá antecipado que não tenho argumentos muito convincentes de natureza positiva para apoiar meus pontos de vista. Se eu tivesse, eu não deveria ter tido tanta dor para apontar as falácias em visões contrárias. Tal evidência como eu darei agora.

Voltemos por um momento à objeção de Lady Lovelace, que afirmou que a máquina só pode fazer o que lhe dissermos. Pode-se dizer que um homem pode "injetar" uma idéia na máquina, e ela responderá até certo ponto e depois cairá em quiescência, como uma corda de piano atingida por um martelo. Outro símile seria uma pilha atômica de tamanho

menor do que o crítico: uma ideia injetada deve corresponder a um nêutron que entra na pilha de fora. Cada um desses nêutrons causará uma certa perturbação que eventualmente desaparece. Se, no entanto, o tamanho da pilha for suficientemente aumentado, a perturbação do pneu causada por um tal nêutron entrará provavelmente aumentando até que toda a pilha seja destruída. Existe um fenômeno correspondente para mentes, e existe um para máquinas? Parece haver um para a mente humana. A maioria deles parece ser "subcrítica", isto é, corresponder nessa analogia a pilhas de tamanho subcrítico. Uma ideia apresentada a tal mente, em média, dará origem a menos de uma ideia em resposta. Uma pequena proporção é supercrítica. Uma ideia apresentada a tal mente que pode dar origem a toda uma "teoria" que consiste em ideias secundárias, terciárias e mais remotas. As mentes dos animais parecem ser muito subcríticas. Aderindo a essa analogia, perguntamos: "Pode uma máquina ser feita para ser supercrítica?"

A analogia da "pele de uma cebola" também é útil. Ao considerar as funções da mente ou do cérebro, encontramos certas operações que podemos explicar em termos puramente mecânicos. Isso que dizemos não corresponde à mente real: é uma espécie de pele que devemos remover se quisermos encontrar a mente real. Mas então, no que resta, encontramos uma outra pele a ser removida, e assim por diante. Procedendo desta maneira, chegamos à mente "real", ou finalmente chegamos à pele que não tem nada nela? Neste último caso, toda a mente é mecânica. (No entanto, não seria uma máquina de estado discreto. Discutimos isso.)

Estes dois últimos parágrafos não pretendem ser argumentos convincentes. Eles devem ser descritos como "recitações que tendem a produzir crença".

O único apoio realmente satisfatório que pode ser dado para o ponto de vista expresso no início do § 6, será aquele fornecido esperando pelo fim do século e depois fazendo o experimento descrito. Mas o que podemos dizer enquanto isso? Que medidas devem ser tomadas agora para que o experimento seja bem sucedido?

Como já expliquei, o problema é principalmente de programação. Devem ser feitos avanços na engenharia também, mas parece improvável que estes não sejam adequados para os requisitos. As estimativas da capacidade de armazenamento do cérebro variam de  $10$  a  $10^{15}$  dígitos binários. Eu me inclino para os valores mais baixos e acredito que apenas uma fração muito pequena é usada para os tipos mais elevados de pensamento. A maior parte é provavelmente usada para a retenção de impressões visuais, eu ficaria surpreso se mais de  $10^9$  fosse necessário para um jogo satisfatório do jogo de imitação, de qualquer forma contra um homem cego. (Nota: A capacidade da Enciclopédia Britânica, 11ª edição, é de  $2 \times 10^9$ ) Uma capacidade de armazenamento de  $10^7$ , seria uma possibilidade muito prática, mesmo pelas técnicas atuais. Provavelmente não é necessário aumentar a velocidade de operação das máquinas. Partes de máquinas modernas que podem ser consideradas análogas às células nervosas funcionam cerca de mil vezes mais rápido que as últimas. Isso deve fornecer uma "margem de segurança" que pode cobrir as perdas de velocidade que surgem de muitas maneiras. Nosso problema então é descobrir como programar essas máquinas para jogar o jogo. Na minha taxa atual de trabalho, eu produzo cerca de mil dígitos de progratiirne por dia, de modo que cerca de sessenta trabalhadores, trabalhando firmemente ao longo dos cinquenta anos, poderiam realizar o trabalho, se nada entrasse na cesta de lixo. Algum método mais expedito parece desejável.

No processo de tentar imitar uma mente humana adulta, somos obrigados a pensar bastante sobre o processo que a levou ao estado em que se encontra. Podemos notar três componentes.

- (a) O estado inicial da mente, digamos no nascimento,
- b) A educação a que foi submetido,
- c) Outra experiência, que não deve ser descrita como educação, a que foi submetida.

Em vez de tentar produzir um programa para simular a mente adulta, por que não tentar produzir uma que simule a criança? Se isso fosse então submetido a um curso apropriado de educação, obter-se-ia o cérebro adulto. Presumivelmente, o cérebro da criança é algo parecido com um notebook, como se o comprasse do papel de parede. Muito pouco mecanismo e muitas folhas em branco. (Mecanismo e escrita são, do nosso ponto de vista, quase sinônimos.) Nossa esperança é que haja tão pouco mecanismo no cérebro infantil que algo parecido possa ser facilmente programado. A quantidade de trabalho na educação que podemos assumir, como uma primeira aproximação, é a mesma que para a criança humana.

Assim, dividimos nosso problema em duas partes. O programa infantil e o processo de educação. Esses dois permanecem muito intimamente conectados. Não podemos esperar encontrar uma boa máquina infantil na primeira tentativa. É preciso experimentar ensinar uma dessas máquinas e ver quão bem ela aprende. Pode-se então tentar outro e ver se é melhor ou pior. Existe uma ligação óbvia entre este processo e a evolução, pelas identificações

Estrutura da máquina infantil = material hereditário

Mudanças da máquina filho = mutação,

Seleção natural = julgamento do experimentador

Pode-se esperar, no entanto, que esse processo seja mais rápido do que a evolução. A sobrevivência do mais apto é um método lento para medir vantagens. O experimentador, pelo exercício da inteligência, deveria ser capaz de acelerá-lo. Igualmente importante é o fato de que ele não está restrito a mutações aleatórias. Se ele pode traçar uma causa de alguma fraqueza, ele provavelmente pode pensar no tipo de mutação que irá melhorá-lo.



Não será possível aplicar exatamente o mesmo processo de ensino à máquina do que a uma criança normal. Ele não será, por exemplo, provido de pernas, de modo que não poderia ser pedido para sair e encher o tanque de carvão. Possivelmente, pode não ter olhos. Mas por mais que essas deficiências possam ser superadas por uma engenharia inteligente, não se pode mandar a criatura para a escola sem que as outras crianças se divirtam excessivamente. Deve ser dado algumas propinas. Não precisamos nos preocupar demais com as pernas, olhos, etc. O exemplo da srta. Helen Keller mostra que a educação pode ocorrer desde que a comunicação em ambas as direções entre professor e aluno possa ocorrer de alguma forma ou de outra.

Normalmente associamos punições e recompensas ao processo de ensino. Algumas máquinas-filho simples podem ser construídas ou programadas nesse tipo de princípio. A máquina tem que ser construída de tal maneira que os eventos que precederam a ocorrência de um sinal de punição dificilmente serão repetidos, enquanto um sinal de recompensa aumentou a probabilidade de repetição dos eventos que levaram a ele. Essas definições não pressupõem nenhum sentimento por parte da máquina, fiz algumas experiências com uma dessas máquinas infantis e consegui ensinar algumas coisas, mas o método de ensino era pouco ortodoxo para que a experiência fosse realmente bem-sucedida.

O uso de punições e recompensas pode, na melhor das hipóteses, fazer parte do processo de ensino. Grosso modo, se o professor não tiver outro meio de comunicação com o aluno, a quantidade de informação que pode alcançá-lo não excede o número total de recompensas e punições aplicadas. Quando uma criança aprender a repetir "Casabianca", provavelmente se sentirá muito dolorida se o texto só puder ser descoberto por uma técnica de "Vinte Perguntas", cada "NÃO" tomando a forma de um golpe. É necessário, portanto, ter outros canais de comunicação "sem emoção". Se estes estiverem disponíveis, é possível ensinar uma máquina por meio de punições e recompensas a obedecer ordens

dadas em alguma língua, por exemplo, uma linguagem simbólica. Estas ordens devem ser transmitidas através do "

As opiniões podem variar quanto à complexidade adequada na máquina infantil. Pode-se tentar torná-lo o mais simples possível de forma consistente com os princípios gerais. Alternativamente, pode-se ter um sistema completo de inferência lógica "embutido". "No último caso, a loja estaria amplamente ocupada com definições e proposições. As proposições teriam vários tipos de status, por exemplo, fatos bem estabelecidos, conjeturas, teoremas matematicamente provados, declarações dadas por uma autoridade, expressões tendo a forma lógica de proposição, mas não valor de crença. Certas proposições podem ser descritas como "imperativas". A máquina deve ser construída de modo que, assim que um imperativo seja classificado como "bem estabelecido", a ação apropriada ocorra automaticamente. Para ilustrar isso, suponha que o professor diga à máquina: "Faça sua lição de casa agora". Isso pode fazer com que "O professor diga 'Faça sua lição de casa agora'" para ser incluído entre os fatos bem estabelecidos. Outro fato pode ser: "Tudo o que o professor diz é verdade". Combiná-los pode eventualmente levar ao imperativo "Faça sua lição de casa agora", sendo incluído entre os fatos bem estabelecidos, e isso, pela construção da máquina, significará que a lição de casa realmente começa, mas o efeito é muito satisfatório. . Os processos de inferência usados pela máquina não precisam ser tais que satisfariam os lógicos mais exigentes. Pode não haver, por exemplo, hierarquia de tipos. Mas isso não significa que as falácias do tipo irão ocorrer mais do que somos obrigados a cair sobre penhascos sem proteção. Imperativos adequados (expressos nos sistemas, que não fazem parte das regras do sistema) como "Não use uma classe a menos que seja uma subclasse de uma que tenha sido mencionada pelo professor" podem ter um efeito similar a "Não vá muito perto da borda. "

Os imperativos que podem ser obedecidos por uma máquina que não tem membros são obrigados a ter um caráter bastante intelectual, como no exemplo (fazendo lição de

casa) dado acima. importante entre tais imperativos serão aqueles que regulam a ordem na qual as regras do sistema lógico em questão devem ser aplicadas, pois em cada estágio quando se está usando um sistema lógico, há um grande número de passos alternativos, qualquer um dos quais é permitido aplicar um, no que diz respeito à obediência às regras do sistema lógico. Essas escolhas fazem a diferença entre um raciocinador brilhante e um patético, não a diferença entre um sonoro e um falacioso. Proposições que levam a imperativos desse tipo podem ser "Quando Sócrates é mencionado, use o silogismo em Bárbara" ou "

A ideia de uma máquina de aprendizagem pode parecer paradoxal para alguns leitores. Como as regras de operação da máquina podem mudar? Eles devem descrever completamente como a máquina reagirá, qualquer que seja sua história, quaisquer que sejam as mudanças que ela possa sofrer. As regras são, portanto, bastante invariantes no tempo. Isso é bem verdade. A explicação do paradoxo é que as regras que se modificam no processo de aprendizagem são de um tipo menos pretensioso, reivindicando apenas uma validade efêmera. O leitor pode traçar um paralelo com a Constituição dos Estados Unidos.

Uma característica importante de uma máquina de aprendizado é que seu professor muitas vezes ignorará completamente o que está acontecendo lá dentro, embora ainda consiga de alguma forma prever o comportamento de seu aluno. Isso deve se aplicar mais fortemente à educação posterior de uma máquina que surge de uma máquina infantil de design (ou programa) bem experimentado. Isto está em claro contraste com o procedimento normal quando se usa uma máquina para realizar cálculos. O objeto da pessoa é então ter uma imagem mental clara do estado da máquina em cada momento da computação. Este objeto só pode ser alcançado com uma luta. A visão de que "a máquina só pode fazer o

que sabemos como ordenar", "parece estranha diante disso. Processos que são aprendidos não produzem cem por cento de certeza de resultado; se o fizessem, não poderiam ser desaprendidos.

Provavelmente, é sensato incluir um elemento aleatório em uma máquina de aprendizado. Um elemento aleatório é bastante útil quando estamos procurando por uma solução de algum problema. Suponhamos, por exemplo, que quiséssemos encontrar um número entre 50 e 200 que fosse igual ao quadrado da soma de seus dígitos, poderíamos começar em 51, depois tentar 52 e continuar até obtermos um número que funcionasse. Alternativamente, podemos escolher números aleatoriamente até obtermos um bom. Esse método tem a vantagem de ser desnecessário manter o controle dos valores que foram tentados, mas a desvantagem de que se pode tentar o mesmo duas vezes, mas isso não é muito importante se houver várias soluções. O método sistemático tem a desvantagem de que pode haver um enorme bloco sem soluções na região que deve ser investigado primeiro. Agora, o processo de aprendizagem pode ser considerado como uma busca por uma forma de comportamento que satisfará o professor (ou algum outro critério). Como provavelmente há um número muito grande de soluções satisfatórias, o método aleatório parece ser melhor que o sistemático. Deve ser notado que é usado no processo análogo da evolução. Mas aí o método sistemático não é possível. Como se poderia acompanhar as diferentes combinações genéticas que foram tentadas, de modo a evitar tentá-las novamente? Deve ser notado que é usado no processo análogo da evolução. Mas aí o método sistemático não é possível. Como se poderia acompanhar as diferentes combinações genéticas que foram tentadas, de modo a evitar tentá-las novamente? Deve ser notado que é usado no processo análogo da evolução. Mas aí o método sistemático não é possível. Como se poderia acompanhar as diferentes combinações genéticas que foram tentadas, de modo a evitar tentá-las novamente?

Podemos esperar que as máquinas acabem por competir com os homens em todos os campos puramente intelectuais. Mas quais são os melhores para começar? Mesmo esta é uma decisão difícil. Muitas pessoas pensam que uma atividade muito abstrata, como o jogo de xadrez, seria melhor. Também pode ser mantido que é melhor fornecer à máquina os melhores órgãos sensoriais que o dinheiro pode comprar, e depois ensiná-la a entender e falar inglês. Este processo pode seguir o ensino normal de uma criança. As coisas seriam apontadas e nomeadas, etc. Novamente, eu não sei qual é a resposta certa, mas acho que ambas as abordagens devem ser tentadas.

Nós só podemos ver uma curta distância à frente, mas podemos ver muito o que precisa ser feito.