1. O Jogo da Imitação

Proponho considerar a questão: 'As máquinas podem pensar?' Isso deve começar com as definições do significado dos termos 'máquina' e 'pensar'. As definições podem ser formuladas 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 como elas são comumente usadas, é difícil escapar da 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 questão por outra, que está intimamente relacionada a ela e é expressa em palavras relativamente inequívocas.

A nova forma do problema pode ser descrita em termos de um jogo que chamamos de 'jogo da imitação'. É jogado com três pessoas, um homem (A), uma mulher (B) e um interrogador (C) que podem ser de ambos os sexos. O interrogador fica em uma sala separada 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 '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, por favor, me diga o comprimento de seu cabelo? Agora suponha que X seja realmente A, então A deve responder. O objetivo de A no jogo é tentar fazer com que C faça a identificação errada. Sua resposta pode, portanto, ser

'Meu cabelo é shingled, e os fios mais longos têm cerca de nove centímetros de comprimento.'

Para que os tons de voz não ajudem o interrogador, as respostas devem ser escritas, ou melhor ainda, datilografadas. O arranjo ideal é ter um teleimpressor 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 respostas dela, mas de nada servirá, pois o homem pode fazer comentários semelhantes.

Agora fazemos a pergunta: 'O que acontecerá quando uma máquina assumir o papel de A neste jogo?' O interrogador decidirá erroneamente com tanta frequência quando o jogo é jogado assim como quando o jogo é 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, interrompendo 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 indistinguível da pele humana. É possível que em algum momento isso possa ser feito, mas mesmo supondo que essa invenção esteja disponível, deveríamos sentir que havia pouco sentido em tentar tornar uma 'máquina pensante' mais humana vestindo-a com tal carne artificial. A forma em que colocamos o problema reflete esse fato na condição que impede o interrogador de ver ou tocar

os outros competidores, ou ouvir suas vozes. Algumas outras vantagens do critério proposto podem ser mostradas por meio de perguntas e respostas exemplares. Por isso:

- P:

  Por favor, escreva-me um soneto sobre a Ponte Forth.

- UMA :

  Conte comigo nessa. Nunca consegui escrever poesia.

- P:

  Adicionar 34957 a 70764

- UMA :

  (Pausa cerca de 30 segundos e depois dá como resposta) 105621.

- P:

  Você joga xadrez?

- UMA :

  sim.

- P:

  Eu tenho K no meu K1, e nenhuma outra peça. Você tem apenas K em K6 e R em R1. É o seu movimento. O que você toca?

- UMA :

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

O método de perguntas e respostas parece ser adequado para introduzir quase qualquer um dos campos da atividade humana que desejamos incluir. 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 vangloriar-se, se o considerarem conveniente, o quanto quiserem de seus encantos, força ou heroísmo, mas o interrogador não pode exigir demonstrações práticas.

O jogo pode ser criticado com base no fato de que as probabilidades pesam demais contra a máquina. Se o homem tentasse fingir ser a máquina, ele claramente faria uma exibição muito ruim. Ele seria entregue imediatamente pela lentidão e imprecisão na aritmética. As máquinas não podem 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 satisfatoriamente o jogo da imitação, não precisamos nos incomodar com essa objeção.

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

3. As Máquinas envolvidas no Jogo

A questão que colocamos no § 1 não será bem definida até que tenhamos especificado o que queremos dizer com a palavra 'máquina'. É natural que desejemos permitir que todo tipo de técnica de engenharia seja usada em nossas máquinas. Desejamos também permitir a possibilidade de um engenheiro ou equipe de engenheiros construir uma máquina que funcione, mas cujo modo de operação não possa ser satisfatoriamente descrito por seus construtores porque eles aplicaram um método amplamente experimental. Finalmente, desejamos excluir das máquinas os homens nascidos da maneira usual. É difícil enquadrar as definições de modo a satisfazer essas três condições. Pode-se, por exemplo, insistir que a equipe de engenheiros deve 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 da técnica biológica merecedora dos mais altos elogios, mas não estaríamos inclinados a considerá-lo um caso de "construção de uma máquina pensante". Isso nos leva a abandonar a exigência de que todo tipo de técnica seja permitida. Estamos mais dispostos a fazê-lo pelo fato de que o atual interesse em 'máquinas pensantes' foi despertado por um tipo particular de máquina, geralmente chamado de 'computador eletrônico' ou 'computador digital'. Seguindo esta sugestão, só permitimos que computadores digitais participem do nosso jogo.mas não estaríamos inclinados a considerá-lo um caso de "construção de uma máquina pensante". Isso nos leva a abandonar a exigência de que todo tipo de técnica seja permitida. Estamos mais dispostos a fazê-lo pelo fato de que o atual interesse em 'máquinas pensantes' foi despertado por um tipo particular de máquina, geralmente chamado de 'computador eletrônico' ou 'computador digital'. Seguindo esta sugestão, só permitimos que computadores digitais participem do nosso jogo.mas não estaríamos inclinados a considerá-lo um caso de "construção de uma máquina pensante". Isso nos leva a abandonar a exigência de que todo tipo de técnica seja permitida. Estamos mais dispostos a fazê-lo pelo fato de que o atual interesse em 'máquinas pensantes' foi despertado por um tipo particular de máquina, geralmente chamado de 'computador eletrônico' ou 'computador digital'. Seguindo esta sugestão, só permitimos que computadores digitais participem do nosso jogo.Seguindo esta sugestão, só permitimos que computadores digitais participem do nosso jogo.Seguindo esta sugestão, só permitimos que 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 necessário um breve relato da natureza e 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), acontecer que os computadores digitais não consigam dar uma boa exibição no jogo .

Já existem vários computadores digitais em funcionamento, e pode-se perguntar: 'Por que não tentar o experimento imediatamente? Seria fácil satisfazer as condições do jogo. Vários interrogadores poderiam ser usados e estatísticas compiladas para mostrar com que frequência a identificação correta era 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 existem computadores imagináveis que se sairiam bem. Mas esta é apenas a resposta curta. Veremos essa questão sob uma luz diferente mais adiante.

4. Computadores Digitais

A ideia por trás dos computadores digitais pode ser explicada dizendo que essas máquinas são destinadas a realizar quaisquer operações que poderiam ser feitas por um computador humano. O computador humano deve seguir regras fixas; ele não tem autoridade para se desviar deles em nenhum detalhe. Podemos supor que essas regras são 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 faz seus cálculos. Ele também pode fazer suas multiplicações e adições em uma 'máquina de mesa', mas isso não é importante.

Se usarmos a explicação acima como definição, estaremos em perigo de circularidade de argumento. Evitamos isso dando um esboço dos meios pelos quais o efeito desejado é alcançado. Um computador digital geralmente pode ser considerado como consistindo de três partes:

- Loja.
- Unidade executiva.
- Ao controle.

O depósito é um depósito de informações e corresponde ao papel do computador humano, seja este o papel em que ele faz seus cálculos ou aquele em que seu livro de regras é impresso. Na medida em que o computador humano fizer cálculos em sua cabeça, uma parte da loja corresponderá à sua memória.

A unidade executiva é a parte que realiza as várias operações individuais envolvidas em um cálculo. O que são essas operações individuais varia de máquina para máquina. Normalmente, operações bastante longas podem ser feitas, como 'Multiply 3540675445 by 7076345687', mas em algumas máquinas apenas as muito simples, como 'Write down 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 zelar para que essas instruções sejam obedecidas corretamente e na ordem correta. O controle é construído de tal forma que isso necessariamente acontece.

As informações na loja geralmente são divididas 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 pode dizer—

'Adicione o número armazenado na posição 6809 ao número 4302 e coloque o resultado de volta na última posição de armazenamento'.

Escusado será dizer que não ocorreria na máquina expressa em inglês. É mais provável que seja 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, *viz* . 'Adicione o número...' Será notado que a instrução ocupa 10 dígitos e assim forma um pacote de informações, muito convenientemente. O controle normalmente levará as instruções a serem obedecidas na ordem das posições em que estão armazenadas, mas ocasionalmente uma instrução como

'Agora obedeça a instrução armazenada na posição 5606, e continue de lá' pode ser encontrado, ou novamente

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

Instruções destes últimos tipos são muito importantes porque tornam possível que uma sequência de operações seja repetida várias vezes até que alguma condição seja cumprida, mas ao fazê-lo obedecer não a novas instruções em cada repetição, mas as mesmas ao longo do tempo. e de novo. Para fazer uma analogia doméstica. Suponha que a mãe queira que Tommy ligue para o sapateiro todas as manhãs a caminho da escola para ver se os sapatos dela estão prontos, ela pode perguntar de novo todas as manhãs. Alternativamente, ela pode afixar um aviso de uma vez por todas no corredor que ele verá quando sair para a escola e que o avise para pedir os sapatos, e também para destruir o aviso quando ele voltar se estiver com os sapatos. .

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

O livro de regras que descrevemos como usando nosso computador humano é, obviamente, uma ficção conveniente. Computadores humanos reais realmente se lembram do que precisam fazer. Se alguém quiser fazer uma máquina imitar o comportamento do computador humano em alguma operação complexa, é preciso perguntar a ele como isso 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 realizar a operação A' significa colocar a tabela de instruções apropriada na máquina para que ela faça A.

Uma variante interessante da ideia de um computador digital é um 'computador digital com um elemento aleatório'. Estes possuem instruções envolvendo o lançamento de um dado ou algum processo eletrônico equivalente; uma dessas instruções pode ser, por exemplo, '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 use essa frase). Normalmente, não é possível determinar pela observação de uma máquina se ela possui um elemento aleatório, pois um efeito semelhante pode ser produzido por dispositivos como fazer as escolhas depender dos dígitos do decimal para π.

A maioria dos computadores digitais reais tem apenas um armazenamento finito. Não há dificuldade teórica na ideia de um computador com armazenamento ilimitado. É claro que apenas uma parte finita pode ter sido usada de cada vez. Da mesma forma, apenas uma quantidade finita pode ter sido construída, mas podemos imaginar cada vez mais sendo adicionado conforme necessário. Tais computadores têm interesse teórico especial e serão chamados de computadores de capacidade infinita.

A ideia de um computador digital é antiga. Charles Babbage, professor lucasiano de matemática em Cambridge de 1828 a 1839, planejou tal máquina, chamada de Máquina Analítica, mas nunca foi concluída. Embora Babbage tivesse todas as ideias 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 100 vezes mais lenta que a máquina de Manchester, ela mesma uma das mais lentas das máquinas modernas. O armazenamento deveria ser puramente mecânico, usando rodas e cartões.

O fato de a Máquina Analítica de Babbage ser inteiramente mecânica nos ajudará a nos livrar de uma superstição. A importância está muitas vezes ligada 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, em certo sentido, equivalentes, vemos que esse uso da eletricidade não pode ter importância teórica. É claro que a eletricidade geralmente entra no que diz respeito à sinalização rápida, de modo que não é surpreendente que a encontremos em ambas as 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 do uso de eletricidade é vista como apenas uma semelhança muito superficial.

5. Universalidade dos 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 há muitos tipos de máquinas que podem ser lucrativamente *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 exemplo de uma máquina de estado discreto, podemos considerar uma roda que gira 120° uma vez por segundo, mas pode ser parada por uma alavanca que pode ser operada de fora; além disso uma lâmpada deve acender em uma das posições da roda. Esta máquina pode ser descrita abstratamente como segue. O estado interno da máquina (que é descrito pela posição da roda) pode ser q_1, q_2 ou q_3. Há um sinal de entrada i_0 ou i_1, (posição da alavanca). O estado interno a qualquer momento é determinado pelo último estado e sinal de entrada de acordo com a tabela

		Last State		
		q_1	q_2	q_3
Input	i_0	q_2	q_3	q_1
	i_1	q_1	q_2	q_3

Abrir em nova guia Baixar slide

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

State	q_1	q_2	q_3
Output	o_0	o_0	o_1

Abrir em nova guia Baixar slide

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

Parece que dado o estado inicial da máquina e os sinais de entrada é sempre possível prever todos os estados futuros. Isso é uma reminiscência da visão de Laplace de que a partir do estado completo do universo em um momento do tempo, conforme descrito pelas posições e velocidades de todas as partículas, deveria ser possível prever todos os estados futuros. A previsão que estamos considerando é, no entanto, mais próxima da praticabilidade do que a considerada por Laplace. O sistema do 'universo como um todo' é tal que erros muito 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 ocorra. Mesmo quando consideramos as máquinas físicas reais em vez das máquinas idealizadas, o conhecimento razoavelmente preciso do estado em um momento produz um conhecimento razoavelmente preciso qualquer número de passos depois.

Como mencionamos, os computadores digitais se enquadram na classe de máquinas de estado discreto. Mas o número de estados de que tal máquina é capaz é geralmente enorme. Por exemplo, o número da máquina que agora funciona 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 deve ser tão imenso. O computador inclui um depósito correspondente ao papel usado por um computador humano. Deve ser possível escrever na loja qualquer uma das combinações de símbolos que possam ter sido escritas no papel. Para simplificar, suponha que apenas dígitos de 0 a 9 sejam usados como símbolos. As variações na caligrafia são ignoradas. Suponha que o computador tenha 100 folhas de papel cada uma contendo 50 linhas com espaço para 30 dígitos. Então o número de estados é $10^{100 \times 50 \times 30}$, *ou seja* , $10^{150.000}$. Trata-se do número de estados de três máquinas Manchester juntas. O logaritmo para a 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 são colocadas juntas, suas capacidades devem ser somadas para obter a capacidade da máquina resultante. Isso leva à possibilidade de declarações como 'A máquina de Manchester contém 64 faixas magnéticas cada uma com capacidade de 2560, oito tubos eletrônicos com capacidade de 1280. O armazenamento variado 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 rapidez suficiente, o computador digital poderia imitar o comportamento de qualquer máquina de estado discreto. O jogo da imitação poderia então ser jogado com a máquina em questão (como B) e o computador digital de imitação (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 suficientemente rápido. Além disso, ele 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 eles são máquinas *universais* . A existência de máquinas com esta propriedade tem a importante consequência de que, à parte as considerações de velocidade, é desnecessário projetar várias novas máquinas para realizar vários processos de computação. Todos eles podem ser feitos com um computador digital, devidamente programado para cada caso. Ver-se-á que, como consequência disso, todos os computadores digitais são, em certo sentido, equivalentes.

Podemos agora considerar novamente o ponto levantado no final do §3. Sugeriu-se provisoriamente que a pergunta 'As máquinas podem pensar?' deveria ser substituído por 'Existem computadores digitais imagináveis que se sairiam bem no jogo da imitação?' Se desejarmos, podemos tornar isso superficialmente mais geral e perguntar 'Existem máquinas de estado discretas que funcionariam

bem?' Mas, em vista da propriedade de universalidade, vemos que qualquer uma dessas questões é equivalente a isso: 'Vamos fixar nossa atenção em um computador digital específico *C*. É verdade que modificando este computador para ter um armazenamento adequado, aumentando adequadamente sua velocidade de ação, e dotando-o de um programa apropriado, *C* pode desempenhar satisfatoriamente o papel de A no jogo da imitação, sendo o papel de B desempenhado por um homem?'

6. Pontos de vista contrários sobre a questão principal

Podemos agora considerar que o terreno foi limpo e estamos prontos para prosseguir com o debate sobre nossa questão, 'As máquinas podem pensar?' e a variante dela citada no final da última seção. Não podemos abandonar totalmente a forma original do problema, pois as opiniões diferirão quanto à adequação da substituição e devemos pelo menos ouvir o que tem a ser dito a esse respeito.

Isso simplificará as coisas para o leitor se eu explicar primeiro minhas próprias crenças sobre o assunto. Considere primeiro a forma mais precisa da pergunta. Acredito que daqui a cerca de cinquenta anos será possível programar computadores, com capacidade de armazenamento de cerca de 10^{9}, para fazê-los jogar o jogo da imitação tão bem que um interrogador médio não terá mais de 70% de chance de fazer a identificação correta após cinco minutos de interrogatório. A pergunta original, 'As máquinas podem pensar!' Eu acredito ser muito sem sentido para merecer discussão. No entanto, acredito que no final do século o uso das palavras e a opinião geral educada terão mudado tanto que se poderá falar de máquinas pensando sem esperar ser contrariado. Além disso, acredito que nenhum propósito útil é servido ao ocultar essas crenças. A visão popular de que os cientistas procedem inexoravelmente de fato bem estabelecido para fato bem estabelecido, nunca sendo influenciados por qualquer conjectura não provada, é bastante equivocada. Desde que fique claro quais são fatos comprovados e quais são conjecturas, nenhum dano pode resultar.

Passo agora a considerar as opiniões opostas às minhas.

(1) A Objeção Teológica

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

Não posso aceitar qualquer parte disso, mas tentarei responder em termos teológicos. Eu acharia o argumento mais convincente se os animais fossem classificados com os homens, pois há uma diferença maior, a meu ver, entre o animal típico e o inanimado do que entre o homem e os outros animais. O caráter arbitrário da visão ortodoxa fica 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 tornar um igual a dois,mas não devemos acreditar que Ele tem liberdade para conferir uma alma a um elefante se Ele achar conveniente? Poderíamos esperar que Ele só exercesse esse poder em conjunto com uma mutação que fornecesse ao elefante um cérebro adequadamente aprimorado para atender às necessidades dessa alma. Um argumento de forma exatamente semelhante pode ser feito para o caso das máquinas. Pode parecer diferente porque é mais difícil de “engolir”. Mas isso realmente

significa apenas que pensamos que seria menos provável que Ele considerasse as circunstâncias adequadas para conferir uma alma. As circunstâncias em questão são discutidas no restante deste artigo. Ao tentar construir tais máquinas, não devemos usurpar irreverentemente Seu poder de criar almas, mais do que estamos na procriação de filhos: ao contrário, estamos, em ambos os casos,instrumentos de Sua vontade provendo mansões para as almas que Ele cria.

No entanto, isso é mera especulação. Não estou muito impressionado com argumentos teológicos, seja lá o que eles podem ser usados para apoiar. Tais argumentos foram muitas vezes considerados insatisfatórios no passado. No tempo de Galileu, foi argumentado que os textos, "E o sol parou ... e não se apressou em se pôr cerca de um dia inteiro" (Josué x. 13) e "Ele lançou os fundamentos da terra, para que não mover a qualquer momento" (Salmo cv. 5) foram uma refutação adequada da teoria copernicana. Com nosso conhecimento atual, tal argumento parece fútil. Quando esse conhecimento não estava disponível, causava uma impressão bem diferente.

(2) A objeção 'cabeças na areia'

"As consequências do pensamento das máquinas seriam terríveis demais. Esperemos e acreditemos que eles não podem fazê-lo".

Este argumento raramente é expresso tão abertamente como no formulário acima. Mas afeta a maioria de nós que pensa sobre isso. Gostamos de acreditar que o Homem é, de alguma forma sutil, superior ao resto da criação. É melhor que se demonstre que ele é *necessariamente* superior, pois então não há perigo de ele perder sua posição de comando. A popularidade do argumento teológico está claramente ligada a esse sentimento. É provável que seja bastante forte em pessoas intelectuais, uma vez que valorizam o poder de pensar mais do que os outros e estão mais inclinados a basear sua crença na superioridade do homem nesse poder.

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

(3) A Objeção Matemática

Há uma série de resultados de lógica matemática que podem ser usados para mostrar que existem limitações para os poderes das máquinas de estado discreto. O mais conhecido desses resultados é conhecido como teorema de Gödel, [1] e mostra que em qualquer sistema lógico suficientemente poderoso podem ser formuladas declarações que não podem ser provadas nem refutadas dentro do sistema, a menos que possivelmente o próprio sistema seja inconsistente. Existem outros resultados, em alguns aspectos semelhantes, devidos a *Church, Kleene, Rosser* e *Turing.*Este último resultado é o mais conveniente de se considerar, uma vez que se refere diretamente a máquinas, enquanto os outros só podem ser usados em um argumento comparativamente indireto: por exemplo, se o teorema de Gödel for usado, precisamos, além disso, ter alguns meios de descrever sistemas 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 capacidade infinita. Ele afirma que há certas coisas que tal máquina não pode fazer. Se estiver preparado para dar respostas a perguntas como no jogo da imitação, haverá algumas perguntas para as quais ele dará uma resposta errada ou deixará de dar uma resposta por mais tempo que seja concedido para uma resposta. Pode haver, é claro, muitas dessas perguntas,e perguntas que não podem ser respondidas por uma máquina podem ser satisfatoriamente respondidas por outra. É claro que estamos supondo por enquanto que as perguntas são do tipo para as quais uma resposta "Sim" ou "Não" é apropriada,

em vez de perguntas como "O que você acha de Picasso?" As perguntas nas quais sabemos que as máquinas devem falhar são desse tipo: “Considere a máquina especificada da seguinte maneira… . Esta máquina responderá 'Sim' a qualquer pergunta?” Os pontos devem ser substituídos por uma descrição de alguma máquina em uma forma padrão, que pode ser algo como o usado no § 5. Quando a máquina descrita tem uma certa relação comparativamente simples com a máquina que está sob interrogação, pode ser mostrada que a resposta está errada ou não está disponível. Este é o resultado matemático:argumenta-se que prova uma deficiência das máquinas à qual o intelecto humano não está sujeito.

A resposta curta a esse argumento é que, embora esteja estabelecido que existem limitações aos poderes de qualquer máquina em particular, apenas foi declarado, sem qualquer tipo de prova, que tais limitações não se aplicam ao intelecto humano. Mas não acho que essa visão possa ser descartada tão levianamente. Sempre que uma dessas máquinas recebe 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. Esse sentimento é ilusório? É, sem dúvida, bastante genuíno, mas não acho que se deva dar muita importância a ele. Muitas vezes, nós mesmos damos respostas erradas a perguntas para justificar nossa satisfação com tal evidência de falibilidade por parte das máquinas. Avançar,*todas as* máquinas. Em suma, então, pode haver homens mais inteligentes do que qualquer máquina, mas também pode haver outras máquinas mais inteligentes, e assim por diante.

Aqueles que se apegam ao argumento matemático estariam, penso eu, principalmente dispostos a 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 está muito bem expresso na Oração Lister *do Professor Jefferson* para 1949, da qual cito. “Até que uma máquina possa escrever um soneto ou compor um concerto por causa de pensamentos e emoções sentidas, e não pela casual queda de símbolos, poderíamos concordar que máquina é igual a cérebro – isto é, não apenas escrevê-lo, mas saber que ele escreveu isto. Nenhum mecanismo poderia sentir (e não apenas sinalizar artificialmente, um artifício fácil) prazer em seus sucessos, tristeza quando suas válvulas se fundem, ser aquecido por lisonjas, ser infeliz por seus erros, ser encantado por sexo, ficar com raiva 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 se pode ter certeza de que uma máquina pensa é *ser* a máquina e sentir-se pensando. Alguém poderia então descrever esses sentimentos para o mundo, mas é claro que ninguém estaria justificado em prestar atenção. Da mesma forma, de acordo com essa visão, a única maneira de saber o que um *homem* pensa é ser esse homem em particular. Na verdade, é o ponto de vista solipsista. Pode ser a visão mais lógica a se sustentar, mas dificulta a comunicação de ideias. A está sujeito a acreditar que '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 de 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 da imitação como um 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 à moda do papagaio". Ouçamos uma parte de tal *viva voce* :

Interrogador: Na primeira linha de seu soneto que diz 'Devo te comparar a um dia de verão', 'um dia de primavera' não faria tão bem ou melhor?

Testemunha: Não iria escanear.

Interrogador: Que tal 'um dia de inverno' Isso iria escanear tudo bem.

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

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

Testemunha: De certa forma.

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

Testemunha: Eu não acho que você está falando sério. Por esfola de inverno entende-se um dia típico de inverno, em vez de um dia especial como o Natal.

E assim por diante. O que diria o professor Jefferson se a máquina de escrever sonetos fosse capaz de responder assim em *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 acho que ele a descreveria como "um artifício fácil". Esta frase é, penso eu, destinada a abranger dispositivos como a inclusão na máquina de um registro de alguém lendo um soneto, com comutação apropriada para ligá-lo de tempos em tempos.

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

Não desejo dar a impressão de que acho que não há mistério sobre a consciência. Há, por exemplo, um paradoxo relacionado a qualquer tentativa de localizá-lo. Mas não acho que esses mistérios precisem necessariamente ser resolvidos antes que possamos responder à pergunta com a qual estamos preocupados neste artigo.

(5) Argumentos de Várias Deficiências

Esses argumentos assumem a forma: “Eu garanto que você pode fazer com que as máquinas façam todas as coisas que você mencionou, mas você nunca será capaz de fazer uma para fazer X”. Numerosas características X são sugeridas nesta conexão. Eu ofereço uma seleção:

Seja gentil, engenhoso, bonito, amigável (p. 448), tenha iniciativa, tenha senso de humor, diferencie o certo do errado, cometa erros (p. 448), apaixone-se, aprecie morangos e creme (p. 448), fazer alguém se apaixonar por ela, aprender com a experiência (p. 456 f.), usar as palavras corretamente, ser sujeito de seu próprio pensamento (p. 449), ter tanta diversidade de comportamento quanto um homem, fazer algo realmente novo (pág. 450). (Algumas dessas deficiências recebem consideração especial, conforme indicado pelos números das páginas.)

Normalmente, nenhum suporte é oferecido para essas declarações. Acredito que eles se baseiam principalmente 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 essas são propriedades necessárias de máquinas em geral. Muitas dessas limitações estão associadas à capacidade de armazenamento muito pequena da maioria das máquinas. (Estou assumindo que a ideia de capacidade de armazenamento é estendida de alguma forma para cobrir outras máquinas além de máquinas de estado discreto. A definição exata não importa, pois nenhuma precisão matemática é reivindicada na presente discussão.) Há alguns anos, quando muito pouco se ouvia sobre computadores digitais, era possível suscitar muita incredulidade em relação a 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, é claro, em grande parte inconscientes. Quando uma criança queimada teme o fogo e mostra que o teme evitando-o, devo dizer que estava aplicando a indução científica. (É claro que eu também poderia descrever seu comportamento de muitas outras maneiras.) As obras e os costumes da humanidade não parecem ser um material muito adequado para aplicar a indução científica. Uma parte muito grande do espaço-tempo deve ser investigada, para obter resultados confiáveis.Caso contrário, podemos (como a maioria das crianças inglesas) decidir que todo mundo fala inglês e que é tolice aprender francês.

Há, no entanto, observações especiais a serem feitas sobre muitas das deficiências que foram mencionadas. A incapacidade de apreciar morangos e creme pode ter parecido frívola para o leitor. Possivelmente uma máquina pode ser feita para saborear este delicioso prato, mas qualquer tentativa de fazê-lo 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 como entre homem branco e homem branco, ou entre homem negro e homem negro.

A afirmação de que “máquinas não podem cometer erros” parece curiosa. Alguém é tentado a replicar: “Eles são piores por isso?” Mas vamos adotar uma atitude mais solidária e tentar ver o que realmente significa. Acho que essa crítica pode ser explicada em termos do jogo da imitação. Alega-se que o interrogador poderia distinguir a máquina do homem simplesmente colocando-lhes uma série de problemas de aritmética. A máquina seria desmascarada por causa de sua precisão mortal. A resposta para isso é simples. A máquina (programada para jogar o jogo) não tentaria dar o *direito*respostas para os problemas aritméticos. Introduziria deliberadamente erros de uma maneira calculada para confundir o interrogador. Uma falha mecânica provavelmente se manifestaria por meio de uma decisão inadequada sobre que tipo de erro cometer na aritmética. Mesmo esta interpretação da crítica não é suficientemente simpática. Mas não podemos permitir o espaço para ir muito mais longe. 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 de forma diferente do que foi projetado para fazer. Nas discussões filosóficas gosta-se de ignorar a possibilidade de tais erros; trata-se, portanto, de 'máquinas abstratas'.Essas máquinas abstratas são ficções matemáticas e não objetos físicos. Por definição, são incapazes de erros de funcionamento. Nesse sentido, podemos dizer verdadeiramente que 'máquinas nunca podem errar'. Erros de conclusão só podem surgir quando algum significado é atribuído 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. Claramente, não há razão alguma para dizer que uma máquina não pode cometer esse tipo de erro. Pode não fazer nada além de digitar repetidamente '0 = 1'. 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 errôneos.

A afirmação de que uma máquina não pode ser o sujeito de seu próprio pensamento só pode ser respondida se puder ser demonstrado que a máquina tem *algum* pensamento com *algum* assunto. No entanto, "o assunto das operações de uma máquina" parece significar alguma coisa, pelo menos para as pessoas que lidam com isso. Se, por exemplo, a máquina estava tentando encontrar uma solução da equação $x^2 - 40x - 11 = 0$ seria tentado a descrever essa equação como parte do objeto da máquina naquele momento. Nesse sentido, uma máquina, sem dúvida, pode ser seu próprio assunto. Pode ser usado para ajudar na elaboração de seus próprios programas ou para prever o efeito de alterações em sua própria estrutura. Observando os resultados de seu próprio comportamento, ele pode modificar seus próprios programas de modo a atingir algum propósito de maneira mais eficaz. Estas são possibilidades de um futuro próximo, ao invés de sonhos utópicos.

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

As críticas que estamos considerando aqui são muitas vezes 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 causará muita impressão. Pensa-se que o método (seja qual for, pois deve ser mecânico) é realmente bastante básico. Compare o parêntese na declaração de Jefferson citada na p. 21.

(6) Objeção de Lady Lovelace

Nossas informações mais detalhadas da Máquina Analítica de Babbage vêm de um livro de memórias de *Lady Lovelace.* Nele ela afirma: “A Máquina Analítica não tem pretensões de *originar* nada. Ele pode fazer *o que nós sabemos como ordenar que ele* execute” (itálicos dela). Esta afirmação é citada por *Hartree*(p. 70) que acrescenta: “Isso não significa que não seja possível construir equipamentos eletrônicos que 'pensarão por si', ou nos quais, em termos biológicos, se possa estabelecer um reflexo condicionado, que serviria como 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”.

Estou totalmente de acordo com Hartree sobre isso. Note-se que não afirma que as máquinas em questão não tinham a propriedade, mas sim que as provas disponíveis para Lady Lovelace não a encorajavam a acreditar que a possuíam. É bem possível que as máquinas em questão tivessem, em certo sentido, essa propriedade. Pois suponha que alguma máquina de estado discreto tenha a propriedade. O Analytical Engine era um computador digital universal, de modo que, se sua capacidade de armazenamento e velocidade fossem adequadas, ele poderia, por programação adequada, imitar a máquina em questão. Provavelmente esse argumento não ocorreu à condessa ou a Babbage. Em qualquer caso, não havia obrigação deles de reivindicar tudo o que poderia ser reivindicado.

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

Uma variante da objeção de Lady Lovelace afirma que uma máquina “nunca pode fazer nada realmente novo”. Isso pode ser aparado 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 fez não foi simplesmente o crescimento da semente plantada nele pelo ensino, ou o efeito de seguir princípios gerais bem conhecidos. Uma variante melhor da objeção diz que uma máquina nunca pode 'nos pegar de surpresa'. Esta afirmação é um desafio mais direto e pode ser enfrentado diretamente. As máquinas me surpreendem com grande frequência. Isso se deve em grande parte porque não faço cálculos suficientes para decidir o que esperar que eles façam, ou melhor, porque, embora faça um cálculo, o faço de maneira apressada, descuidada, correndo riscos. Talvez eu diga a mim mesmo: 'Suponho que a voltagem aqui deve ser a mesma que lá:de qualquer forma, vamos supor que seja.

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 confissões me deixam aberto a palestras sobre meus modos viciosos, mas não lançam nenhuma dúvida sobre minha credibilidade quando testemunho as surpresas que experimento.

Não espero que esta resposta silencie meu crítico. Ele provavelmente dirá que tais surpresas se devem a algum ato mental criativo de 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 requer tanto um “ato mental criativo” se o evento surpreendente se origina de um homem, um livro, uma máquina ou qualquer coisa senão.

A visão de que as máquinas não podem causar surpresas se deve, acredito, a uma falácia à qual filósofos e matemáticos estão particularmente sujeitos. Esta é a suposição de que assim que um fato é apresentado à mente, todas as consequências desse fato surgem na mente simultaneamente com ele. É uma suposição muito útil em muitas circunstâncias, mas facilmente se esquece que é falsa. Uma consequência natural de fazer isso é que se assume que não há virtude na mera elaboração de consequências a partir 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 atinge 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 da imitação, o interrogador não poderá tirar proveito dessa diferença. A situação pode ficar mais clara se considerarmos alguma outra máquina contínua mais simples. Um analisador diferencial funcionará 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.) Algumas delas fornecem suas respostas de forma digitada e, portanto, são adequadas para participar do jogo. Não seria possível para um computador digital prever exatamente quais respostas o analisador diferencial daria a um problema, mas seria bem capaz de dar o tipo certo de resposta. Por exemplo, se solicitado a fornecer o valor de π(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). Nessas 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 pretendam descrever o que um homem deve fazer em todas as circunstâncias concebíveis. Pode-se, por exemplo, ter uma regra de que deve-se parar quando se vê um semáforo vermelho e ir se vê um verde, mas e se por alguma falha ambos aparecerem juntos? Pode-se talvez decidir que é mais seguro parar. Mas algumas outras dificuldades podem surgir dessa decisão mais tarde. Tentar estabelecer regras de conduta para cobrir todas as eventualidades, mesmo aquelas decorrentes de semáforos, parece ser impossível. Com tudo isso eu concordo.

Se substituirmos "leis de comportamento que regulam sua vida" por "leis de conduta pelas quais ele regula sua vida" no argumento citado, o meio não distribuído não é mais insuperável. Pois acreditamos que não só é verdade que ser regulado por leis de comportamento implica ser algum tipo de máquina (embora não necessariamente uma máquina de estado discreto), mas que, inversamente, ser tal máquina implica ser regulado por tais leis. No entanto, não podemos nos convencer tão facilmente da ausência de leis de comportamento completas como de regras de conduta completas. A única maneira que conhecemos para encontrar tais leis é a observação científica, e certamente não conhecemos nenhuma circunstância sob a qual poderíamos dizer: 'Procuramos o suficiente. Não existem tais leis.' Pois acreditamos que não só é verdade que ser regulado por leis de comportamento implica ser algum tipo de máquina (embora não necessariamente uma máquina de estado discreto), mas que, inversamente, ser tal máquina implica ser regulado por tais leis. No entanto, não podemos nos convencer tão facilmente da ausência de leis de comportamento completas como de regras de conduta completas. A única maneira que conhecemos para encontrar tais leis é a observação científica, e certamente não conhecemos nenhuma circunstância sob a qual poderíamos dizer: 'Procuramos o suficiente. Não existem tais leis.'

Podemos demonstrar com mais força que tal afirmação seria injustificada. Pois suponha que poderíamos ter certeza de encontrar tais leis se existissem. Então, dada uma máquina de estado discreto, certamente seria possível descobrir por observação suficiente sobre ela para prever seu comportamento futuro, e isso dentro de um tempo razoável, digamos mil anos. Mas este não parece ser o caso. Configurei no computador Manchester um pequeno programa usando apenas 1.000 unidades de armazenamento, em que a máquina fornecida com um número de dezesseis dígitos responde com outro em dois segundos. Eu desafiaria qualquer um a aprender com essas respostas o suficiente sobre o programa para poder prever quaisquer respostas para 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, *viz.* telepatia, clarividência, precognição e psicocinese. Esses fenômenos perturbadores parecem negar todas as nossas ideias científicas usuais. Como gostaríamos de desacreditá-los! Infelizmente, a evidência estatística, pelo menos para a telepatia, é esmagadora. É muito difícil reorganizar as próprias idéias de modo a encaixar esses novos fatos. Uma vez aceitas, não parece um grande passo acreditar em fantasmas e bichos. A ideia de que nossos corpos se

movem simplesmente de acordo com as leis conhecidas da física, juntamente com algumas outras ainda não descobertas, mas um pouco semelhantes, seria uma das primeiras a desaparecer.

Este argumento é, a meu ver, bastante forte. Pode-se dizer em resposta que muitas teorias científicas parecem permanecer viáveis na prática, apesar de colidirem com a PES; que, de fato, podemos nos dar muito bem se nos esquecermos disso. Isso é um conforto bastante frio, e teme-se que o pensamento seja exatamente o tipo de fenômeno em que a PES pode ser especialmente relevante.

Um argumento mais específico baseado em PES poderia ser o seguinte: “Vamos jogar o jogo da imitação, usando como testemunhas um homem que é bom como receptor telepático e um computador digital. O interrogador pode fazer perguntas como "A que naipe pertence a carta na minha mão direita?" O homem por telepatia ou clarividência dá a resposta certa 130 vezes em 400 cartas. A máquina só pode adivinhar aleatoriamente, e talvez acerte 104 vezes, então o interrogador faz a identificação correta.” Há 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 qualquer questionamento, por clarividência. Com ESP tudo pode acontecer.

Se a telepatia for admitida, será necessário apertar nosso teste. A situação poderia ser considerada análoga à que ocorreria se o interrogador estivesse falando consigo mesmo e um dos competidores estivesse ouvindo 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 sustentar minhas opiniões. Se eu tivesse feito isso, não teria me esforçado tanto para apontar as falácias em pontos de vista contrários. As provas que tenho, darei agora.

Voltemos por um momento à objeção de Lady Lovelace, que afirmava que a máquina só pode fazer o que mandamos. Pode-se dizer que um homem pode "injetar" uma ideia na máquina, e que 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 inferior ao crítico: uma ideia injetada é corresponder a um nêutron entrando na pilha de fora. Cada um desses nêutrons causará uma certa perturbação que eventualmente desaparecerá. Se, no entanto, o tamanho da pilha for suficientemente aumentado, a perturbação causada por tal nêutron que entra muito provavelmente continuará aumentando até que toda a pilha seja destruída. Existe um fenômeno correspondente para as mentes e existe um para as máquinas? Parece haver um para a mente humana.A maioria deles parece ser 'subcrítico',*ou seja* , para corresponder nesta analogia a pilhas de tamanho subcrítico. Uma ideia apresentada a tal mente dará, em média, menos de uma ideia em resposta. Uma pequena proporção é supercrítica. Uma ideia apresentada a tal mente pode dar origem a toda uma 'teoria' consistindo de ideias secundárias, terciárias e mais remotas. As mentes dos animais parecem ser definitivamente subcríticas. Aderindo a essa analogia, perguntamos: 'Pode uma máquina 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. Dizemos que isso não corresponde à mente real: é uma espécie de pele que devemos despir se quisermos encontrar a mente real. Mas então, no que resta, encontramos mais uma pele a ser arrancada, e assim por diante. Procedendo desta forma, alguma vez chegamos à mente 'real', ou eventualmente chegamos à pele que não tem nada dentro dela? 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 deveriam ser descritos como 'recitações que tendem a produzir crença'.

O único suporte realmente satisfatório que pode ser dado para a visão expressa no início do § 6, será aquele fornecido pela espera do fim do século e então fazendo o experimento descrito. Mas o que podemos dizer enquanto isso? Que medidas devem ser tomadas agora para que a experiência seja bem-sucedida?

Como já expliquei, o problema é principalmente de programação. Avanços na engenharia também terão que ser feitos, mas parece improvável que eles não sejam adequados para os requisitos. As estimativas da capacidade de armazenamento do cérebro variam de 10^{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 altos de pensamento. A maior parte é provavelmente usada para a retenção de impressões visuais. Eu ficaria surpreso se mais de 10^{9} fossem necessários para jogar satisfatoriamente o jogo da imitação, pelo menos contra um cego. (Nota—A capacidade da *Encyclopaedia Britannica*, 11ª edição, é 2×10^{9}.) Uma capacidade de armazenamento de 10^{7}seria uma possibilidade muito praticável mesmo pelas técnicas atuais. Provavelmente não é necessário aumentar a velocidade das operações 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 poderia cobrir as perdas de velocidade decorrentes de várias maneiras. Nosso problema então é descobrir como programar essas máquinas para jogar o jogo. No meu ritmo atual de trabalho, produzo cerca de mil dígitos de programa por dia, de modo que cerca de sessenta trabalhadores, trabalhando de forma constante ao longo dos cinquenta anos, podem realizar o trabalho, se nada for para a cesta de papéis. Algum método mais rápido parece desejável.

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

- O estado inicial da mente, digamos no nascimento,
- A educação a que foi submetido,
- Outra experiência, a não ser qualificada de educação, a que tenha sido submetida.

Em vez de tentar produzir um programa para simular a mente adulta, por que não tentar produzir um que simule a da criança? Se este fosse então submetido a um curso adequado de educação, obter-se-ia o cérebro adulto. Presumivelmente, o cérebro infantil é algo como um bloco de notas, comprado nas papelarias. 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 podemos supor, como primeira aproximação, ser praticamente a mesma que para a criança humana.

Assim, dividimos nosso problema em duas partes. O programa-criança e o processo educativo. Estes dois permanecem muito intimamente ligados. Não podemos esperar encontrar uma boa máquina infantil na primeira tentativa. É preciso experimentar ensinar uma dessas máquinas e ver como 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 filha	= Material hereditário
Alterar " "	= Mutações
Seleção natural	= Julgamento do experimentador

Abrir em nova guia

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, deve ser capaz de acelerá-lo. Igualmente importante é o fato de que ele não está restrito a mutações aleatórias. Se ele puder traçar a causa de alguma fraqueza, provavelmente poderá pensar no tipo de mutação que a melhorará.

Não será possível aplicar exatamente o mesmo processo de ensino à máquina que a uma criança normal. Não será, por exemplo, provido de pernas, de modo que não possa ser solicitado a sair e encher o balde 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 zombem demais dela. Deve ser dada alguma instrução. Não precisamos estar muito preocupados 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 por algum meio ou outro.

Normalmente associamos punições e recompensas ao processo de ensino. Algumas máquinas-filho simples podem ser construídas ou programadas com esse tipo de princípio. A máquina deve ser construída de tal forma que os eventos que logo precederam a ocorrência de um sinal de punição dificilmente se repitam, enquanto um sinal de recompensa aumenta a probabilidade de repetição dos eventos que levaram a ele. Essas definições não pressupõem nenhum sentimento por parte da máquina. Fiz alguns experimentos com uma dessas máquinas infantis e consegui ensinar algumas coisas, mas o método de ensino era muito heterodoxo para que o experimento fosse considerado realmente bem-sucedido.

O uso de punições e recompensas pode, na melhor das hipóteses, fazer parte do processo de ensino. Grosso modo, se o professor não tem outro meio de comunicação com o aluno, a quantidade de informações que pode chegar até ele não excede o número total de recompensas e punições aplicadas. No momento em que uma criança aprende a repetir 'Casabanca', 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 alguns outros canais de comunicação 'sem emoção'. Se estes estiverem disponíveis, é possível ensinar uma máquina por punições e recompensas a obedecer a ordens dadas em algum idioma, *por exemplo* ,uma linguagem simbólica.

Essas ordens devem ser transmitidas através dos canais 'sem emoção'. O uso dessa linguagem diminuirá muito o número de punições e recompensas exigidas.

As opiniões podem variar quanto à complexidade que é adequada na máquina filha. Pode-se tentar torná-lo o mais simples possível consistentemente com os princípios gerais. Alternativamente, pode-se ter um sistema completo de inferência lógica 'embutido'. [1] Neste último caso, a loja estaria amplamente ocupada com definições e proposições. As proposições teriam vários tipos de status, *por exemplo* ,fatos bem estabelecidos, conjecturas, teoremas matematicamente provados, declarações dadas por uma autoridade, expressões tendo a forma lógica de proposição, mas não de valor de crença. Certas proposições podem ser descritas como 'imperativos'. A máquina deve ser construída de tal forma que assim que um imperativo for classificado como 'bem estabelecido' a ação apropriada automaticamente ocorra. Para ilustrar isso, suponha que o professor diga à máquina: 'Faça sua lição de casa agora'. Isso pode fazer com que "Professor diz 'Faça sua lição de casa agora'" seja incluído entre os fatos bem estabelecidos. Outro fato pode ser,

"Tudo o que o professor diz é verdade". A combinação destes pode eventualmente conduzir ao imperativo, 'Faça já o seu dever de casa', sendo incluído entre os factos bem estabelecidos, e isto, pela construção da máquina, fará com que o dever de casa realmente comece, mas o efeito é muito satisfatório . Os processos de inferência usados pela máquina não precisam ser tais que satisfaçam os lógicos mais exigentes. Por exemplo, pode não haver hierarquia de tipos. Mas isso não significa necessariamente que ocorrerão falácias de tipo, assim como não estamos fadados a cair de penhascos não cercados. Imperativos adequados (expressos *dentro* dos sistemas, não fazendo parte das regras *de*o sistema) como 'Não use uma classe a menos que seja uma subclasse de uma que foi mencionada pelo professor' pode ter um efeito semelhante a 'Não vá muito perto da borda'.

Os imperativos que podem ser obedecidos por uma máquina que não tem membros são obrigados a ser de caráter bastante intelectual, como no exemplo (fazer lição de casa) dado acima. Importante entre tais imperativos serão aqueles que regulam a ordem em que as regras do sistema lógico em questão devem ser aplicadas. Pois em cada estágio quando se está usando um sistema lógico, há um número muito grande de passos alternativos, qualquer um dos quais é permitido aplicar, no que diz respeito à obediência às regras do sistema lógico. Essas escolhas fazem a diferença entre um raciocinador brilhante e um raciocinador insignificante, não a diferença entre um raciocínio sólido e um falacioso. Proposições que levam a imperativos desse tipo podem ser "Quando Sócrates é mencionado, use o silogismo em Bárbara" ou "Se um método provou ser mais rápido que outro,não use o método mais lento". Alguns deles podem ser 'dados por autoridade', mas outros podem ser produzidos pela própria máquina,*por exemplo* , por indução científica.

A ideia de uma máquina de aprendizado 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 possa sofrer. As regras são, portanto, bastante invariantes no tempo. Isso é bem verdade. A explicação do paradoxo é que as regras que são alteradas no processo de aprendizagem são de um tipo bem 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 aprendizagem é que seu professor muitas vezes ignora o que está acontecendo lá dentro, embora ainda seja capaz de prever, até certo ponto, o comportamento de seu aluno. Isso deve se aplicar mais fortemente à educação posterior de uma

máquina resultante de uma máquina infantil de projeto (ou programa) bem testado. Isso está em claro contraste com o procedimento normal ao usar uma máquina para fazer cálculos: o objetivo de alguém é, então, ter uma imagem mental clara do estado da máquina em cada momento do cálculo. Este objetivo só pode ser alcançado com uma luta. A visão de que 'a máquina só pode fazer o que sabemos ordenar que ela faça', [1]parece estranho diante disso. A maioria dos programas que podemos colocar na máquina resultará em algo que não podemos entender, ou que consideramos um comportamento completamente aleatório. O comportamento inteligente, presumivelmente, consiste em um afastamento do comportamento completamente disciplinado envolvido na computação, mas bastante leve, que não dá origem a um comportamento aleatório ou a loops repetitivos inúteis. Outro resultado importante de preparar nossa máquina para seu papel no jogo da imitação por um processo de ensino e aprendizagem é que a 'falibilidade humana' provavelmente será omitida de uma maneira bastante natural, *ou seja* ,sem 'coaching' especial. (O leitor deve conciliar isso com o ponto de vista das pp. 24, 25.) Os processos que são aprendidos não produzem cem por cento. certeza do resultado; se o fizessem, não poderiam ser desaprendidos.

Provavelmente é aconselhável incluir um elemento aleatório em uma máquina de aprendizado (ver p. 438). Um elemento aleatório é bastante útil quando estamos procurando a solução de algum problema. Suponha, por exemplo, que queiramos encontrar um número entre 50 e 200 que seja igual ao quadrado da soma de seus dígitos, podemos começar em 51 e depois tentar 52 e continuar até obtermos um número que funcione. Alternativamente, podemos escolher números aleatoriamente até obtermos um bom. Este método tem a vantagem de não ser necessário acompanhar os valores que foram tentados, mas a desvantagem de poder 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 bloco enorme sem soluções na região que deve ser investigada primeiro.

Podemos esperar que as máquinas venham a 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 jogar xadrez, seria melhor. Pode-se também sustentar que é melhor fornecer à máquina os melhores órgãos dos sentidos que o dinheiro pode comprar, e então ensiná-la a entender e falar inglês. Este processo poderia seguir o ensino normal de uma criança. As coisas seriam apontadas e nomeadas, etc. Novamente, não sei qual é a resposta certa, mas acho que ambas as abordagens devem ser tentadas.

Só podemos ver uma curta distância à frente, mas podemos ver muito que precisa ser feito.

[1] Possivelmente esta visão é herética. São Tomás de Aquino (*Suma Teológica* , citado por Bertrand Russell, p. 480) afirma que Deus não pode fazer um homem sem alma. Mas isso pode não ser uma restrição real aos Seus poderes, mas apenas um resultado do fato de que as almas dos homens são imortais e, portanto, indestrutíveis.

[1] Os nomes dos autores em itálico referem-se à Bibliografia.

[1] Ou melhor 'programado' para que nossa máquina-filho seja programada em um computador digital. Mas o sistema lógico não terá que ser aprendido.

[1] Compare a declaração de Lady Lovelace (p. 450), que não contém a palavra 'somente'.

BIBLIOGRAFIA

Samuel
Mordomo
,
Erewhon
,
Londres
,
1865
.
Capítulos 23, 24, 25
,
O livro das máquinas
.

Google Scholar

Alonzo
Igreja
, “
Um problema insolúvel da teoria elementar dos números
”,
Americano J. de Matemática.
,
58
(
1936
),
345
–
363
.

Google Scholar

Referência cruzada

K.
Gödel
, “
Über formal unentscheildbare Sätze der Principia Mathematica und verwandter Systeme, I
”,
Monatshefle für Math, und Phys.
, (
1931
),

173
–
189
.

Google Scholar

RD
Hartree
,
Calculando Instrumentos e Máquinas
,
Nova Iorque
,
1949
.

SC
Kleene
, “
Funções recursivas gerais de números naturais
”,
Americano J. de Matemática.
,
57
(
1935
),
153
–
173
e
219
–
244
.

Google Scholar

Referência cruzada

G.
Jefferson
, “
A Mente do Homem Mecânico”. Oração Lister para 1949
.

Jornal médico britânico
, vol.
eu
(
1949
),
1105
–
1121
.

Google Scholar

Referência cruzada

Condessa de Lovelace
, '
Notas do tradutor para um artigo sobre o Engiro Analítico de Babbage
',
Memórias Científicas
(ed. por
R.
Taylor
), v.
3
(
1842
),
691
–
731
.

Google Scholar

Bertrand
Russel
,
História da filosofia ocidental
,
Londres
,
1940
.

Google Scholar

SOU
Turing
, “
Em números computáveis, com uma aplicação ao Entscheidungsproblem
”,
Proc. Matemática de Londres. Soc.
(
2
),
42
(
1937
),
230
–
265
.

Google Scholar

Universidade Victoria de Manchester.