

A estranha inversão de raciocínio de Turing

Alguns dos maiores e mais revolucionários avanços na ciência têm feito notar a sua expressão inicial em termos atractivamente modestos, sem qualquer fanfarra. Charles Darwin conseguiu comprimir toda a sua teoria num só resumo de um parágrafo que qualquer leigo pode facilmente acompanhar, em todos os seus detalhes:

Se, sob condições de vida variáveis, os seres vivos apresentam diferenças individuais em quase todas as partes da sua estrutura (facto indisputável); se, porque aumentam de número em progressão geométrica, passam por uma forte luta pela sobrevivência em algum momento das suas vidas, idade, época ou ano (o que também não pode ser contestado); então, considerando a infinita complexidade de relações que existe entre todos os seres vivos, e entre cada um deles e as suas condições de vida, relações essas que conduzem a uma infinita diversidade de estrutura, constituição e hábitos que lhes é vantajosa, então, dizíamos, seria um facto deveras extraordinário se jamais tivesse havido uma variação útil para o bem-estar de cada ser vivo, quando são tantas as variações que ocorreram e são proveitosas para o homem. Mas se considerarmos que alguma vez surgiram num organismo variações que lhe são benéficas, então temos de aceitar que esse organismo tem mais hipóteses de vencer a luta pela sobrevivência e de ser preservado. E, devido ao forte princípio da hereditariedade, esse organismo tenderá a produzir descendentes semelhantes, transmitindo-lhes os caracteres que lhe deram vantagem. A este princípio de conservação dos caracteres, ou da sobrevivência do mais apto, dei o nome de selecção natural. (*Origem das Espécies*, fim do capítulo 4)

Francis Crick e James Watson concluíram o seu artigo sobre a estrutura do ADN que marcou toda uma época com a frase deliciosamente modesta:

Não escapou à nossa atenção que os emparelhamentos específicos que postulámos imediatamente sugerem um possível mecanismo de cópia para a unidade de vida replicante. (Watson e Crick, 1953: 738)

E Alan Turing criou um novo mundo de ciência e tecnologia, montando a cena para resolver um dos mais desconcertantes *puzzles* que permaneciam na ciência, o problema mente-corpo, com um frase declarativa ainda mais curta a meio do seu artigo de 1936 sobre números computáveis:

É possível inventar uma só máquina que pode ser usada para computar qualquer sequência computável. (Turing. 1936: 241)

Turing não apenas intuiu que este feito notável seria possível; ele mostrou exactamente como fazer uma máquina deste tipo. Com essa demonstração a era computacional tinha nascido. É importante lembrar que havia entidades chamadas computadores antes de Turing ter falado nesta ideia – mas esses eram pessoas, trabalhadores administrativos com suficientes capacidades matemáticas, paciência e orgulho no seu trabalho para gerar resultados confiáveis de horas e horas de computação, dia após dia. Muitos deles eram mulheres.



Dryden Flight Research Center E49-0053 Photographed 10/49
Early "computers" at work. NASA photo



Milhares deles eram empregados em engenharia e comércio, nas forças armadas e noutros postos, calculando tabelas para usos em navegação, artilharia e outros empreendimentos técnicos. Uma boa maneira de compreender a ideia revolucionária de Turing sobre a computação é colocá-la em justaposição com a de Darwin sobre a evolução. O mundo pré-Darwiniano era sustentado, não pela ciência, mas pela tradição: todas as coisas no universo, desde a mais exaltada (“homem”), à mais humilde (a formiga, o calhau, e a gota de água), seriam criações de uma coisa ainda mais exaltada, Deus, um criador inteligente onnipotente e onnipresente – que apresentava uma semelhança impressionante com a segunda coisa mais exaltada. Chamemos a isso a teoria da

criação pelo espalhamento das riquezas a partir do topo. Darwin substituiu-a pela teoria da criação como “borbulhar a partir do solo”. Um dos críticos de Darwin do século XIX apresenta-a brevemente:

Na teoria com a qual temos de lidar, a Absoluta Ignorância é o artífice; então devemos enunciar como princípio fundamental de todo o sistema que, DE MODO A FAZER UMA MÁQUINA PERFEITA E BONITA, NÃO É REQUISITO NECESSÁRIO SABER COMO FAZÊ-LA. Esta proposição será encontrada, com exame cuidadoso, como expressando, de modo condensado, a intenção essencial da Teoria, e como expressando em poucas palavras o sentido de Mr. Darwin; que, por uma estranha inversão de raciocínio, parece pensar que a Absoluta Ignorância é totalmente qualificada para tirar lugar à Absoluta Sabedoria em todas as conquistas da capacidade criativa. (MacKenzie, 1868)

Foi, de facto, uma estranha inversão de raciocínio. Até hoje muitas pessoas não conseguem entender a inquietante ideia de que um processo sem propósito e sem mente ocorra durante eões, gerando organismos cada vez mais subtis, eficientes e complexos sem qualquer sopro vital ou entendimento do que está a fazer.

A ideia de Turing era uma semelhante – aliás, extraordinariamente semelhante – estranha inversão de raciocínio. O mundo pré-Turing era um mundo no qual os computadores eram pessoas que tinham que entender matemática de modo a realizar os seus trabalhos. Turing percebeu que isto simplesmente não era necessário: era possível pegar nas funções que realizavam e espremer as últimas gotas do entendimento, deixando apenas acções brutas e mecânicas. PARA SER UMA MÁQUINA DE COMPUTAÇÃO PERFEITA E BELA NÃO É UM REQUISITO SABER O QUE É ARITMÉTICA.

O que Darwin e Turing descobriram, de modos diferentes, foi a existência de *competência sem compreensão* (cf. Dennett, 2009, do qual material nos parágrafos anteriores foi esboçado, com revisões). Isto inverteu a suposição profundamente plausível de que a compreensão é de facto a *fonte* de toda a competência avançada. Afinal, por que insistimos em mandar as nossas crianças para a escola, e por que franzimos a sobrancelha aos métodos antiquados de aprendizagem habitual? Nós esperamos que a competência crescente das nossas crianças flua da sua crescente compreensão; o lema da educação moderna deve ser: “compreender *a fim de ser* competente”. E para nós, membros da espécie *Homo sapiens*, esta é quase sempre a melhor maneira de olhar e esforçar pela competência. Eu suspeito que este muito amado princípio da educação é um dos motivadores primários do cepticismo em relação tanto à evolução como à sua prima do mundo de Turing, a Inteligência Artificial. A própria ideia de que a mecanicidade sem mente possa gerar competências de nível humano – ou de nível divino! - atinge muitos como filistina, repugnante, um insulto às

nossas mentes e à mente de Deus.

Consideremos agora como Turing fez a sua prova. Ele utilizou computadores humanos como o seu modelo. Eles sentaram-se nas suas secretárias, fazendo apenas um passo simples e altamente confiável de cada vez, verificando o seu trabalho, escrevendo os resultados intermediários em vez de confiar nas suas memórias, consultando as suas instruções quantas vezes achassem necessário, tornando o que parecia à primeira vista uma tarefa difícil numa rotina que poderiam quase desempenhar enquanto dormiam. Turing sistematicamente reduzia os passos simples a passos ainda mais simples, removendo todos os vestígios de discernimento e compreensão. Poderia um computador humano ter dificuldade em distinguir o número 9999999999 do número 999999999? Depois, reduziu o problema perceptual de *reconhecer o número* em problemas mais simples, distribuindo actos mais fáceis e estúpidos de discriminação em múltiplos passos. Então ele preparou um inventário de *building blocks* básicos dos quais construir o algoritmo universal que pudesse executar qualquer outro algoritmo. Ele mostrou como esse algoritmo iria permitir ao computador (humano) computar qualquer função, e notou que:

O comportamento do computador em qualquer momento é determinado pelos símbolos que ele está a observar e pelo seu “estado de espírito” nesse momento. Podemos supor que há um limite L ao número de símbolos ou quadrados que o computador pode observar nesse momento. Se ele deseja observar mais, ele deve usar observações sucessivas. ... A operação realmente executada é determinada ... pelo estado de espírito do computador e os símbolos observados. Em particular, eles determinam o estado de espírito do computador depois da operação ser executada.

Ele notou depois, calmamente:

Podemos agora construir uma máquina para fazer o trabalho deste computador (p.251)

Aqui vemos uma redução de *toda a computação possível* a um processo sem mente. Podemos começar com os *building blocks* simples que Turing isolou, e construir camada sobre camada de computação mais sofisticada, restaurando gradualmente, a inteligência que Turing havia tão habilmente excluído das práticas de computadores humanos.

Mas o que dizer do génio de Turing e de programadores menores mais tardios, cuja compreensão inteligente foi manifestamente a fonte dos *designs* que podem unir os *building blocks* sem mente de Turing em competências úteis? Esta dependência não reintroduz apenas a perspectiva “espalhamento a partir do topo” acerca da inteligência, com Turing no papel de Deus? Não menos que um pensador como Roger Penrose expressou o seu cepticismo sobre a possibilidade de que a Inteligência Artificial pudesse ser fruto de nada além de processos algorítmicos sem mente.

Eu sou um forte crente no poder da selecção natural. Mas eu não vejo como a selecção natural em si mesma possa fazer evoluir algoritmos que pudessem ter o tipo de julgamentos conscientes da validade de outros algoritmos que nós parecemos ter. (1989, p.414)

Ele admite:

Na minha maneira de pensar ainda há algo misterioso acerca da evolução, com o seu aparente “fazer tentativas” direccionamento a um propósito futuro. As coisas pelo menos *parecem* organizar-se a si mesmas de um modo algo melhor do que elas “precisam”, apenas com base na evolução casual e selecção natural. (1989, p.416)

De facto, pareceria improvável que uma só cadeia de eventos de selecção natural, ocorrendo durante biliões de anos, fosse capaz de criar uma cadeia de zeros e uns que, assim que lidos por um computador digital, fossem um “algoritmo” para “juízos conscientes.” Mas como Turing percebeu perfeitamente, não havia nada que evitasse que o processo da evolução se copiasse a si mesmo em várias escalas, montasse discernimento e julgamento. O passo recursivo que fizesse com que a bola rolasse – conceber um computador que conseguisse imitar qualquer outro computador – poderia ele mesmo ser reiterado, permitindo computadores específicos melhorar os seus próprios poderes *redesenhando-se a si mesmos*, deixando o seu *designer* original para trás. Já em “Computing Machinery and Intelligence”, o seu artigo clássico em *Mind*, 1950, ele reconheceu que não havia contradição no conceito de um computador (não-humano) que pudesse aprender.

A ideia de uma máquina que aprende pode parecer paradoxal a alguns leitores. Como podem as regras de operação da máquina mudar? Elas deveriam descrever completamente como a máquina irá reagir qualquer que seja o seu historial, quaisquer que sejam as mudanças por que passam. As regras são, então, bastante invariantes no tempo . Isto é realmente verdade. A explicação do paradoxo é que as regras que podem ser mudadas no processo de aprendizagem são de um tipo bastante menos pretensioso, requerendo apenas uma validade efêmera. O leitor pode fazer um paralelo com a Constituição dos Estados Unidos. (Cf. Suber, 2001, não publicado, para uma discussão valiosa desta passagem e do chamado paradoxo do auto-aperfeiçoamento).

Ele viu claramente que toda a versatilidade e auto-mutabilidade do pensamento humano – aprendizagem e re-avaliação, e linguagem e resolução de problemas, por exemplo – poderiam em princípio ser construídas a partir desses *building blocks*. Chamemos a isto a teoria da mente “borbulhar a partir do solo”, e contrastemo-la com as várias teorias da mente “espalhamento a partir do topo” de pensadores como René Descartes até John Searle (e incluindo, notavelmente, Kurt

Gödel cuja prova foi a inspiração para o trabalho de Turing) que começam com a consciência humana na sua maior reflexão, e depois são incapazes de unir tais poderes mágicos com os meros mecanismos dos corpos e cérebros humanos.

Turing, tal como Darwin, decompôs o mistério da inteligência (ou Design Inteligente) no que podemos chamar de passos atômicos de acasos não inteligentes que, quando acumulados aos milhões, se somam num tipo de pseudo-inteligência. A Unidade de Processamento Central de um computador não sabe *realmente* o que é a aritmética, nem compreende o que é a adição, mas ela “compreende” o “comando” para adicionar dois números e colocar a sua soma numa registadora – no sentido mínimo de que ela adiciona com segurança quando é chamada para adicionar e coloca a soma no local correcto. Digamos que ela *como que* compreende a adição. Alguns níveis acima, quando os *building blocks* são montados aos bilhões e triliões, o programa de jogo de xadrez não compreende *realmente* que a sua rainha está em perigo, mas *como que* compreende isso, e o Watson do IBM no jogo Jeopardy *como que* compreende as questões a que responde.

Porque havemos de permitir esta conversa do “*como que*”? Porque quando analisamos – ou sintetizamos – este monte de níveis cada vez mais competentes, temos de nos manter a par de dois factos sobre cada nível: o que *é* e o que *faz*. O que *é* pode ser descrito em termos da organização estrutural das partes das quais é feito - desde que assumamos que as partes funcionam como são supostas funcionar. O que *faz* em alguma função (cognitiva) que (como que) executa – bem o suficiente para que no nível seguinte possamos assumir que temos no nosso inventário um bloco construtor mais inteligente que executa apenas essa função – como que, bom o suficiente para utilizar. Esta é a chave para quebrar a espinha dorsal da questão complexa incompreensível de como uma mente poderia alguma vez ser composta de mecanismos materiais.

Examinando o Trabalho e o seu Impacto tardio

O que poderíamos chamar de operador *como que* é, em ciência cognitiva, o paralelo do gradualismo de Darwin nos processos evolutivos. Antes de haverem bactérias havia *uma espécie de* bactéria, antes de haver mamíferos havia *uma espécie de* mamíferos e antes de haver cães havia *uma espécie de* cães, e assim por diante. Precisamos do gradualismo de Darwin para explicar a grande diferença entre um robot humanóide e uma máquina calculadora. O símio e a maçã são feitos dos mesmos ingredientes básicos, diferentemente estruturados e explorados numa cadeia de vários níveis de competências funcionais diferentes. Não há numa linha de princípio que separe *uma espécie de* símio de um símio. O robot humanóide e a máquina calculadora são ambos feitos dos mesmos blocos-básicos-de-Turing, que não pensam e não sentem, mas quando os compomos em estruturas maiores e mais competentes, se tornam depois os elementos de estruturas ainda mais

competentes em níveis mais elevados, eventualmente chegando a peças tão *como que* inteligentes que podem ser reunidas em competências que merecem ser chamadas de compreensão. Utilizamos a estratégia intencional (Dennett, 1971, 1987) para manter a pista das crenças e desejos (ou “crenças” e “desejos” ou *espécies de* crenças e desejos) dos agentes *como que* racionais em todos os níveis, desde as bactérias mais simples através de todos os circuitos discriminatórios, assinaladores, comparativos e de memória que compõem os cérebros dos animais, desde estrelas-do-mar a astrónomos. Não há nenhuma linha divisória acima da qual a verdadeira compreensão possa ser encontrada – mesmo no nosso caso. A criança pequena *como que* compreende a sua própria frase “O papá é um médico”, e eu *como que* compreende “ $E = mc^2$ ”. Alguns filósofos resistem a este anti-essencialismo: ou acreditamos que a neve é branca ou não; ou somos conscientes ou não; nada conta como uma aproximação de qualquer fenómeno mental – é tudo ou nada. E para tais pensadores, os poderes das mentes são mistérios impossíveis de solucionar porque são “perfeitos”, e perfeitamente improváveis de serem encontrados em meros mecanismos materiais.

Ainda não chegámos à compreensão “real” em robots, mas estamos a aproximar-nos. Essa, pelo menos, é a convicção daqueles que foram inspirados pelo *insight* de Turing. Os teóricos do “espalhamento a partir do topo” sentem estar completamente certos de que nenhuma quantidade de construção avançada poderá levar-nos à coisa real. Eles pensam que uma *res cogitans* Cartesiana, uma coisa pensante, não pode ser construída com os *building blocks* de Turing. E os criacionistas estão similarmente certos que nenhuma quantidade de “baralhar” e cópia e selecção pode sequer chegar a coisas vivas (reais). Eles estão errados, mas podemos apreciar o desconforto que motiva a sua convicção.

A estranha inversão de raciocínio de Turing, tal como a de Darwin, vai contra a corrente de milénios de pensamento anteriores. Se a história de resistência ao pensamento de Darwin é uma boa medida, podemos esperar que no futuro, muito depois de todos os triunfos do pensamento humano serem atingidos ou ultrapassados por “meras máquinas”, já não haverá pensadores a insistir que a mente humana funciona de maneiras misteriosas que nenhuma ciência pode compreender.

Referências

Darwin, C., 1859. *On the Origin of Species*, John Murray, London.

Dennett, D.C., 1971. Intentional systems. *J. Phil.*, 68, 87–106.

Dennett, D.C., 1987. *The Intentional Stance*, MIT Press, Cambridge, MA.

Dennett, D.C., 2009. Darwin’s Strange inversion of reasoning. *PNAS*, 106, 10061–10065.

MacKenzie, R.B., 1868. *The Darwinian Theory of the Transmutation of Species Examined*, Nisbet & Co, London.

Penrose, R., 1989. *The Emperor’s New Mind*, Oxford University Press, Oxford.

Suber, P., (unpublished). Saving Machines From Themselves: The Ethics of Deep Self Modification, preprint, 30 November 2001. <http://www.earlham.edu/~peters/writing/selfmod.htm>

Turing, A., 1936. On computable numbers, with an application to the Entscheidungsproblem. Proc. Lond. Math. Soc. 42, 23–265, and erratum (1937) 43, 544–546.

Turing, A., 1950. Computing machinery and intelligence. Mind, LIX, 2236, 433–460.

Watson, J.D. and Crick, F.H.C., 1953. A structure for deoxyribose nucleic acid. Nature 171, 737–738.